2010

Robust Dialog Management Through A Context-centric Architecture

Victor C. Hung
University of Central Florida

ROBUST DIALOG MANAGEMENT
THROUGH A CONTEXT-CENTRIC
ARCHITECTURE

by

VICTOR C. HUNG
B. S. Carnegie Mellon University, 2002
M. S. Carnegie Mellon University, 2002

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Electrical Engineering and Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2010

Major Professor: Avelino J. Gonzalez

ABSTRACT

This dissertation presents and evaluates a method of managing spoken dialog interactions with a robust attention to fulfilling the human user's goals in the presence of speech recognition limitations. Assistive speech-based embodied conversation agents are computer-based entities that interact with humans to help accomplish a certain task or communicate information via spoken input and output. A challenging aspect of this task involves open dialog, where the user is free to converse in an unstructured manner. With this style of input, the machine's ability to communicate may be hindered by poor reception of utterances, caused by a user's inadequate command of a language and/or faults in the speech recognition facilities. Since a speech-based input is emphasized, this endeavor involves the fundamental issues associated with natural language processing, automatic speech recognition and dialog system design. Driven by Context-Based Reasoning, the presented dialog manager features a discourse model that implements mixed-initiative conversation with a focus on the user's assistive needs. The discourse behavior must maintain a sense of generality, where the assistive nature of the system remains constant regardless of its knowledge corpus. The dialog manager was encapsulated into a speech-based embodied conversation agent platform for prototyping and testing purposes. A battery of user trials was performed on this agent to evaluate its performance as a robust, domain-independent, speech-based interaction entity capable of satisfying the needs of its users.

Dedicated to Mom & Dad.

ACKNOWLEDGMENTS

First and foremost, I want to show gratitude to my family and closest friends for their never-ending support of all of my endeavors. I would like to thank Drs. Avelino Gonzalez, Ronald DeMara, Kenneth Stanley, Lotzi Bölöni, and Richard Gilson for serving on my dissertation committee. Much appreciation goes out to my Intelligent Systems Laboratory colleagues – Johann, Viet, Gary, Carlos, Raul, Miguel, J. R., Feras, and Ruben. I would also like to thank AT&T Inc., the Link Foundation and the University of Central Florida Institute of Simulation and Training for their fellowships to make this important research a reality. A show of thanks goes out to Alex Schwarzkopf, Rita Rodriguez, Edward Clancy, Rathindra DasGupta, Donald Senich, Glenn Larsen, Greg Misorek and the rest of the National Science Foundation for their continued moral support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AI	Artificial Intelligence
AIML	Artificial Intelligence Markup Language
ALICE	Artificial Linguistic Internet Chat Entity
ASR	Automatic Speech Recognition
BDI	Belief-Desire-Intention
CBR	Case-Based Reasoning
CCBR	Conversational Case-Based Reasoning
CxBR	Context-Based Reasoning
DSS	Decision Support System
EA	Evolutionary Algorithm
ECA	Embodied Conversation Agent
FSM	Finite State Machine
GA	Genetic Algorithm
GUI	Graphical User Interface
HCI	Human-Computer Interaction
IE	Information Extraction
IR	Information Retrieval
ISL	Intelligent Systems Laboratory
ISU	Information State Update
I/UCRC	Industry/University Cooperation Research Center
ML	Machine Learning

NLP	Natural Language Processing
NLU	Natural Language Understanding
NN	Neural Network
NSF	National Science Foundation
PDA	Personal Digital Assistant
POS	Part-of-Speech
QA	Question-Answering
RFID	Radio-Frequency Identification
UCF	University of Central Florida
WER	Word-Error Rate
WoZ	Wizard-Of-Oz
WWW	World Wide Web

CHAPTER ONE: INTRODUCTION

This dissertation deals with the development of a prototypical speech-based dialog management system for the sake of serving as an information deployment tool in the form of an assistive conversation agent. A novel aspect of this prototype is the centralized use of *contextualization* to drive a dialog exchange. The proposed system, the CONtext-centric Corpus-based Utterance Robustness dialog manager, or CONCUR, was built with three major themes in mind: 1) overcoming the limitations of automatic speech recognition (ASR), 2) supporting a domain-independent knowledge management system, and 3) providing an open dialog discourse style. The purpose of this dissertation is to provide an account of the journey in which CONCUR came to existence. To begin weaving this story, this chapter introduces the fundamental issues that inspired the development of a context-based dialog manager.

Text-based exchanges have been the de facto standard of effective communication between humans and computers. These interactions often exist as a person's sequence of keyboard taps and mouse clicks in response to the demands of the software interface. With the emergence of ELIZA (Weizenbaum, 1966), the computing world was introduced to conversation-based human-computer interaction (HCI). While quite limited in its utility and execution, ELIZA provided a glimpse into the future of machine interfaces, where humans naturally converse with computers instead of performing keystroke commands and double-clicks. Additionally, ASR technology has emerged to someday replace the keyboard and mouse with a microphone. The melding of conversation agents with ASR has made the idea of speech-based agents a reality.

Conceptually, combining conversation agents, or *chatbots*, with voice technology seems simple enough as a winning formula for creating fully autonomous talking computers. Mass media has led us into thinking this is the near future of computing, with Kubrick's HAL, Knight Rider's KITT, Star Trek's Commander Data and Lucas' C3PO in mind. (Breazeal, 2005) Historically, however, this recipe has been executed without much success. For one, the development of agents capable of *natural*, or human-like conversations has not made significant enough progress beyond ELIZA. Secondly, ASR technology has also not been perfected, making human voice a less-than-reliable method of interacting with a computer. In this dissertation, these bounds of chatbot limitations and ASR imperfection are addressed with the help of context-based methods. The research presented here contributes a generalized architecture for empowering speech-based *embodied conversation agents* (ECA), or chatbots with a physical presence beyond that of a text-based chatroom, with open dialog understanding in lieu of error-prone ASR.

The main problem being tackled in this dissertation is overcoming poor ASR performance when trying to conduct an effective spoken conversation with an assistive ECA. The following list outlines the major themes that are touched upon in addressing this problem:

- The design of **conversation agents** with respect to discourse modeling as well as input/output handling
- Support for providing an **open dialog** such that a user is not constrained as to how s/he can respond
- The role of **automatic speech recognition** in speech-based conversation agents and its impact as a weak link in relaying reliable user utterance information

- The importance of **domain-independent knowledge management** when developing ECA technology

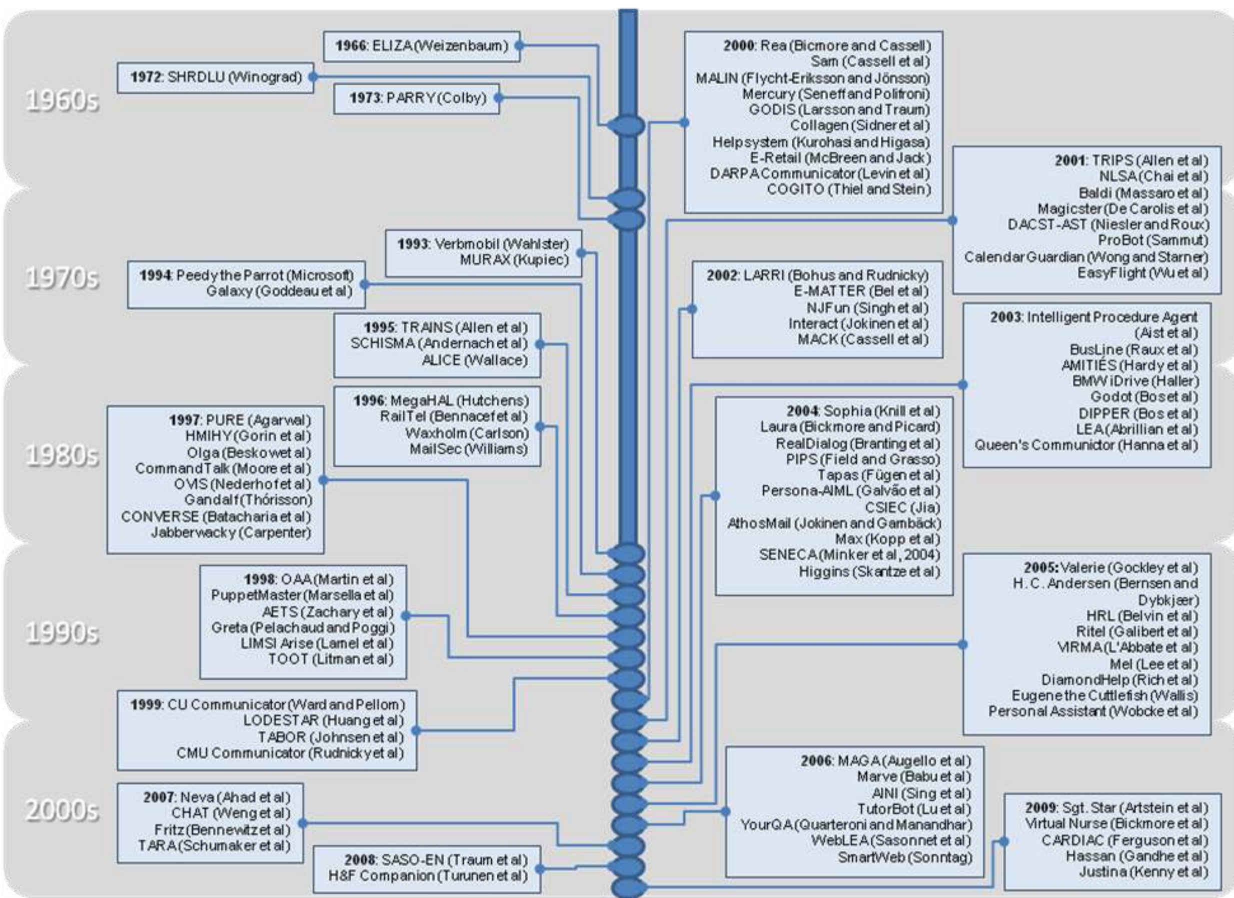
This dissertation proposes the use of a context-centric dialog management system as a solution to the aforementioned problem. For the sake of this system, the concept of *contextualization* permeates each of the above items, serving as a common thread that essentially embodies the whole point of this work. The remainder of this chapter expands on framing the overall problem by giving a brief historical introduction to these themes.

Conversation Agents

The idea of a conversation-based HCI began as a text-based endeavor with the creation of ELIZA (Weizenbaum, 1966) in the late 1960's. ELIZA initiated the chatbot movement with its ability to take in conversational text exchanges and return a coherent, human-like response if only for a brief period of time. This natural language input-output processing, however, turned out to be a template-based output utterance look-up table triggered by a matched word pattern. Despite its simplicity, Weizenbaum's work sparked the progression toward natural dialog HCI. The formation of such a fantastic machine that could "understand" and converse painted a picture of an incredibly promising horizon for the computing world.

Fifty years would pass, however, and the major steps to move beyond ELIZA toward C3PO have yet to be taken. ELIZA, despite its shortcomings as a prototype chatbot, still exists in a modernized form as the Artificial Linguistic Internet Chat Entity (ALICE). (Wallace, 2002) In the decade following ELIZA's inception, PARRY (Colby, 1973), added a statistically-powered dimension to chatbot technologies, offering more authentic speech actions through a corpus built from a machine learning (ML) analysis of paranoia patient essays. Years later, Winograd's SHRDLU (1980) excited the artificial intelligence (AI) world with its seemingly conscious grasp of the world, using a highly domain-constrained model of its own Blocks World to answer highly domain-constrained questions – a concept to be later emulated in the next century (Stoness et al, 2005).

The decades following ELIZA saw gradual improvement in augmenting the realism and complexity of chatting agents. Figure 1 gives a timeline of chatbot technology.



Compiled from: (Agarwal, 1997) (Ahad et al, 2007) (Allen et al, 1996; 2001) (Andernach et al, 1995) (Artstein et al, 2009) (Atwell and Shawar, 2007) (Augello et al, 2006) (Babu et al, 2006) (Ball, 1999) (Batacharia et al, 1997) (Belvin et al, 2005) (Bernsen and Dybkjaer, 2005) (Bickmore and Cassell, 2000) (Bickmore and Picard, 2004) (Bickmore et al, 2009) (Bohus and Rudnicky, 2002; 2003) (Branting et al, 2004) (Cassell et al, 1999; 2000) (Chai et al, 2001) (Charfuelán et al, 2002) (Colby, 1973) (Ferguson et al, 2009) (Field and Grasso, 2004) (Flycht-Eriksson and Jönsson, 2000) (Fügen et al , 2004) (Galibert et al, 2005) (Galvão et al, 2004) (Gandhe et al, 2009) (Gockley et al, 2005) (Gorin et al, 1997) (Hardy et al, 2003; 2004) (Hassel and Hagen, 2005) (Huang et al, 1999) (Jia, 2004) (Johnsen et al, 2000) (Jokinen and McTear, 2009) (Jokinen, 2003) (Kenny et al, 2007; 2009) (Kopp et al, 2005) (Krenn et al, 2004) (Kupiec, 1993) (Kurohashi and Higasa, 2000) (L'Abbate et al, 2005) (Lamel et al, 1998) (Lee et al, 2005) (Levin et al, 2000) (Litman et al, 1998) (Lu et al, 2006) (LuperFoy et al, 1998) (Massaro et al, 2001) (McBreen and Jack, 2000) (Minker et al, 2004) (Nederhof et al, 1997) (Niesler and Roux, 2001) (O'Shea et al, 2008) (Quarteroni and Manandhar, 2007) (Rich et al, 2005) (Sagae et al, 2009) (Sammut, 2001) (Sansonnet et al, 2006; 2007) (Sidner, 2002) (Sing et al, 2005; 2006) (Skantze et al, 2006) (Sonntag, 2006) (Thiel and Stein, 2000) (Thórisson, 1999) (Turunen et al, 2008) (Wallis, 2005) (Weizenbaum, 1966) (Williams, 1996) (Winograd, 1980) (Wobcke et al, 2005) (Wong and Starner, 2001) (Yan and Zheng, 2004)

Figure 1: Conversation agent timeline

In this figure, there exists a large gap of conversation agent inactivity starting in the early 70's when the attention of AI research turned mostly to knowledge-based systems and their associated issues. The 1990's saw the revival of chatbot technology, with special interest in specific applications, as well as the conception of the *embodied* conversation agent.

ECAs refer to those machines that exist to interact with humans from beyond the chatroom, typically helping users perform specific tasks for a variety of domains. (Ahad et al, 2007) These conversation agents include a physical embodiment, such as a virtual talking head (Massaro et al, 2001) or simply a voice on the phone (Gorin et al, 1997) (Béchet et al, 2004) or a full-blown animated character (McBreen and Jack, 2000) (Catrambone, 2002). A few recent ECA enrichments have included: cross-cultural concerns (Krenn et al, 2004), social skills (Wallis and Pelchaud, 2005) and gestural input (Bernsen et al, 2004) (Bernsen and Dybkjær, 2005).

Some of the more advanced displays of ECA technology are those affiliated with the University of Southern California's Institute for Creative Technologies (ICT). Sergeant Blackwell (Kenny et al, 2007), Sergeant Star (Artstein et al, 2009), and Hassan (Gandhe et al, 2009) each incorporate a fully animated character with full speech input and output capabilities. Sergeants Blackwell and Star are both used as military recruitment tools, while Hassan serves as a training agent for tactical questioning. This work has contributed greatly to the exploration of integrating ECAs into interactive immersive experiences. While their overall ECA design work is quite cutting edge, providing realistic agent imagery and voice output, their dialog management techniques differ greatly from those reflected in this dissertation. The main difference is that Sergeant Blackwell, Sergeant Star and Hassan insist on a human user-based response corpus for assistance in knowledge management and in ASR disambiguation. Creating this data repository calls for a laborious effort in gathering human subject testing over a long period of time before rolling out the final production ECA. In this dissertation, the ASR and the dialog discourse model benefit from a quick knowledge acquisition turnover because of the proposed encyclopedia corpus-based knowledge management system that also allows for domain independence. Further discussion of Sergeant Blackwell, Sergeant Star and Hassan will be featured throughout this work.

The point to be made about conversation agents is that after fifty years of developing this technology, the original prototype in ELIZA has not experienced the radical paradigm shift needed to make the jump to truly natural conversation-based dialog. While this dissertation makes no claim to provide such a drastic leap, it does provide a solution that is framed within the bounds of modern conversation agent technology. The next chapter covers a deeper exploration of recent chatbot capabilities.

Open Dialog

A recent feature of ECA design is the use of voice input to drive human-machine conversation. These speech-based agents have been implemented in a wide range of degrees of complexity, from the simple automated telephone operator (Karis and Dobroth, 1991) (Gorin et al, 1997) (Béchet et al, 2004) (Gustafson et al, 2008) to the potentially risky virtual air traffic controller. (DeMara and Moldovan, 1993) (Schaefer, 2001) (Ragnarsdottir and Hvannberg, 2006) The common thread among these agents is that each machine is constrained to a single track of domain expertise. By framing the focus, certain expectations on the inputs can be instantiated, providing the machine with a better chance at recognizing the user's semantic intentions.

An *open* dialog describes those HCI exchanges that do not adhere to a specific topic or predetermined goal. (Harabagiu et al, 2000) (Pan, 2002) Developing a machine that engages in *completely* open conversation remains an AI pipedream, since total semantic disambiguation is expected for such a system, and current advances in chatbot have yet to reach such linguistically accurate approaches to natural language understanding (NLU). (Jokinen and McTear, 2009) Additionally, open dialog lends itself to a very natural style of interaction, rather than the aforementioned single-track HCI systems. Early attempts at open HCI had been limited to question-answering (QA), as seen in search engine-based querying. (Harabagiu et al, 2000) The open dialog strategy is one that is not frequently utilized in recent conversation agent designs. (Jokinen and McTear, 2009)

The work presented here strives to capture the essence of open dialog as a means to create a more natural conversational interface, especially in speech-based agents. Early attempts at spoken chatbots employed explicit *response* expectations to assist in discourse. (Karis and Dobroth, 1991) (DeMara and Moldovan, 1993) (Gorin et al, 1997) (Schaefer, 2001) (Béchet et al, 2004) (Ragnarsdottir and Hvannberg, 2006) (Gustafson et al, 2008) The result of this effort was a very structured and menu-like HCI interaction with little resemblance to human-to-human dialog. This dissertation proposes a *context* expectation infrastructure to guide its speech actions. Such a framework maintains a sense of free response to the user while allowing the machine to have some control of the overall dialog. The next chapter provides some insight on how the open-dialog problem has been addressed by recent researchers.

Automatic Speech Recognition

Achieving the speech-based interactions seen in Knight Rider is still far away, as ASR technology is still in its relative infancy. The current state of conversation agents still has a long journey to maturation before it can completely satisfy this ideal view of HCI. A challenging aspect of open systems is processing user input, especially when speech-based recognition is involved. (Kang et al, 2009) Environmental factors, such as interfering background noise, can compromise a machine's ability to pick up spoken utterances. Speech recognizer systems have seen word error rates (WER) optimize at 30%, under controlled conditions. (Kang et al, 2009) Additionally, a non-native speaker could give a grammatically inaccurate response, causing even more confusion when an agent must determine the user's needs. It can be argued that even a human interlocutor can encounter these same transmission problems, such as speaking on a weak-signaled cellular phone or conversing with someone with limited language skills such as a non-native speaker. Furthermore, ASR systems must also be able to handle filler words such as 'uh' and 'um,' that are so prevalent in human conversation. (Clark and Fox Tree, 2002) These issues in speech recognition create an obstacle to developing natural speech-based ECAs when seeking to maintain an open conversation.

This dissertation proposes a method to mitigate the effects of poor ASR performance. The use of contextual information helps to overcome this obstacle, as the speech recognition only needs operate sufficiently enough to provide the system with an indication of the user's intended context. This reflects the idea that native English speakers often succeed in communicating with others that have a poor grasp of the language by identifying the context. The next chapters will expand upon this concept, supplying a historical background of context-based speech recognition followed by a description of the proposed context-centric dialog manager.

Domain-Independent Knowledge Management

Historically, ECAs have been designed with specific domains in mind. Different agents have served as virtual medical assistants (Bickmore et al, 2009), military officers (Artstein et al, 2009), and robotic receptionists (Babu et al, 2006). This variety of expertise has lead researchers to independently produce knowledge bases as part of the entire development of the conversation agent itself, where each individual conversation agent is built entirely from the ground up. ECAs would benefit from a reusable knowledge management system, similar to the separation of domain knowledge from interface architecture found in expert systems. (Gonzalez and Dankel, 1993) Such a knowledge management infrastructure would only require an injection of domain-specific expertise into the overall dialog manager, promoting domain independence.

The Artificial Intelligence Markup Language (AIML) (Galvão et al, 2004) (Lu et al, 2006) allows programmers to provide custom knowledge files to an ELIZA-based conversation agent. Large text corpuses have been autonomously mined to provide an ML-based agent response bank for chatbots. (Huang et al, 2007) While these methods do capture the spirit of domain-independent knowledge management, a rare feature found in modern ECA designs, they go about it in a somewhat inefficient manner. AIML knowledge is authored using a highly structured pattern-matching syntax, a process that can be painstakingly meticulous, often entailing a manual transformation of individual pieces of knowledge into specialized templates. The ML-based knowledge acquisition is wholly dependent on the accuracy of the learning model and the volume of its source data, which may cause flaws in the final agent response corpus. These imperfections may require an added layer of human involvement to verify the knowledge integrity.

In this dissertation, the idea of domain independence is grasped by feeding encyclopedia entry-style data into the knowledge manager. Different entries representing different domains can be readily used by the system with little human effort. The simplicity of using encyclopedia entries from already established sources eliminates the tedious knowledge modeling process found in AIML authoring. Additionally, the direct transfer of accurate entries minimizes any surprise responses that would otherwise occur from ML-driven knowledge bases. In the chapters to follow, further analysis of knowledge management in ECAs is explored, with special attention to domain independence.

Chapter Summary

This dissertation presents a context-based approach that improves the robustness of assistive spoken dialog systems. In this case, robustness refers to the ability to respond to a user's input at any given time with maximum utility and minimum conversational awkwardness. This work strives to achieve the seamless and natural spoken interaction between humans and computers that remains an ultimate goal in AI. Hollywood's articulate androids and loquacious computers forever permeate the human race's vision of its cybernetic future. Unfortunately, science fiction's portrayals of such high-tech expectations have not exactly been fulfilled, as the current state of reality has yet to produce such fanciful technology. In fact, modern AI researchers are still struggling with many of the same fundamental questions that were posed by their predecessors. (Minsky, 1961; 2006)

A speech-based dialog manager for an ECA has been proposed, one in which contextual information is used to fill in the gaps left by inaccurate ASR systems. This architecture appeals to the idea of creating a sense of open dialog, and it facilitates agent expertise modularity through domain-independent knowledge management. This chapter has identified four major themes of this dissertation: conversation agent design, open dialog, automatic speech recognition, and domain-independent knowledge management. Each of these general topics plays an important role in the realization of the proposed dialog management system. The next chapter brings forth a survey of the background issues involved in these themes, with special emphasis on the most recent research endeavors. Chapter Three presents a formal introduction of the problem, while Chapter Four details the approach used in its solution. A prototype of the system is provided in Chapter Five, followed by its evaluation results in Chapter Six. This dissertation finishes with a conclusion of the research in Chapter Seven.

CHAPTER TWO: REVIEW OF THE RELEVANT LITERATURE

This chapter describes the current state of the art techniques associated with speech-based ECAs. The previous chapter identified four major themes: conversation agent design, open dialog, overcoming automatic speech recognition, and domain-independent knowledge management. Each of these items is affected by three fundamental technologies covered in this chapter: natural language processing (NLP), dialog systems, and context-based methods. These topics concentrate on a specific aspect of the behavior of the proposed speech-based dialog management system. Within the NLP section, an examination of various domain-specific applications shows its flexibility and versatility in practical situations, with special regard to knowledge management and conversation agent design. The dialog systems portion delivers the modern-day advances in conversation agents since the days of ELIZA. (Weizenbaum, 1966) This discussion also mentions the use of open discourse in early dialog systems. The final part regarding context-based methods gives insight into how context has played an important role in resolving the semantic ambiguity issues in NLP, with emphasis on its impact in ASR.

Natural Language Processing

NLP refers to the branch of AI in which a human agent interfaces with a machine in her/his own native tongue, whether through text-based entry or through speech input. Wilks (2005) identifies four major issues associated with NLP: linguistic systems, knowledge representation structures, information corpora, and statistical and quantitative methods. These are considered quintessential research topics in NLP, and the following sections provide an in-depth look at each.

Linguistic Systems

Linguistic systems refer to those NLP approaches that interpret user input at the grammatical level. An important piece of a linguistic system is its parser, which serves as the first line of defense in interpreting a human's intent. What makes designing the parser especially challenging is the handling of multiple sources of linguistic ambiguities. These ambiguities exist in speech recognition, syntactic processing and semantic understanding. (Baker et al, 1994)

Speech recognition issues can appear when phonetically similar words can be mistaken for the speaker's originally intended words. Lieberman et al (2005) use the excellent example of "wreck a nice beach" as a homonym for "recognize speech." Spoken to a machine, both phrases are *phonetically* identical, but they are far from being *semantically* identical. Resolution of speech recognition ambiguities often employs contextual cues to constrain the number of possible matching words for the user's utterances.

Syntactic confusion occurs when sentence parts can be interpreted in a variety of permutations. For example, the sentence "the woman ate the cake with the fork" may be interpreted as a woman eating dessert with the aid of a fork, or it could be understood as a lady devouring a particular cake decorated with a metal eating utensil. This ambiguity often causes confusion in humans, and much more bewilderment to the most naïve of linguistic systems. Wilks (2005) refers to this problem as *center-embedding*. Again, contextual recognition remains the key in maintaining conversational sanity, as combining knowledge of the current state of the environment with the current conversation can go a long way in resolving syntactic ambiguities.

Semantic ambiguity results when the meanings of the sentence parts may be understood in multiple ways. An instance of this type of confusion occurs in the following: "the pitcher put the batter in the refrigerator.” In this sentence, the semantic intent of ‘batter’ is in question, since it could mean either a baseball player or a food item. There is no real justification to choose one or the other without the help of any additional information. To correctly interpret these semantic questions, it is necessary to be equipped with contextual cues related to the current state of affairs.

These sources of conversational ambiguity make up the brunt of difficulties encountered in creating linguistics-based NLP systems. The common remedy to resolving these sources of confusion lies in maintaining a good grasp of the current situational context associated with the linguistic utterance.

Knowledge Representation Structures

Language has been viewed as a trivial issue once knowledge is established in a proper representation. (Wilks, 2005) Traditionally, this knowledge representation is expressed in logic-based syntax. Upon creating a predicate logic rule base, sentences can easily be formed by simply reading off each individual rule. Conversely, a natural language sentence could be converted into a logical expression, and thus, effortlessly added to the existing knowledge base. This latter process, however, has proven to be quite a difficult problem in NLP.

Wilks (2005) mentions three viewpoints on the relationship of language and logic statements. The first dictates that logical inferences must be derived from conversation. Instead of parsing a sentence for face value, the meaning of the utterance may have logical attachments that must be inferred from a back-end knowledge model. The second stance maintains that meaning can exist outside of logic. This assumes some sort of a priori association between words that is not established using logic alone. Earlier work on *word primitives* (Wilks and Fass, 1992) and *scripts* (Schank and Abelson, 1977) adhered to this type of natural language understanding through the explicit pre-programming of machines with pre-processed bodies of knowledge. The third viewpoint says that both logic and language suffer from the same problem of ambiguity, and that knowledge representation based on predicate logic is a practice based on completely arbitrary decisions. (Nirenburg and Wilks, 2001) Simply put, the symbols found in both logical statements and in language statements are essentially cut from the same cloth of vagueness.

AI researchers are still divided in how conceptual knowledge translates to lingual semantics. These three camps of knowledge representation for the sake of NLP exemplify the fact that fundamental questions of symbolic manipulation remain unresolved. Unsurprisingly, little progress has been made in advancing the relationship between logic and language in practical applications.

Information Corpora

Information corpora refer to those massive data sources that provide extensive knowledge of individual words or phrases for the sake of semantic resolution. Early work saw the transfer of thesaurus data into IBM punch cards (Masterman, 1957) (Spärck Jones, 1964) for the purpose of miscellaneous linguistic disambiguation tasks. The computational infancy of that era, however, prevented any major breakthroughs in NLP. Since then, vocabulary corpora have evolved to include not only word meanings, but also statistical information regarding usage and distribution. (Rayson and Garside, 1999) Moreover, thesaurus data can now be used to drive semantic analysis for real-time applications, such as chatbots (Zdravokova, 2000), or data-intensive functions, such as automatic keyphrase indexing. (Medelyan and Witten, 2006)

Ontologies

A classic type of NLP vocabulary corpora is the *ontology*. Ontologies are pre-configured dictionaries that provide such linguistic information as related synonyms and word groups, parts of speech, and multiple definitions. (Gruber, 1993) Ontologies have been used to classify documents (Cheng et al, 2003), a function usually executed by a simple set of keywords. Several ontologies may also be used in a single system. Magnini et al (2002) contributed work on a management policy between different ontologies to result in an agreement of contextualized linguistic meanings. Goh et al (2006) produced a crisis communication agent that drew knowledge from multiple information sources for different domains. Ontologies have also made strides in attempting to learn content during spoken dialog interactions. (Loos, 2006)

WordNet serves as the gold standard ontology. WordNet is an ongoing project in which researchers continuously augment its entries, while knowledge engineers utilize WordNet to aid in their NLP endeavors. (Miller et al, 1990) Some recent enhancements to WordNet include extending it to support combinations of words, or *phrasets*. (Bentivogli and Pianta, 2004). Harabagiu and Moldovan (1996) used WordNet to parse the Internet as a means for feedback-loop text processing. This use of WordNet hints at some of the work developed by the Semantic Web project, which is described later.

Cyc (Lenat, 1995) and ConceptNet (Liu and Singh, 2004) take ontologies a step further by providing a common sense database. This knowledge incorporates all of the information that humans apparently take for granted, but machines must be explicitly told. For example, the utterance “the man drove to Orlando” contains a plethora of common sense assumptions that are not immediately established by a computer. These assumptions might include such ideas as “the man drove a car” or “the man drove on the road,” and so forth. Cyc and ConceptNet provide an enormous amount of data to assist in making these additional statements. Cyc is implemented using a web of logic-based constructions. ConceptNet is built as a semantic network, similar to WordNet. The main difference between the two is that ConceptNet focuses on *concept*-based look-up, while WordNet is a *word*-based system. As also seen in Ozcan and Aslandogan’s work (2004), the practice of using a conceptually organized infrastructure has been featured in a predecessor system in KL-ONE (Forrest, 1991).

Tree Banks

Tree banks add a structural dimension to information corpora. They parse bodies of text to include both annotated and syntactic sentence information in the form of a tree structure consisting of parts-of-speech (POS) tags. Penn Tree Banks serve as the quintessential type of tree bank, whose primary data source is the Wall Street Journal.

A specialized use of tree banks involves memory-based parsing. Canisius and van den Bosch (2004) created a memory-based shallow parser for Dutch tree banks. Kitano and Higuchi (1991) presented a parallelized text parsing process rather than the traditional serial parsing method. Lee et al (1995) performed Korean/English translation using a modified memory-based technique, emphasizing the parallelized passing of markers. Zavrel and Daelemans (2003) implemented Memory-Based classification for information extraction that yielded better results than Hidden Markov Models.

Semantic Web

In past endeavors, selection of certain pieces of literature for inclusion in information capture was a deliberate process. (Wilks, 2005) Rote encyclopedia sections were used to provide content for NLP endeavors. (Kupiec, 1993) Now, this has drastically changed with the rapidly increasing growth of the World Wide Web (WWW), as researchers are turning to the Internet as a massive textual repository .(Hammer et al, 1997) (Huhns and Singh, 1997) (Huck et al, 1998) (Eliassi-Rad and Shavlik, 2003) (Neumann and Xu, 2003) (L'Abbate et al, 2005) (Gatterbauer et al, 2007) (Soderland and Mandhani, 2007) The Semantic Web project was recently conceived to utilize the entire Internet as a source of ontological data (McIlraith et al, 2004). This work provides many promising avenues for NLP, especially when encyclopedic sites such as Wikipedia become increasingly popular. (Milne et al, 2006) (Ruiz-Casado et al, 2006) (Suh et al, 2006) (Zesch et al, 2007)

Statistical and Quantitative Methods

Since the 1960's, statistical and quantitative methods were touted as possible solutions to handling NLP. Researchers also noticed that most Western languages were characteristically redundant in sentence elements, as this would compliment nicely with said methods. (Wilks, 2005) The main idea was to derive algorithms that determine a mathematical correlation or statistical relationship between massive amounts of data points made up of language elements.

This notion hints at the eventual practice of ML in NLP endeavors and natural language applications, such as chatbots (Batacharia et al, 1997), QA systems (Hermjakob et al, 2000), machine translation (Levin et al, 2003), and knowledge corpus management (Cardie, 1994) (Peñas et al, 2001) (Perez-Marin et al, 2006) (Thomas and Sheth, 2006).

ML-based NLP drastically differs from its knowledge-based counterparts. In knowledge-based NLP, pre-defined meanings and relationships of words must be established in an a priori manner before a machine can understand text. ML, on the other hand, uses a purely numerical treatment of textual relationships, such as word pairing frequencies and phrase occurrence analysis. It has penetrated the NLP realm in many forms, ranging from automated POS tagging to dialog management (Wilks, 2005) to disambiguation resolution (Roth, 1998) to word similarity categorization. (Lee, 1997) (Means et al, 2004) Recent endeavors in this statistical treatment of language have resulted in advancements in Information Extraction (IE) techniques. IE exploits pattern matching to derive data relationships from large knowledge sources, such as newspaper text. Hence, it is can be observed that ML-based NLP does not require linguistic proficiency. (Vogel, 2003) (Fleischman and Roy, 2005) Because of this lack of explicit knowledge, ML can be easily applied as a general solution for any domain. This absence of expertise, however, could prove flawed, as false relationships between words can easily be formed if the ML system's source corpora are tainted with misleading textual data.

Hybrid NLP

Despite the extensive use of ML in NLP applications, knowledge-based methods often provide more accurate solutions, albeit requiring extensive hand-tailored efforts to achieve this success.

(Barrows et al, 2000) Recent NLP research has seen the merging of ML with knowledge-based methods to provide a sort of *hybrid* solution. This appears to be the natural evolution of NLP as both techniques, each of which has achieved mild successes, combine to create a more formative tool. (Haav, 2003) (Jiang and Zhai, 2007) Some examples of hybrid NLP include domain knowledge-enhanced ML-based case-based reasoning (Brüninghaus and Ashley, 1998), ontology enhancement through automated text mining and manual expert inspection (Parekh et al, 2004), and NLP parsing using fuzzy pattern matching of extracted example trees from a tree bank (Streiter, 2001). Each of these endeavors saw performance improvements when combining the forces of both ML and knowledge-based NLP practices.

This section discussed the general NLP research challenges: linguistic systems, knowledge representation structures, information corpora, and statistical and quantitative methods. As a supplement to this theoretical treatment, the next section reviews some specific applications of NLP in different domains.

Domain-Specific NLP Applications

As with any technological endeavor, NLP techniques have been studied and applied to facilitate the execution of everyday tasks. As a result, NLP tools exist for a plethora of domains, each with different user-oriented goals in mind. The following section describes recent research in domain-specific applications for NLP.

Medicine has driven many researchers toward natural language solutions, such as those needed to parse biomedical documents (Rindflesch, 1995) (Friedman, 2000) (Xiao and Rösner, 2003) (Popowich, 2005) (Al-Mubaid, 2006) (Simperl and Schlangen, 2006) and doctor's notes

(Barrows et al, 2000), as well as aid in meditative relaxation training (Ravichandran and Karthik, 2004). NLP also has found its way into law, where it aids in patent information retrieval (Babina, 2006) as well as in the indexing of crime-scene photography (Pastra et al, 2003). The foreign language community has greatly embraced NLP technology. Many language translation technologies have been developed for Dutch (Canisius and van den Bosch, 2004), Chinese (Fung and Yee, 1998), French (Goulian et al, 2003), Korean (Lee et al, 1995) (Han et al 2002), and German (Neumann and Xu, 2003). Similarly, NLP technology has surfaced in a variety of other disciplines, such as engineering (Sawyer et al, 2002) (Pazienza et al, 2005), chemistry (Townsend et al, 2005) and genealogy (Walker, 2004). The point of this voluminous listing of technologies is to exemplify the widespread applicability of NLP in real-world domains.

NLP techniques have also been incorporated in multi-modal HCI systems. The concept of incorporating language-based tools in different input modalities is often used in ECA projects. Ahad et al (2007) developed a virtual librarian conversational agent. Their system used text-based interaction, complete with radio frequency identification (RFID) user detection, an individual user rapport database and a suggestion engine powered by Amazon web services. Wong and Starner (2001), Augello et al (2006), and Santangelo et al (2006) each developed helper agents to be used on Personal Digital Assistants (PDA). Gockley et al (2005) produced a robotic secretary that would interact with humans using a keyboard-based input system, while more substantial user systems involving air traffic control (ATC) have been developed using NLP techniques with speech recognition software. (Schaefer, 2001) (Ragnarsdottir and Hvannberg, 2006)

The common thread in each of these research projects is that they all touch upon a very small functional niche, where each NLP system has a very specialized goal. This directed specialization has allowed each system to succeed in its own particular application. In the case of the librarian, the air traffic controller, and the personal assistants, each agent had very well-defined tasks that could be accurately modeled by human programmers. This method of agent design, while functionally adequate, does not provide a generalized answer for domain-independent solutions. A domain independent infrastructure is desirable for HCI systems because it provides a reusable agent shell that can be quickly changed for building agents with different task expertise.

Dialog Systems

Dialog systems represent a specialized field of NLP. Their purpose is to provide an HCI method that resembles a conversational exchange between two humans. Additionally, *assistive* dialog systems exist to serve a particular purpose, such as providing information to its users or aiding them with completing a certain task. The channel of communication between the user and the dialog system may be based on text, speech and gestures, or on a combination of all three.

This dissertation focuses on speech-based (spoken) dialogs. Spoken dialog systems exist with a wide range of complexity, especially when dealing with the level of "openness" that the conversation domain can support. Specifically, dialog systems range in different expected response complexities, from single-word utterances to full-on natural language sentences. Often times, these differences can mean the trade-off between real-time processing ability and performance robustness. (Allen et al, 1996; 2001) Clearly, dialog system architects are pitted with many other design choices from which to select. Flycht-Eriksson and Jönsson (2000) list the four major components of typical dialog systems as the Interpreter, the Generator, the Domain

Knowledge Manager, and the Dialog Manager. The Interpreter is essentially the speech recognition unit, and the Generator is the speech synthesis, or text-to-speech (TTS) module. The remainder of this section is devoted to describing the other two components and the design decisions associated with each of them.

Domain Knowledge Manager

The Domain Knowledge Manager acts as the back-end database for a dialog system, supplying it with various facts and figures. Reasoning constructs, such as predicate logic rule bases, may also exist within this module. In essence, this combination of a rule base and a fact base resembles the infrastructure of an expert system or a decision support system (DSS), such as that featured in the SimCity dialog system. (Augello et al 2009) The Domain Knowledge Manager can be paired with a full-blown database application, such as a library card catalog (Cenek, 2001), a time card tracker (LuperFoy et al, 1998), or an interactive inventory (Thiel and Stein, 2000) (Chai et al, 2001) (Owda et al, 2007). Often times, missing data items in this knowledge base can drive the dialog system behavior. This is known as slot-filling and will be discussed in the Dialog Manager section.

Flycht-Eriksson and Jönsson (2000) cite two main design issues with these systems: data accessibility and relevant knowledge retrieval. The data accessibility problem occurs when multiple sources of domain knowledge is present. The Domain Knowledge Manager must be able to recognize how much, if any, of that knowledge is needed for that particular instance of time. This is done by a specialized component in the Domain Knowledge Manager known as the *domain task model*. Relevant knowledge retrieval becomes apparent when the dialog system is presented with insufficient or erroneous information. It is up to the Domain Knowledge Manager to decide what knowledge in the knowledge base is sufficiently relevant to coherently become included in the current conversation.

Dialog Manager

The Dialog Manager serves as the conversation control mechanism. Its primary operation is to compose a coherent response when given a user utterance. The responses supplied by the Dialog Manager come in a variety of flavors. A data-driven response may simply parrot back a piece of information residing within the Domain Knowledge Manager. The Dialog Manager may pose clarification questions if incomplete or unintelligible input has been detected. Data-driven questions may also be asked as a means to direct the conversation toward the Dialog Manager's liking. These responses may relate to some missing data that the system wishes to retrieve from the user. It is clear that the responses from the Dialog Manager are quite varied. Flycht-Eriksson and Jönsson (2000) suggest three sub-systems of the Dialog Manager that work together to determine these responses. These modules are known as the Dialog Model, the System Task Model, and the Dialog History. The Dialog History simply keeps accounts of all human-computer interactions for future reference. The System Task Model defines how the *system* itself executes tasks that assist the user in accomplishing her/his own tasks. This model differs from the *user* task model in that the system is only interested in providing a verbal, information-based action, while the user is looking to achieve a real-world accomplishment. For example, if the user task model involves the human participant trying to take a bus to get to the airport for a 3 p.m. flight, the system task model would provide a schedule for all buses before that flight departure time (Flycht-Eriksson and Jönsson, 2000).

The remaining Dialog Manager component, the Dialog Model, describes the general structure of how a dialog is to be executed. It gives a plan for what response should be outputted given a particular input, and it also dictates what major actions should take place at each conversational crossroad. Flycht-Eriksson and Jönsson (2000) suggest that intention-based models and structurally based models represent the two main styles of dialog modeling. The structural type does not require an explicit definition of the user's goals. These dialog models resemble a very linear, scripted conversation. Bui (2006) describes this as *finite state-based* model. Each of the agent's dialog moves is pre-determined and the system itself controls the conversational initiatives.

The intention-based dialog model, on the other hand, recognizes the user's desired end results, and plans its conversational responses accordingly. Plan operators and domain knowledge are employed to accomplish the necessary goals. Bui (2006) suggests that *frame-based* flow be used for dialogs driven by user intention. The Information State Update (ISU) (Lemon, 2004) dialog method is based on this principle, where grammar switching is motivated by data flow. Frame-based, or *slot-filling*, dialogs drive conversation with the goal of completing some sort of data class that has missing data points. (McTear, 2002) These models can support mixed-initiatives, where either the user or the system is capable of controlling the flow of the conversation. Figure 2 portrays the block diagram for general dialog system architecture as proposed by Flycht-Eriksson and Jönsson (2000).

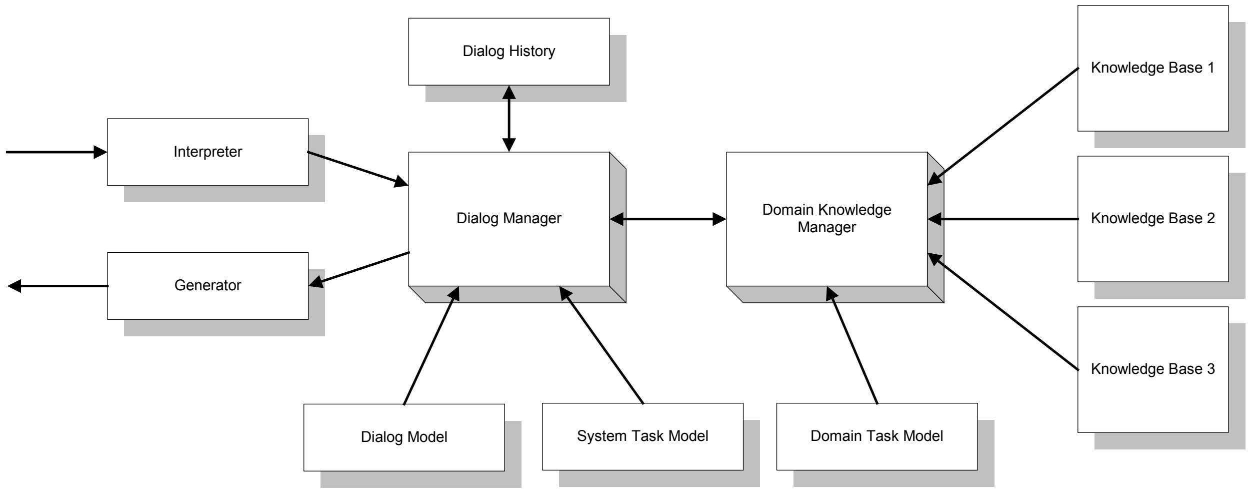


Figure 2: Dialog system architecture block diagram (Flycht-Eriksson and Jönsson, 2000).

The preceding section divulges the inner-workings of a dialog system. Essentially, these systems serve as behavior models of speech actions for an autonomous agent. The next section describes some specific examples of dialog systems built by other researchers.

Dialog System Applications

As seen in the previous section describing specific NLP applications, a vast number of domains have benefited from NLP technology. Classified as a specialized usage of NLP, dialog systems have propagated this spirit of practicality. Dialog systems can use spoken or non-spoken input, the latter of which usually involve text-based user entry. The following section describes a series of innovations in domain-specific dialog systems.

Andernach et al (1995) provide a text-based dialog manager, SCHISMA, for theater information and ticket booking. SCHISMA consists of two major components: an error handler for user input and a semantic parser. The error handler, or morphological analysis and failure (MAF) portion, pre-processes the user's keyboard-based query entry. The MAF essentially inspects the input for typing errors and conceptual mistakes before entering the parsing component. The parser, PARS, breaks down the user query into a set of word groupings that are treated as feature structures. Domain-specific grammar rules are applied to these groupings to aid in semantic resolution. Performance results were not provided by the authors, as the SCHISMA project was not in full practical use. Andernach et al's (1995) work could be improved by incorporating the error analyzer for speech-based input, rather than solely typing-based interfaces. This suggestion would be useful for more recent endeavors in ASR-based agent research.

Ball (1999) examines the use of scripted patterns in a speech-based open dialog interface. He uses a Microsoft Agent-based time management helper application to test these findings. In this system, both domain-specific and open techniques are utilized. A system of patterns, using the SpeakEasy pattern system, is heavily used for determining the user's semantic intent. Patterns consist of three components: a primary representative input response, a list of ordered word patterns, and a list of key words or phrases. The user's utterances to the agent are matched against the different SpeakEasy pattern entries. Additionally, using task templates under the NLPWIN system incorporates a deeper level of pattern matching. This system combines a robust English language parser with the MindNet semantic knowledge base to determine semantic equivalence between different input sentences. The idea of sentence similarity matching can later be seen in O'Shea et al's conversation agent work. (2008) Ball's research provides some simple techniques for providing a simple dialog system using readily available software packages. His work, however, does not address how the agent can determine the user's intended task goals after determining her/his semantic intent.

Sonntag (2006) presents the SmartWeb system, a text-based dialog system that uses a hybridized contextualization of information. This research deals with the case in which multiple user goals co-exist, especially when many sources of information are present in the database. The key to SmartWeb's functionality lies in its ML-based feature extraction of data. This process allows the system to identify the different contexts in which it may work. During the information retrieval procedure, SmartWeb matches the user query with a best-fitting context. Sonntag's work gives insight into a data-fetching system that can operate over a plethora of domains. The main contribution from Sonntag's SmartWeb lies in its ability to identify a single unique user information focus, or goal. Extending this work would entail enabling support of multiple user goals.

Retrieving travel data has been a popular application for spoken dialog systems. The initial prototype stages of Larsson and Traum's (2000; 2002) TRINDI dialog system and the DARPA Communicator (Levin et al, 2000) served as speech-based travel booking software. Levin et al (2000) provided some preliminary performance data in both quantitative and qualitative forms, noting an ASR word accuracy of 72.73% when system goals were completed, and 63.25% for incomplete tasks. They noted the system was more often used as a menu-driven entity, rather than the conversational agent that it was designed to be. Railway information could be conversationally fetched using LIMSI Arise (Lamel et al, 1998) and Litman et al's AMTRAK system (1998). LIMSI Arise provided quantitative results, yielding an 11% error rate when making reservations and a 16% failure rate for deploying information. Niesler and Roux (2001) used spoken dialog technology to drive a hotel room reservation process. Misu and Kawahara's Virtual Kyoto (2007) gives local visitor information using spoken queries, which operates at a failure rate of 50% when using an open environment. A 61.4% success rate was achieved when the system functioned with contextual relevance information. Navigational systems from Skantze et al (2006), Belvin et al (2005), Baca et al (2003) and Takeda et al (2005) have used speech-based input to provide real-time location data. Baca et al's automotive system (2003) saw a decrease in error rate from 49% to 3% when semantic frames were used to assist in ASR interpretation. Hardy et al (2004) present their model for an automated travel information call center, observing a 70% overall success rate with a WER of approximately 20%. Agarwal (1997), Huang et al (1999) and Johnsen et al (2000) propose PURE, LODESTAR and TABOR, respectively, as spoken dialog systems that also retrieve travel information. TABOR (Johnsen et al, 2000) saw 10.7% of its responses to be illegal turns, while 13.3% were turn errors. LODESTAR (Huang et al, 1999) managed to achieve a system correct response rate of 90.9%.

In each of these travel data systems, a slot-filling technique is used to drive the agent's actions. Additionally, the performance data for some of these systems yield very successful results. This success can be attributed to the fact these travel systems deal with a very narrow domain. Since travel information is straightforward, a simple linear flow of data is requested by each of the automated systems. The obvious goal of attaining enough information to determine the user's travel plans has already been determined before the conversation is initiated. These travel information agents exemplify information retrieval driven by a specific goal. Nevertheless, they do not need to automatically determine the needs of the user. This requirement is usually found in dialog systems in which the there may be multiple sources of contextualized data for each user.

Bohus and Rudnicky (2003) present the RavenClaw spoken dialog manager, as featured in their aircraft maintenance system, LARRI (2002). RavenClaw specializes in operation of domain-specific applications. It separates the domain-dependent functionality from the generic dialog components. In this setup, the domain may be interchanged for a variety of tasks while the domain-independent dialog operations may exist without alteration. RavenClaw was tested on five different domain applications, each of which did not require major re-structuring of any core modules. These applications, however, reflected a linear, slot-filling execution, where the number of expected slots was pre-determined. An improvement to RavenClaw would be the incorporation of domains that include dynamic expectation agendas.

Lemon (2004) uses the ISU method to control speech-based dialog with a grammar switching approach. A development framework, DUDE, facilitates the authoring of ISU-based interactions. (Lemon and Liu, 2006) Lemon uses context-based method in that only certain linguistic groups are pertinent for specific situations. Hence, these groups, or subgrammars, serve as the different contexts. Lemon proclaims that the use of context sensitive techniques improves a speech recognition system's recognition error rate and concept error rate. A 68.9% concept recognition rate was achieved using contextual assistance. The basis of his research deals with the idea of storing data as dialog-specific structures rather than domain-specific elements. Hence, Lemon chooses to focus on conversational constructs, rather than a linear database, to control the flow of information within a dialog. His work could be improved by showing how his ISU method can be used for a significantly large number of contexts, as to exhibit its strength as a more generalized solution.

Schaefer (2001) presents the Cognitive Controller Model (CMM), a speech-driven behavior model that predicts a user's verbal output in ATC situations. The CMM relies heavily on context-sensitive methods, where the system must dynamically determine the relevant syntax of the current state of a conversation. The difficulty of establishing the situational context in ATC lies in the unpredictable nature and inherent complexity of ATC. Schaefer implements a human problem solving cognition prediction model into his CMM. Hence, the CMM uses a behavior model as the underlying decision maker for the system's linguistic response. This model uses a standard phraseology set and achieved a recognition rate of 93%. The system became slightly more error-prone upon extending the size of the phraseology. Schaefer notes that CMM can only process one instruction at a time, whereas human operators often deliver multiple instructions in a single exchange. The CMM presents a novel approach to speech recognition with the use of a cognitive model to provide expectations on the users' behavior. This method works well for ATC because air traffic conversations are simple enough to follow a well-defined structure and dictionary. Thus, Schaefer's work is best used for highly-constrained conversations, as more complex conversations would require a much longer CMM development time.

Jokinen et al (2001; 2003), Pallotta (2004), and Pallotta and Ballim (2001) propose the components needed to devise a dynamic dialog manager, regardless of spoken or non-spoken user inputs. Jokinen et al suggest that communication flexibility using ML techniques will enhance the dialog exchange experience between human and computer. This flexibility would allow the machine to exhibit a sense of diversity between different users. Additionally, they propose that dynamic knowledge bases can be extracted from these dialog interactions, rather than employing the traditional slot-filling information acquisition techniques. They specify that clustering and classification methods should be implemented to accomplish such knowledge extraction. Pallotta proposes a cognition-based dialog manager. In this work, he identifies the importance of using context when dealing with partially complete information. Hence, he claims that contextual cues should be used to help dialog systems infer data points. This is a concept that Pallotta refers to as *computational cognition*. Both Jokinen et al and Pallotta shed some light on utilizing context-driven methods for dialog management that is explored and expanded upon in this dissertation.

Question-Answering Systems

Question-Answering (QA) is an active sub-topic of dialog systems. Much of today's web search engine algorithms operate using the principles derived from this body of research. The basic structure of a QA system consists of three parts: question analysis, information retrieval (IR) and answer deployment. (Kwok et al, 2001) Various researchers have added their own utilities to each of these components to enhance QA performance. The following section describes some key examples of such endeavors.

Brill et al (2002) analyze the AskMSR QA performance for accuracy and predictability of being incorrect. In this system, a question is first rephrased into a set of declarations to facilitate the matching process in a search engine. The idea of query expansion has been proposed in other QA endeavors. (L'Abbate and Thiel, 2002) (Varges et al, 2009) The declarative sentences result in a list of possible query answers in the form of word sequences of size n, or n-grams, derived from page summaries. This list is subjected to a set of 15 manually written filters, starting with a question-type filter. The QA process concludes with n-gram tiling, where a collection of related short answers are combined to form a single larger answer. AskMSR depends on the integrity of its data, where higher redundancy yields stronger confidence. This emphasis on the actual data, rather than linguistic methods, allows for a QA system that may opt for a non-answer in place of a wrong answer. AskMSR presents an effective model for answering user queries, especially in instances of unstructured speech. The dependence of manually configured filtering rules, however, suggests that the system may not be robust enough for larger scale projects.

Cao et al (2005) create and compare two QA techniques based on NLP parsing and pattern matching, respectively. Their work deals with retrieving answers from information sourced in video format, a task that will later be addressed using Karanastasi et al's OntoNL (2007) In Cao et al's work, the video content is first transcribed to written form via speech recognition methods, complete with phonetics-based and domain knowledge error correction. In performing the QA interaction, Cao et al's NLP approach converts the question to a declarative sentence for matching within the existing sentences in the knowledge base. Their pattern matching technique, on the other hand, identifies in each question the question *type* (i. e., who is, where is, what is) and the remaining questions *parts* for matching within the video transcripts, a method similar to Ravichandran and Hovy's (2002) surface text patterning. It was determined that the NLP method worked best in domain-specific corpora, while the pattern matching was most effective in the presence of large volumes of information. Cao et al present two approaches to their information retrieval issue, both of which brought forth different advantages for different situations. Their work would be improved by providing a hybrid solution that mixes methods derived from both the NLP and pattern matching techniques.

Cimiano et al (2007) create the ORAKEL system, which can be used by a domain expert with no NLP knowledge. Logical Description Grammars (LDG) are heavily used to store question data that are originally in natural language format. ORAKEL uses LDG's to break down a factoid question into a tree, structures that have been featured in previous works for QA purposes. (Punyakanok et al, 2004) Three source lexicons are used to interpret various parts of the tree: a domain independent lexicon, a domain-specific lexicon and an ontological lexicon.

The domain-specific lexicon is customizable in such a manner that a domain expert that does not have NLP training understanding may easily add her/his knowledge to the existing data. The intent of ORAKEL is to provide domain-specific support for an NLP system, without the need to be domain dependent. Cimiano et al's work would benefit by identifying a set of results that demonstrate which domains ORAKEL could operate with little or no change to its core NLP infrastructure.

Kwok et al (2001) presents the MULDER system, which performs QA using information found on the Web. An NLP-based approach of mixing the Maximum Entropy-Inspired (MEI) parser with the PC-Kimmo lexical analysis tool was used to parse questions. Following the parsing phase, the type of question was determined, which in turn, sparked a parallelized Google search containing re-formulations of the original question. These queries vary in specificity levels, which can be controlled by manipulating verbs, expanding query concepts, maintaining noun phrases, and using grammatically equivalent speech. Upon retrieving search results and extracting web page summaries, MULDER scores each summary with respect to closeness of significant keywords. Clustering is performed on the best answer candidates and the champion of the winning cluster is selected. Kwok et al showed that MULDER was able to perform better than Google and AskJeeves in web-based QA. The effectiveness of MULDER relies heavily upon its access to multiple processing units to execute its parallelized procedures. This necessity for processing horsepower is the primary limitation of MULDER. Nevertheless, this hardware requirement may be viewed as a handy alternative for producing a more efficient, less processor-intensive software solution.

Mollá et al (2003) present ExtrAns, which serves as a *technical* QA system, a realm where concrete responses are expected. The authors assert that such a system must be based on

NLP techniques, rather than ML methods, because of the heavy use of technical jargon and domain-specific terms. ExtrAns functions with four major components: a terminology database, an NLP-intensive parsing and interpretation system, a logical annotator to incorporate minimal semantic information, and a user-friendly display. Mollá et al showed that ExtrAns outperforms the legacy IR system, SMART, 0.63 versus 0.46, using the Mean Reciprocal Rank (MRR) metric. The MRR simply measures how accurately a system can make a known answer the number one retrieval ranking. Although Mollá et al make a good argument for not using ML in their QA method, augmenting ExtrAns with such techniques would possibly enhance its performance.

Nyberg et al (2005) extends JAVELIN, a QA system, to use domain-specific knowledge. The use of local corpora for QA was also featured in Lita and Carbonell's (2002) work, and later in Hickl et al's FERRET (2006). The particular domain that JAVELIN deals with is weapons of mass destruction (WMDs). JAVELIN consists of six components: question analyzer, retrieval strategist, information extractor, answer generator, execution manager and planner. The question analyzer profiles the question for type and answer type. The retrieval strategist performs a string of queries to solicit related information. The information extractor retrieves knowledge excerpts to supply ensuing answers. The answer generator prioritizes the potential machine answers. The execution manager sets up these components for real-time execution. The planner polices the execution of each of the aforementioned subsystems. In enhancing JAVELIN to support domain-specific functionality, Nyberg et al added features derived from the Identifinder and ASSERT text annotation systems, the WordNet ontology and a WMD ontology. They provided support for using both predicates *and* keywords in the question analyzer, retrieval strategist, information extractor, answer generator, and execution manager, where the original system only utilized keywords. JAVELIN was also fitted with type hierarchies to resolve relationships between objects, as well as the Also Known As (AKA) Extraction tool that handles synonyms. An improvement to Nyberg et al's work would be to provide performance results for additional domains outside of WMD's to exhibit JAVELIN's strength as a domain-independent QA system.

Surdeanu et al (2006) perform QA for speech transcriptions, based on NLP components and information retrieval methods. The proposed architecture implements a traditional NLP QA structure, complete with a Question Processor (QP), Passage Retriever (PR) and Answer Extractor (AE). The advantage of Surdeanu et al's system remains in the marriage of an information retrieval-based algorithm (where keywords and surrounding word contexts are heavily utilized) with two NLP tools: a POS tagger and a name entity recognizer and classifier (NERC). This work would benefit by providing results for a set of various domains, emphasizing the system's versatility for different contexts.

Chen et al (2006) propose a method of QA that uses a language model to assist in prioritizing an answer candidate list. This architecture consists of three steps: 1) learning the language model using the Web, 2) using the language model to perform the answer prioritization, and 3) eliminating duplicate candidates. The key mechanism to the language model is the Web-based discovery of words that most likely occur with the original search items. Once these helper terms are established, a two-word sequence, or bigram, and a two-term, or biterm, model were implemented to reflect these relationships and to create a ranking system for answer candidates. Chen et al determined that this language model method performed better than the unigram model and the Vector Space Model (VSM) techniques. This research could see improvement if the search engine content that was queried was more closely monitored, as to prevent the establishing of any false term relationships that may occur as a result of errant Web data.

Recent developments have been made to integrate speech-based interfaces to QA systems. Dialog systems (Denecke and Yasuda, 2005) and NLP tools (Litkowski, 1999) (Bernardi et al, 2003) made just for answering questions are graying the line between assistive conversation agents and QA systems. Schofield and Zheng (2003) transformed the AnswerBus QA tool into an open, spoken dialog system. SPIQA (Hori et al, 2003), Ritel (Galibert et al, 2003) (Rosset et al, 2006) and González-Ferreras et al (2008) also provide a similar functionality, offering three other examples of speech-enabled question answerers. The common issues that plague each of these projects are similar to those of any spoken interface system. In comparing a speech-based QA-based conversation agent with that of an ECA, it must be noted that the QA discourse model exhibits much more simplified behavior than that of a full-blown ECA. Sergeant Star (Artstein et al, 2009), while existing as a full-body interactive avatar, could still be considered just another spoken dialog QA system, mainly because of its simplistic answer-response pair style of discourse. While the QA agent only needs to provide rote answers to its user, an ECA's conversational manner must incorporate a more personalized HCI experience. The next section describes this next class of agents that reflect such a personal nature.

Conversation Agents

Conversation agents refer to those software programs designed to interact with a human user in natural language, using typed words or spoken utterances. These programs receive the user's input, interpret her/his request, and attempt to produce a coherent response. Weizenbaum's psychotherapist-based ELIZA (1966) represents the pioneering chatting agent software. The following section describes in detail a selection of the more recent examples of modern conversation agents.

Galvão et al (2004) present Persona-AIML, which gives personality to agents based on the AIML (Wallace, 2002). This text-based system consists of four parts: the Categories Base, the Personality Component, the Dialogue Log, and the Reasoning Component. The Categories Base maintains the vanilla AIML interaction patterns. The Personality Component is a repository of the personal rules that defines the chatting agent's values. The Dialogue Log matches conversational contexts with a user, and the Reasoning Component serves as the response engine, integrating the contributions of the other parts of the system. Galvão, et al's main accomplishment with Persona-AIML was to create an AIML-based architecture that can take on different personality traits. However, they did not explicitly establish how adding unique personality can quantitatively enhance the chatting experience with a human user.

Stede and Schlangen (2004) explore the design of conversation agents when multiple relevant topics are present. Specifically, they suggest the use of an *information-seeking* approach to drive the agent's responses. This paradigm prefers the entire domain topic over narrowly defined tasks as the motivator for conversational behavior. These behaviors are delineated by Stede and Schlangen as a set of 15 well-defined speech acts, which include: ask-more-general, ask-more-specific, ask-more-attribute, reply-pos, reply-neg, tell-topic-general, tell-topic-spec-attribute, rule-out-topic, switch-topic, noncommittal, digression, bye, opening/closing, help and garbage. Stede and Schlangen also consider the notion of separating the domain-specific knowledge from the dialog control mechanisms, allowing for a more flexible infrastructure for a variety of different domains. They admit that their work would be enhanced by including a more extensive results section that displays the efficacy of their chatting agent.

Sansonnet et al (2006) also proposed a conversation agent architecture that emphasizes the requirement for domain-independent operation to maintain genericity. In particular, this work deals with *assisting agents*, fully interactive interfaces to help novices deal with a specific task. Assisting agents are composed of three main parts: the dialogical agent, the rational agent and the embodied agent. The dialogical agent deals with understanding the user's inquiries, while the rational agent handles the actual answering of the question. The embodied agent is the actual physical presence of the assisting agent. The basis of Sansonnet et al's work exists only as a theoretical infrastructure. They could reinforce their findings by implementing and evaluating their architecture in an actual application-based agent in a real world setting.

Field and Grasso (2004) developed a conversation agent called Personalised Information Platform for Health and Life Services (PIPS) that pays special attention to persuasive language, also known as argument-based dialog. A starting basis for PIPS exists in a data structure known as the *rhetorical schema*. The six arguments for such a container includes: N (the name of the schema), C (the supporting claim), O_c (the rule-based constraints), A_c (acceptability constraints), R_c (relevance constraints), and S_c (sufficiency constraints). This six-tuple has been used in formalized analysis of arguments, yet it has not been successful at creating arguments. In response to this weakness, PIPS incorporates a two-stage planner whose states include present and the future expectations. The goal state is achieved in a manner similar in operation to theorem prover systems (Field and Ramsay, 2007), which heavily rely on a logical reasoning process. PIPS presents an ideal view of how dialog management should be conducted, but it has yet to see experimental results, as it is still in a state of infancy. A later effort by Schulman and Bickmore (2009) also embraces the use of persuasive techniques, with similarly inconclusive findings.

Hoshino et al (2005) created a speech-based chatbot architecture that incorporates current news topics into conversation. This method was tested on the ASIMO humanoid robot platform. The idea of this endeavor was to enhance the conversational experience in a jovial manner while keeping the user engaged in relevant dialog. By including this conversation control into the ASIMO platform, Hoshino et al were able to give a more human-like vehicle of interaction. While this research proposes a well-equipped form of multi-modal news information distribution, it still uses an antiquated conversation agent model. Similarly, Nakano et al (2005) also used ASIMO as a conversational robot platform. In their work, the focus was not on conversational engagement, but rather, the ability to recognize and service a user's intentions. Geib and Steedman (2007) put together a similar system, where NLP techniques were used for plan recognition. Nakano et al present the Multi-Expert-based Behavior and Dialog Planning (MEBDP) as their agent behavior model. MEBDP consists of two layers: the upper layer being a general task planner and the lower layer is a network of expert sub-mechanisms. The overall idea behind the MEBDP is to provide a finite state machine (FSM) whose main operations exist within the upper layer, calling upon the lower layer experts to take care of the fine-grained functionalities. The inputs to the FSM are dictated by ISU (information state update), essentially boiling the HCI experience between ASIMO and its human user into a slot-filling exercise. While this discourse method is not groundbreaking, what Nakano et al have contributed is a robot-based assistive platform that can provide both physical service as well as spoken information.

Huang et al (2007) propose a method of knowledge accumulation based on extracting information from online forums. The idea behind this research is to find suitable question-answer pairs to serve as a response basis for a chatbot. Similar methods have been done for other corpuses, such as the British National Corpus (BNC) (Shawar and Atwell, 2005), or even a data bank of past conversations (Voth, 2006). In Huang et al's work, forum thread titles act as the questions, and the ensuing replies are the answers. They use Support Vector Machines (SVM) to rank the various forum entries for a certain thread inquiry. The top replies are retained for use into the conversation agent's knowledge base. This system may be improved by logically grouping the extracted response knowledge into a set of categorized domains.

Jia (2003) examines the use of automated conversation agents as language tutors. Tutoring chatbots have been proposed for other realms, such as shipboard damage training (Pon-Barry et al, 2004) and tactical questioning (Gandhe et al, 2009). Jia's work adapts the ALICEBOT to serve as a foreign language chatting partner for English learners. It was concluded that this application of chatbots did not provide very productive results for its human users. The pattern-matching methods used by the conversation agent did not provide a sufficiently immersive foreign language experience for its users. Jia (2004; 2009) responds to this deficiency by presenting the Computer Simulator in Educational Communication (CSIEC). CSIEC consists of six parts: a parser, knowledge representation, knowledge base, a common sense engine, a personality module, and a response manager. The advantage of CSIEC over its pattern-matching predecessors is that it actually handles more complicated grammar situations, rather than selecting a reply from a set of programmed responses. By improving the quality of the conversation grammar, Jia asserts that a better foreign language chatting partner can be accomplished. Nevertheless, a set of quantitative results should have been provided to back this claim.

In a related research effort, Lu et al (2006) use an AIML-based agent, TutorBot, to serve as an instant messaging chatting tutor for English learners. This system contains three sub-systems: RRMBot, ClassifyBot and AIMLBot. RRMBot provides a repository of reference material collected from published English-learning printed resources. ClassifyBot uses the OpenNLP toolkit to provide a grammar-based interactivity utility for the benefit of the user. The AIMLBot component simply provides the chatting platform for the English student. The combination of these three parts comprises the TutorBot user interface environment. Lu et al concede that TutorBot's strength lies in the instantiation of OpenNLP in the AIML architecture, yet they do not provide results on how this contribution is measurably beneficial.

L'Abbate and Thiel (2002) inspect the use of query expansions to help conversation agents retrieve information in a more effective manner. They use five search strategies for this endeavor: empirical, analytical, browsing, analogy, and bibliographical. A dossier of the agent and user interactions is maintained in the Structured Dialogue History, which keeps track of all utterances and the relevant contexts of these instances between the two parties. These historical data are then fed to the query engine to provide a user-centered expansion term list. Such a system allows for a sort of memorized rapport between the chatting agent and its human users. L'Abbate and Thiel could improve their findings by providing data that exemplify the advantage of having different profiles for different users.

Vrajitoru and Ratkiewicz (2004) provided an evolutionary algorithm (EA) that forms new sentences from previously encountered dialogs. This allows for a more diverse array of agent responses. The EA's solution, however, did not guarantee a fully coherent result. Montero and Araki (2006) present a similar technique, in which a genetic algorithm (GA) is used to produce responses, known as *trivial dialog phrases*, to user inputs to a chatbot. Each phrase is considered a gene, and within each gene reside chromosomes consisting of words. The initial population of phrases was generated from a pre-determined selection, known as the Phrase Database.

Mutations occurred for each gene where approximately only one chromosome was altered per phrase, and crossover was not used for the reproduction operations. Candidate phrases were evaluated on the frequency of its associated n-grams within the WWW. Frequently used phrases were considered *completely natural*, while *usable* phrases denoted those that were somewhat commonplace. Phrases that simply were not found in conversational vernacular were deemed *unnatural*. Montero and Araki's work did have its weaknesses, as they note that the critical part of their GA's success is the operation of the mutation mechanism. Their concern lies in the issue that the newly mutated word may stray from the conversational relevance. Additionally, they understand that the expansive nature of the WWW directly affects the fitness function, and it may disturb the consistency of their system's performance.

Montero and Araki (2007) utilize the *data crystallization* method to recognize a shift of domain relevance in chatting agent conversations. Data crystallization is a ML technique that creates bridges, called *dummy items*, to connect clusters of closely related data points. At the conversation level, the dummy items translate to transitional utterances that shift the topical focus in a seamless manner. Montero and Araki assert that an intelligent effect can be produced from these domain-shifting responses. Experiments were executed to measure the performance of their data crystallization-based conversation agent. They used a qualitative evaluation of their conversation agent, asking their users to rank the effectiveness of the dialog flow. Based on seven users, Montero and Araki were able to reduce the percentage of vague replies judgments from 21.11% to nearly 7% when their chatbot was augmented with data crystallization. A suggestion for future experiments would be to provide more quantitative evidence of the system's competency, as well increase the number of test users.

Quarteroni and Manandhar (2007) investigate how effective their chatbot, YourQA, can perform QA tasks. YourQA fetches the top 20 Google search hits for text analysis and re-ranks those pages accordingly. An interactive dialog exchange is executed to further refine the human user's information search needs. Evaluation experiments were conducted to measure the swiftness, realism, and general effectiveness of the system. Quarteroni and Manandhar received mixed notions of success from their test group. One weakness that was cited was the long document retrieval time. Nevertheless, the users hinted at an affinity for a chat-based information search system.

Schumaker et al's (2007) Terrorism Activity Resource Application (TARA) Project investigated the application of chatbots for edifying the general public of terrorism-related information. TARA is based on the ALICE software, which uses a pattern-matching technique for answering questions. Both TARA and ALICE use the AIML architecture, yet TARA was fitted to specialize on terrorism topics. The goal of TARA was to provide a terrorism information source that was both user-friendly and conveniently automated. Three styles of TARA were developed: Dialog, Domain and Both. Dialog TARA remained free of domain expertise, where the user is expected to speak with the freedom of open dialog. Domain TARA dealt solely on terrorism-related contexts, and Both TARA combined the lexical databases of Domain TARA and Dialog TARA. Schumaker et al determined that users preferred the Both TARA style of conversation because of the mix of user-centered capability of Dialog TARA and the effective transfer of terrorism information from Domain TARA. It was also noted that the lone efforts of Domain TARA and Dialog TARA proved ineffective.

Shah and Henry (2005) show how the *Confederate Effect* comes into play in conversation agent design. The Confederate Effect refers to the case where a human user (known as a

Confederate) is mistaken for a machine-based chatting agent by another human user (the Judge). This is often caused by an expert-level grasp of subject matter on the part of the Confederate that can be confused for a machine's pre-programmed knowledge base. Shah and Henry point out that the Confederate Effect played an enormous role in the 2003 Loebner Contest, where the Jabberwocky conversation agent achieved victory because of the expertise of the Confederate participants. The lesson to be learned from this work lies in the idea that an all too thorough display of domain expertise detracts from the effectiveness of both human and machine conversation.

Wallis (2005) advocates the use of *intention maps* to produce more believable conversation agents. In his work, he asserts that conversations with chatbots must exhibit some sort of goal-driven behavior. This conversational goal can be motivated by either party, with the majority of the burden resting on the human user. Intention maps exist to determine whether or not the conversation is goal-oriented. They designate four states of a conversation based on the existence of goals: 1) neither party has a goal, 2) both parties have a goal, 3) only the human has a goal, and 4) only the chatting agent has a goal. In the first instance, the chatting agent must conjure up a topic of discussion. For the state in which both parties have a goal in mind, the conversation agent will concede control to the human user. In the final two cases, the conversations are forced into the single existent goal, unless it is deemed unacceptable. Wallis' work could be better re-affirmed if it is implemented in a live, practical demonstration.

Case-based reasoning (CBR) describes the process of problem solving using a repository, or *case library*, of relevant historical data (Bain, 1986). In CBR, a set of situational variables is described, and the system finds the best matched set of circumstances in its library to determine the proper solution for that case. Aha et al (2005) discuss the concept of *conversational* case-based reasoning (CCBR). For this style of CBR, each individual situational variable is recognized using a question-and-answer session. The CCBR system queries the human user for information until it collects or infers enough data to match on an existing case library instance. Branting et al (2004) pair CCBR with a dialog manager. The key feature to their work lies in a user goal stack, known as the Discourse Goal Stack Model (DGSM). During the question-and-answer exchange, as the CCBR agent directs the conversation toward resolving a case library match, some sub-tasks may occur that may delay this case resolution. The DGSM steps in to manage the execution of these processes, servicing any missing data requests from the agent or any informational clarifications from the user that may occur along the way. Aha et al (2005) points out some outstanding issues in CCBR. These include proper conversation termination recognition, presentation of reasoning justification to the user, and use of background information for conversational motivation. From this work in CCBR, it is observed that a question-and-answer conversation can be driven by a data collection process, as necessitated by CBR. Furthermore, CCBR agents can be adapted to support a non-linear dialog path. (Branting et al, 2004) An enhancement to this work would be to incorporate this style of dialog management into a fully-featured conversation agent and evaluate it quantitatively.

Each of the conversation agents presented above share some common shortcomings. A large number of these projects are not only evaluated by quantitative metrics, but also by qualitative human judgments. This arbitrary style of software validation is often deemed unacceptable, especially in commercial engineering practices. Another problem that plagues conversation agent software is the narrow domain applicability of each individual system. A single, generic chatting architecture has not yet been unanimously established in the NLP community. For those researchers that have proposed such architectures, they often do not provide a domain-specific application in which their system can effectively perform. These current chatbot issues are a strong consideration for the sake of this dissertation.

The collection of conversation agents presented in this section each contributed a stepping-stone toward more effective chatbot technology. Galvão et al (2004) provided a method of customizing an agent's demeanor by devising a modularized personality scheme. Both Stede and Schlangen (2004) and Sansonnet et al (2006) extended this modularization idea toward domain expertise by separating the knowledge base from the discourse mechanism. While each of these research efforts did not provide a compelling set of experimental results, they did lead the way toward more modularized dialog system designs. Varying examples of discourse models were also provided, ranging from FSM-based (Geib and Steedman, 2007), persuasive language (Field and Grasso, 2004), CCBR (Aha et al, 2005) to pattern-matching (Jia, 2003) (Lu et al, 2006) (Schumaker et al, 2007). In these cases, user initiative was the driving force in the conversation flow. Wallis' (2005) work with intention maps brought forth an agent that could express its needs in the dialog, reflecting the idea of mixed-initiative. This section also mentioned the QA work of Quarteroni and Manandhar (2007). Findings from this research advocated the use of chat-based entities for assistive information deployment. In short, the various conversation agents from the recent past have provided insights on dialog system design that embrace modularity, mixed-initiative, and assistive-natured chatting. These elements were to be expanded upon in this dissertation.

Embodied Conversation Agents

A specialized conversation agent whose physical presence is incorporated into the HCI is known as an *embodied* conversation agent. The notion of these interactive agents has existed since the inception of the computing age. Idealistic visions of these agents are rife with extraordinary capabilities, yet state-of-the-art technology yields only a sliver of these expectations. Thórisson's Galdalf (1999) introduced the notion of the interactive character, using both spoken and gesture-based inputs to animate a chatbot. The 2000's presented an evolution of these embodied agents, beginning with the work of Cassell et al (2000), whose conversational playmate, Sam, gave insight into the effectiveness of an HCI experience in a physically immersive environment. The key thing to note about Sam was that it was actually controlled by a human behind the scenes. This handler would react to the child's cues using a script. Although Sam was not an autonomous entity, Cassell et al exhibited the idea that even a child could feel comfortable when interacting with a machine-based being. Tarau and Figa (2004) would eventually extend this idea and create a virtual storytelling ECA. Bickmore and Picard (2004) presented their studies using Laura, a personal trainer agent. As an early prototype of dialog-based agents, Laura's interactions with the user could be considered a one-sided question and answer session, with the agent controlling the 'conversation.' The primary result of Bickmore and Picard's work was the concept that a *caring* embodied agent proved more effective than one of indifference. This work would eventually make its way into successor ECAs rooted in the healthcare sector, such as Turunen et al's fitness companion (2008), Virtual Nurse (Bickmore et al, 2009), CARDIAC (Ferguson et al, 2009) (Galescu et al, 2009), and Justina the Virtual Patient (Kenny et al, 2009).

The latter half of the decade sought more ambitious goals in creating interactive agents. Alm et al (2005) incorporated emotion into their agent, using ML-based methods to provide this affective dimension. Lee et al (2005) experimented with using robots as conversational agents, where an animatronic penguin, Mel, posed as a spokesman for a hypothetical product. This work supported the notion that humans could indeed effectively interact with a physically engaging and conversationally interactive machine. Lee et al's idea was further demonstrated in Kenny et al's work with the Sergeant Blackwell ECA (2007), as well as his successor, Gandhe et al's Hassan (2009). With more sophisticated dialog systems than Mel, Sergeant Blackwell and Hassan's conversational capabilities provide the user with a more natural human-computer interaction. It is noted, however, that Blackwell, and Hassan lack a cognitive model within its dialog discourse model. Specifically, Sergeant Blackwell's response generation involved a mechanical QA pairing matching system (Robinson et al, 2008), a method that lacks cognitive empowerment.

Cassell et al (1999) formulate generalized guidelines for developing ECAs. They declare three requirements involved in building these entities: 1) the capability to perform in real-time, 2) an infrastructure for both input processing and output response decision-making, and 3) an experience resembling human-to-human interaction. Bickmore and Cassell (2000) present the Real Estate Agent system (REA) to exhibit the fulfillment of these properties. Employing both verbal and non-verbal interactivity, REA allows the user to interrupt her primary task of giving a house tour to ask questions. An ELIZA-like (Weizenbaum, 1966) action system is instantiated to manage the interaction, where a word pattern-detection system triggers a template-based response from the agent. REA's knowledge is instantiated with the KQML format, which lends itself to slot-filling, a technique used to direct conversational actions using information-seeking agent initiatives. Bickmore et al's (1999) work lays out the basic requirements necessary to define an ECA. Upon inspecting her dialog exchanges, REA's speech actions often prompt the user for very specific answers, such as 'yes' and 'no,' or just simple acknowledgements of affirmation. Despites its limited complexity, REA (Bickmore and Cassell, 2000) serves as a proof-of-concept platform for such HCI experiences.

Kenny et al (2007) describe a generalized architecture for ECAs. They list three primary goals for such entities: 1) believability, 2) responsiveness, and 3) interpretability. Believability refers to how natural an agent can make its behaviors within an interactive session. Responsiveness corresponds to the effectiveness of its performance within the bounds of its environment. Interpretability responds to how well the user can relate to the conversation agent's behaviors. With these attributes in mind, Kenny et al lay down their three-layered "Virtual Human" architecture, consisting of the Cognitive Layer, the Virtual Human Layer, and the Simulation Layer. The Cognitive Layer is described as a belief-desire-intention (BDI) engine, where the actual agent decisions are made when given a set of input signals. The Virtual Human Layer is seen as the avatar embodiment of the agent, where the input and output actuators are instantiated. The Simulation Layer reflects the overall environment in which the agent exists.

Sidner (2002) presents her findings on creating ECAs, as supported by four separate collaborative agent projects: an e-mail assistant, a scheduler, a VCR recording helper, and an entertainment center agent. In each of these examples, a standard agent archetype is used – one where the user and the machine employ a conversation-based interaction system to accomplish the human's goals, whether it be to set up an e-mail, schedule a meeting, program a VCR, or record a television broadcast, respectively. These agents are predecessors to the more complex DiamondHelp-based graphical user interfaces (GUI), developed by Mitsubishi Electric Research Laboratories (Rich et al, 2005) as well as the speech-based car electronics control field (Minker, et al, 2004) and the virtual personal assistant agents (Williams, 1996) (Nguyen, 2005) (Wobcke et al, 2005) (Babu et al, 2006). Sidner (2002) calls out the similarities between each of the agents, starting with the heavy use of context-based language understanding. Since each project has a well-defined task domain, there already exists an implied contextual bias toward the user's utterances. She also points out that all four agents use a mixed-initiative system. This means that both the user and the machine have the capability to trigger activity. From the user, s/he can initiate action using GUI-based interactions (not dialog-based interruptions), while the machine can accomplish tasks by directly asking the user questions through the text-based conversation interface. The initial human versus computer triggering of tasks, however, is different in each of the four scenarios. In the e-mail and VCR projects, the machine responds to the user's initial demands, while the scheduling agent first solicits the user for direction. The entertainment center agent allows for both initiation by the user and the machine, mainly because of its more complex menu of abilities, as compared to the other agents. Sidner's primary finding from this research remains her assertion that domain-specific ECAs work best using a *subset* language. She insists that the design of these entities requires a contextually-biased approach when dealing with the natural language understanding portion of these machines. Such a feature, however, tends to eliminate the chance for a truly open dialog between humans and computers, since the agent's repertoire of understandable vocabulary is severely pruned. These subset languages, however, tend to be the most efficient way to interact with a machine if the user is cognizant of the specialized command-style jargon. In some pseudo-open dialog systems, the user will eventually shape her/his utterances in order to operate effectively with the agent. (Tomko and Rosenfeld, 2004)

The imitation game, now known as the Turing Test, places a human judge at one computer terminal chatting with another terminal. The agent at the other computer may be a machine or another human. The judge must guess whether s/he is talking to another person or an

AI entity. Pattern-matching chatting agents, such as ELIZA, have often been used with mild success in the imitation game. Sing et al (2006) use an ECA, Artificial Intelligent Neural-network Identity (AINI), to play the imitation game. Their ECA used a state machine to drive their conversational behavior. In the case of AINI, five states comprise its response generation engine. The top state is a natural language understanding of the user inquiry as a means to extract a proper response from the WWW. If the Internet fails to give a proper answer, AINI will revert to a local Frequently-Asked Questions (FAQ) corpus to retrieve information. Failing this state, the Metadata Index Search is summoned, which provides a listing of possible sources of information for the user to sort through. If this stage does not bring success, AINI assumes a trick question has been posed, in which case it responds accordingly. As a default, the system will provide a random answer. A human expert continuously monitors this final state to refine its performance. Sing et al assert that AINI provides an equally effective chatting agent as its ALICE-based predecessors. Their work could be improved by quantifying AINI's utility in a number of various domain-specific applications.

Kopp et al (2005) present their museum guide ECA called Max. This agent is unique in that, for almost two years, it had operated inside of a publicly accessible museum, rather than inside a laboratory. Max's dialog management resembles that of most conversation agents, using a rule-based dialog manager to model its speech responses, a system similar to the ELIZA chatbot. (Weizenbaum, 1966) A version of JAM, a BDI engine, was used to create this mechanism. Kopp et al extend the rule-based method by incorporating contextual cues and long-term planning. Moreover, Max utilizes a multi-modal interface, where both verbal and non-verbal interactions are accounted for. To avoid errors in user interactions, a keyboard-based input system was instantiated. This eliminates many issues when collecting user utterances in a voice-recognition-based system. Kopp et al's work backed the idea that human users were willing to interact with an ECA in a social manner. While the research associated with Max was innovative in that an actual production system was used, the technologies associated in the system's development were not beyond the capabilities of the existing methods of its time. Additionally, Kopp et al were unable to address the effect that speech-based inputs have on dialog management, since they used a keyboard-based entry system.

Since the inception of Weizenbaum's ELIZA (1966), conversation agent technology has both spanned a wider range of applications as well as taken on new techniques and methods to improve upon its fundamental components. The evolution of chatbots progressed toward adding physical embodiments with the introduction of ECAs at the turn of the century. This section described some key developments in conversation agent research, covering both spoken and non-spoken types, and embodied and non-embodied styles. The next section discusses the use of contextualization to aid in both NLP and chatbot technologies. These context-based methods describe a group of techniques that employ contextual data to aid in agent behavior. For the purposes of this dissertation, context information helps to facilitate ASR disambiguation, knowledge management, and conversation agent discourse.

Context-Based Methods

Dialog system design can benefit from the tenets of contextualization. Context-based methods refer to the techniques used by a machine to drive behavior based on the immediate environmental state, also called the current *context*. Context-based reasoning (CxBR) formalizes this concept as a paradigm for agent behavior in which only a subset of an entity's total knowledge is active at any one time. (Gonzalez and Ahlers, 1998) (Stensrud et al, 2004) (Gonzalez et al, 2008) This architecture reflects the idea that humans themselves operate in a contextually relevant manner, where only a fraction of their knowledge is needed for different situations. The actual knowledge needed to function by a CxBR agent (and a human also) is a function of the state of its internal and external environments.

Context-Based Methods in NLP

Resolving semantic ambiguity remains a classic problem in NLP that continuously sees major advancements. One particular research avenue involves reducing the word or phrase identification search space by incorporating clues from the ambient conversational contexts. Contextualization effectively adds an extra layer of knowledge-based input to any reasoning system. ML-based semantic analysis methods have been enhanced by introducing context-based information into their training routines. (Mooney, 2006) In general, NLP problems can easily be enhanced through contextually-driven methods (Porzel and Strube, 2002), such as those found in spoken language translation (Levin et al, 1995) and knowledge modeling (Porzel et al, 2006). Perhaps the group of natural language protocol researchers that has benefited the most from contextualization is the ASR community. The following section presents related works in context-based speech recognition.

Young (1989) presented the MINDS system that incorporated context to make predictions and expectations on user's speech input. This infrastructure depends on four knowledge bases to create a context-based search space for predictions: dialog structure, task semantics, general world knowledge, and user knowledge. MINDS uses knowledge from the dialog structure in the form of an AND-OR tree. This tree delineates all of the goals, sub-goals, and domain subjects that will be relevant for the conversation. Task semantics refer to the knowledge associated with specific domains. General world knowledge fills the void for non-specific knowledge. The final knowledge source comes from a user model, which maintains data regarding the user's own personal goals and fact base. Using these four knowledge sources, search space constraints are established for the list of possible speech inputs. The predictions are structured in a layered manner as a measure to preserve probabilistic flexibility. The MINDS context-based grammar system was tested against a non-predictive grammar configuration. Young was able to see a 100% semantic accuracy with MINDS, while the non-predictive grammar produced only 85%. Semantic accuracy was defined as devising a correct database query. This reported success might seem suspicious, as a perfect accuracy score often raises questions of statistical integrity. His experimental results presentation does not give great detail about the testing, and his sample set included only ten trials, only two of which were female voices. Young could benefit from more experiments using a varied array of domains and a larger sampling of people. Adding these measures would support the credibility of his 100% semantic accuracy claims.

While not directly affiliated with speech recognition technology, Kladke (1989) and Towhidnejad (1990) present methods of using contextual data to perform automatic semantic clarification in a knowledge base. Kladke (1989) provides a context-based solution for determining missing or erroneous details in component descriptions found in a computer-aided design (CAD) representation of process controls. Her work focuses on narrowing down the list of possible component candidates using constraints induced by the contextual cues inferred from the existing information. Towhidnejad (1990) extends Kladke's work by performing a similar constraint-based information population mechanism for actual CAD design drawings.

Serridge (1997) improves phoneme-driven, or segment-based, speech recognition with contexts from domain information. A Viterbi search was modified to traverse a search space pruned by a context-dependent speech model. These models calculated location-based relationships using the corpus data to determine contextual associations. Several types of such models were examined, such as word-dependent phones, left and right biphones, and triphones. Serridge determined that word-dependent phones and right biphones performed best by delivering 20% fewer errors than its context-independent counterpart. Another model, the boundary model, used the corpus training data to generate segment transition groupings. This computationally-expensive model was tested, yielding a 40% reduction in word errors. Thus, Serridge showed that the use of contexts does not necessarily equate to better results. The trade-off in using boundary models over context-based models lies in the requirement of large amounts of computing power. Nevertheless, this work shows promise that context-dependent models can be used as inexpensive ways to reduce speech recognition errors, and with some improvements, they may be able to perform as well as boundary models.

King et al (1998) used phonetics with Hidden Markov Models (HMMs) to recognize speech. The system they present uses a neural network (NN) to detect phonetic features and an HMM-based decoder that transforms these features into phonemes. The NN was trained using the TIMIT database, which provides a plethora of labeled phonemes and word boundaries. The network's goal was to output the correct phonetic label. While speech recognition at the phoneme level may seem like an intuitive approach, the extensive amount of processing power required made such an effort impractical at the time. A NN would be better utilized for a higher-level recognition task, such as training a network to for determining patterns within a conversation.

A direct correlation between context usage and conversation agent discourse exists in Sammut's ProBot (2001). Using a Prolog expression rule base, user inputs are matched to fire off resulting output responses. Sammut's use of contexts comes into play when unexpected utterances are received, requiring the use of contextually-organized hierarchical information to align ProBot for an appropriate response. This hierarchy notation comes in the form of an activation level. When the user has expressed interest in a certain context, all relevant rules for this context are promoted to a higher hierarchy level. The basic idea that Sammut's work promotes is the use of focusing the agent's speech action engine toward a particular subset of its entire knowledge base. While this is a valid notion, ProBot's manually annotated rules base proves to be too laborious of a task to keep the system portable and generic.

Goulian et al (2003) present the *Romus* system to fortify French speech understanding against speech anomalies. These imperfections in speech patterns occur when a speaker includes random pauses or attempts to fix a previously stated word. Romus operates using two components, the chunker and the link-grammar parser. The chunker portion acts like a traditional NLP sentence parser, complete with tagging and segmenting capabilities. An additional feature to Romus' chunker, however, is the inclusion of tagging for anomalous speech points, such as repairs, repeats and hesitations. The link-grammar parser assigns each chunk to a data point consisting of three elements: a category, a lexical head, and morphological cues. Romus uses a dictionary whose entries are encoded in a similar three-point structure. The link-grammar parser will use both relational and logical requirements to determine whether a set of chunks can be properly parsed. Romus was evaluated using three types of problematic speech: spontaneous, unorthodox word ordering and complex uttering. In each of these cases, error rates of 7.4%, 5.1% and 5.1% were achieved, respectively. When all three speech types were combined, a test most resembling human-to-human conversations rather than human-machine interactions, Romus performed at a 39.5% error rate. Goulian et al do not clearly explain this substantial increase in error rate, but they do claim that it is comparable with that of similar natural dialog systems. Nevertheless, they presented an effective technique in preventing language misunderstanding in a general conversational setting. Their work could be improved by providing more detail and expanded testing on how the aggregation of the three speech types affects Romus' overall performance.

Fügen et al (2004) reduce speech recognition errors by a knowledge-based dialog manager. They used *Ibis* as the speech recognizer, which supports context-free grammars and traditional n-gram language modeling. *Tapas* was chosen as the dialog manager, which allows for close coupling with Ibis. Its implementation effort was minimal in that only domain-dependent information was required for operation. The interaction between Tapas and Ibis induced a system of information sharing between the language parser and the dialog processor. Fügen et al saw a 3.3% reduction in word errors from 23.5% for close proximity conversation, and a 9.9% reduction from 34.9% for interactions between distant interlocutors. This work could be improved by showing results for a wider array of different domains.

Yan and Zheng (2004) use contexts to serve as dialog constraints to provide more effective speech recognition. They point out four primary functions of a speech-based interface system: speech recognition, language parsing, dialog management, and speech synthesis. Yan and Zheng recognized that the appropriate interaction of these systems requires an emphasis on pragmatic approach, especially when dealing with real-world applications. In designing this architecture, they asserted that conversation topics exist as *expected focuses*. For each expected focus, a set of associated words comprises the rules that negotiate the understanding of the dialog. The relationship of these rules makes up the expected focus' finite state network (FSN). This conversational architecture was implemented in the *EasyFlight* airline booking system. The expected focus concept draws many parallels to contexts. Yan and Zheng chose to use pattern recognition for creating contextual rules, rather than a set of manually prescribed rules. A hybrid rule set using both techniques would enhance the performance of their system.

Sarma and Palmer (2004) implement a context-based prediction model to improve recognition and repair of speech inputs. They recognize that Automatic Speech Recognition (ASR) technologies pose a growing trend that is not without its early-life flaws. Speech-based information sources, such as telephone calls and news broadcasts are often targets for ASR transcription, but current advances in such techniques contain many word errors. Some researchers have attempted to skirt the ASR problem using discourse-based tactics, such as form-filling (Kang et al, 2009) and partial speech recognition (Sagae et al, 2009). These systems depend on domain-specific methods outside of the ASR results to formulate a set of conversational actions. Sarma and Palmer's approach, on the other hand, uses linguistic methods directly upon the ASR output. They present an ASR error-reduction method that comprises three parts. The first component analyzes raw ASR output, producing patterns of word co-occurrences. These patterns, or contexts, give their system a set of surrounding words for each singular vocabulary word. The second component deals with user information requests, in which a single search term is associated with its context words when performing the query. The final portion of their system finds words that are phonetically close to the search term, but are not affiliated with the context words. While Sarma and Palmer's approach heavily depends on pattern recognition theory, an added level of semantic knowledge could enhance their work. This addition could involve the use of domain-specific ontologies, or even pre-specified word patterns.

Lieberman et al (2005) use common sense knowledge to improve speech recognition. They recognize that traditional context-based speech disambiguation methods were purely statistical processes, where clusters of words are analyzed and attributed with certain probabilities. Lieberman et al developed a semantic relations database named ConceptNet, whose knowledge is based on the Massachusetts Institute of Technology (MIT) common sense database, Open Mind. With this tool, they built a machine capable of understanding conceptual relationships amongst different English words. ConceptNet was combined with the Microsoft Speech Engine, and the resulting software produced a re-ordered list of candidate hypotheses for a user's voice input. This new, context-based ordering scheme saw a 17% error prevention rate. Lieberman et al describe a system that functions under an enormous rule base in ConceptNet. This is acceptable for a proof-of-concept, but under more practical circumstances, this extraordinary expense in data cannot be maintained. A more suitable approach may be to selectively incorporate subsets of ConceptNet's knowledge as contexts.

The context-based techniques in this previous section all demonstrate how contextual cues can be exploited to enhance NLP tasks. Speech recognition was a primary focus because of its direct applicability in speech-based NLP agents. The common theme in these works was that the addition of contextual information enhanced the semantic disambigutation accuracies for ASR systems. It is conjectured that employing context-based methods to other NLP efforts, such as dialog management, will cause similar gains in system performance. Thus, this dissertation proposes a speech-based dialog management system that extends the use of context beyond speech recognition, applying its principles in dialog and knowledge management.

Chapter Summary

Current assistive dialog systems are meant to allow human users to complete a single task in a serial manner. This style of conversation comes off as awkward, as typical human-to-human conversations do not always follow a linear train of thought. Ideally, an open style of interaction is preferred, as the user is not constrained to reply with certain responses. Additionally, robustness must also be applied to the system's ability to respond in the presence of error-prone ASR.

This chapter delineated the current state of the art of technologies involved in designing such a dialog management system, with direct focus on NLP, dialog systems, and context-based methods. Each of these items was presented with a set of real-world research applications. These three topics align to the four themes pertaining to the context-centric dialog management system that is the subject of this dissertation: conversation agents, open dialog, ASR and domain-independent knowledge management. Conversation agent design has been affected by both NLP and dialog system techniques. Open dialog has been featured by NLP applications and QA-based dialog systems. ASR has utilized both NLP and context-based methods, while domain-independent knowledge management can be considered in terms of NLP and dialog system design. The next chapter gives a concise yet formal introduction of the problem addressed by this dissertation in light of these four major dialog management themes.

CHAPTER THREE: PROBLEM DEFINITION

Popular media has introduced us to notional machines that can converse with humans using spoken natural language. These share the characteristic of using spoken human language to interact in a very natural manner. In reality, as pointed out in the previous two chapters, this is a very difficult problem to solve, not only from the standpoint of 'catching' the spoken words, but of making sense out of them and composing an intelligent response based on knowledge. To further complicate the issue, natural conversations often take on a non-linear discourse pattern, where different thought processes can surface and re-surface at any time during an exchange. From this discussion, it is established that the development of naturalness in speech-based conversation agents is a path fettered by technical challenges. Hence, the general problem in this dissertation is to elevate the level of speech-based discourse to a new level of naturalness in ECAs by carrying an open dialog.

Specific Problem

The general problem declared in the previous paragraph yields three specific problems to be addressed by this dissertation:

- Overcoming the limitations of ASR technology, with the full expectation of relatively low speech recognition rates
- Developing a knowledge management system that is domain-independent
- Formulating a dialog discourse model that allows for user response openness

This list describes three issues that affect natural language understanding in an ECA with a less than ideal speech recognition environment. The first specific problem deals with ASR limitations, whose main obstacle is the current state of speech recognition technology. Speech recognition has not matured to the point of having reliable accuracy, causing any subsequent processes dependent on ASR output to inherit these inaccurate results. Two main factors hindering this performance of speech recognizers are noise-ridden environments and non-native speakers.

The second specific problem addresses the need for a generalized infrastructure for ECA knowledge. Equipping conversation agents with domain expertise would benefit from a domain-independent response system. Many ECA designers spend a large amount of time custom tailoring their agents' output utterance behavior. A quicker agent development lifecycle would result if the expert knowledge base could be modeled as a plug-and-play style system. Moreover, tying the knowledge management with a speech recognition system can enhance an ECA's ability to identify a user's conversational intentions. Enhancing the knowledge manager with domain independence would extend its generalized capabilities into the ASR system.

Open dialog remains the third specific problem, as current conversation agent technology has fallen complacent upon discourse models that only support a single train of thought at a time. Often times, an ECA will have certain expectations on what the user can say for each turn, a constraint that limits verbal input mobility. By allowing the user a wider range of possible utterances through open dialog, the HCI experience can have a more natural feel. While an open discourse model makes for more natural conversation, it expands the challenges of speech recognition, as the system must be ready to identify all sorts of unexpected user utterances.

Hypothesis

It is hypothesized that a context-based dialog manager for an effective ECA spoken interface can overcome ASR-related errors, implement domain-independent knowledge management, and provide open dialog discourse.

Contributions

The research described here differs from all other work on spoken dialog systems because it strongly addresses the use of context-driven behavior to overcome the challenges of open dialog and domain independence in conversational flexibility. The specific contributions are as follows:

- A robust method of servicing open user input in spite of weak-performing speech recognition facilities.
- A complete mixed-initiative architecture based on CxBR that manages an assistive dialog using speech-based communication.
- A domain-independent knowledge management system for information-deployment conversation agents.
- A prototype metrics system used to evaluate the devised approach.
- Publishable results and data evaluating the effectiveness of a context-centric discourse-based dialog manager.

CHAPTER FOUR: APPROACH

This chapter describes the overall conceptual approach of the dialog manager featured in this dissertation. Architecting the system incorporated three major design decisions: user input processing method, knowledge management, and agent response discourse mechanism. These features describe the defining challenges in dialog management. This chapter provides an explanation of how dialog managers of the past have handled them and describes how the dialog manager in this dissertation resolves them. Figure 3 shows the flow of information for a generic dialog manager between each of these primary components.

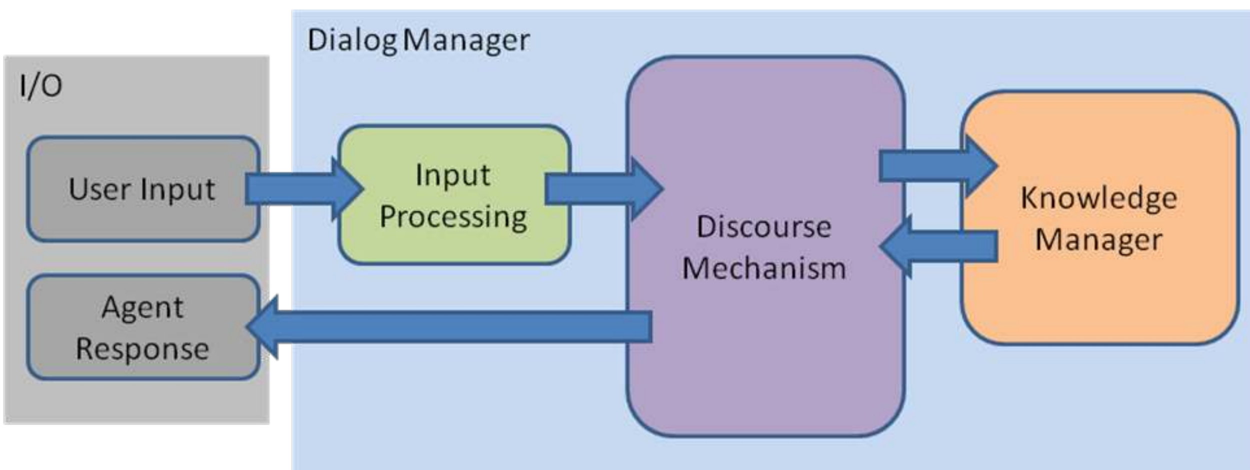


Figure 3: Major dialog manager sub-systems

For each of the major blocks (Input Processing, Discourse Mechanism, and Knowledge Manager), a series of known techniques are described, discussing of their advantages and disadvantages. From these approaches, a design decision is chosen and justified. The aggregation of all three sub-systems selections results in the final approach of the proposed context-centric dialog manager.

User Input Processing Method

The first task for a dialog manager is to interpret the incoming user utterance. The method in which this is done is often dictated by the data representation in which the remainder of the dialog manager handles the input. For this dissertation, only text-based entries will be considered, where the text is either directly typed from a keyboard, or, for speech-based systems, produced by an ASR module. The alternative to this assumption would be the permission of phoneme-based inputs, which would be beyond the scope of this work, as the aim of this dissertation is to handle *conversational* ASR errors using high-level software-based algorithms. Overcoming phoneme identification issues implies the use of lower-level signal processing and hardware-based solutions.

Dialog systems have effectively demonstrated four types of input processing methods: direct query, semantic NLP, statistical categorization, and expectation-based matching. The next section is devoted to describing each of these styles.

Direct Query

Direct query refers to the simplest method of parsing a user's utterance. This is done by simply passing along the raw input, unaltered. Many primitive dialog systems can utilize this method if they seek simple answers, such as posing yes and no questions (Bickmore and Cassell, 2000) (Lee et al, 2005) (Kenny et al 2009) or offering highly-constrained options (Litman et al, 1998) (Johnsen et al 2000). For open dialogs, this method can prove flawed, especially if an ASR does not provide accurate dictation results. Additionally, no semantic information is attached to the incoming input, which does not address ambiguity issues. Table 1 gives some examples of potential flaws in direct query processing when dealing with speech-based systems.

Table 1: Direct query input processing examples

User Utterance	Direct Query Processing Result	Hypothetical ASR Result	Direct Query Processing Result
"Yes, thank you."	yes thank you	"Yes think U"	yes think u
"I have three dollars."	i have three dollars	"I halves free dollar"	i halves free dollar
"Where can I buy the yellow book?"	where can i buy the yellow book	"Wear kennel by the yelling brook"	wear kennel by the yelling brook

From these examples, it can be seen that the ASR result could yield an entirely different semantic meaning from the actual utterance. This, in turn, causes the direct query input processing to pass on the flawed results.

Semantic NLP

Andernach et al's box office attendant, SCHISMA (1995), Huang et al's travel agent LODESTAR (1999), and Skantze et al's walking navigator, Higgins (2006), process user utterances using semantic information. Adding a semantic computation layer to the direct query method lends itself to NLP-based input processing. Using whatever means of NLP on a user utterance can help in the disambiguation process for the rest of the dialog manager. The level of NLP involvement, however, can drastically vary between systems. Additionally, added NLP functionality means higher computational cost. The three main forms of NLP analysis, in order of ascending complexity, are: POS (parts-of-speech) tagging, phrase chunking, and parse tree generation.

POS Tagging

A simple POS tagging could be used to identify the nouns, verbs, and adjectives. This method is most applicable when the dialog manager is looking for a particular POS type, with little need for the other supporting input words. A poorly recognized ASR output, however, may result in sub-standard POS tagging results, since only a portion of the user utterance may be accurately recognized.

An example of POS tagging can be illustrated by a user input of “My car is red.” The tagger could recognize just the adjective ‘red,’ or the noun ‘car,’ depending on whether the dialog manager is expecting to process an adjective or a noun. The remaining input words can be discarded, assuming the system is only interested in nouns and adjectives. A weak ASR, however, may pick up the words “Milk car is red,” or even worse, “Milk care instead.” These instances could pose discourse problems for the dialog manager, as the nouns “milk,” and “care” could be mistakenly processed through the system. Table 2 gives examples of POS tagging for hypothetical utterances and their ASR outputs.

Table 2: POS tagging processing examples

User Utterance	POS Tagging Result	Hypothetical ASR Result	POS Tagging Result
“Yes, thank you.”	(yes PARTICLE) (thank VERB) (you PROPER NOUN)	“Yes think U”	(yes INTERJECTION) (think VERB) (u PROPER NOUN)
“I have three dollars.”	(i NOUN) (have VERB) (three CARDINAL NUMBER) (dollars PLURAL NOUN)	“I halves free dollar”	(i NOUN) (halves VERB) (free ADJECTIVE) (dollar NOUN)
“Where can I buy the yellow book?”	(where WH-ADVERB) (can MODAL) (I NOUN) (buy VERB) (the DETERMINER) (yellow ADJECTIVE) (book NOUN)	“Wear kennel by the yelling brook”	(wear NOUN) (kennel NOUN) (by PREPOSITION) (the DETERMINER) (yelling VERB) (brook NOUN)

These POS taggings show the effect of using flawed ASR results, where mistaken word pairs (i. e., yellow and yelling, three and free, thank and think) will lead to mistaken POS pairs.

Phrase Chunking

Phrase chunking of the input adds complexity to POS tagging, since entire phrases are identified. These can be noun phrases, verb phrases, prepositional phrases or adverb phrases. Phrase chunking simply groups grammatically related adjacent atomic words. While more computationally-expensive than POS tagging, phrase chunking is useful when the dialog manager is looking for specific groups of words.

The input sentence “My red car is broken,” consists of two phrases – the noun phrase “my red car,” and the verb phrase “is broken.” A dialog manager could use either of these phrases, which also includes separate POS tags for each individual word, for further processing. Nevertheless, an error-prone ASR could drastically alter these results with by detecting the string “Myron’s card has spoken.” NLP chunking would yield the noun phrase “Myron’s card” and the verb phrase “has spoken.” These phrases could cause the dialog manager to perform ineffectively. Additionally, there is a possibility that the NLP chunker could falter from ambiguity problems. For example, the sentence “He ate the chicken in the cart,” could be chunked as the noun phrase “the chicken in the cart,” or as the noun phrase “the chicken” with the prepositional phrase “in the cart.”

Table 3 gives examples of phrase chunking, with and without ASR usages. From these instances, it is exhibited that an inaccurate ASR result will cause discrepancies between the machine’s chunking results and the user’s actual utterance chunking results.

Table 3: Phrase chunking processing examples

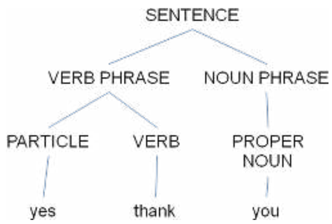
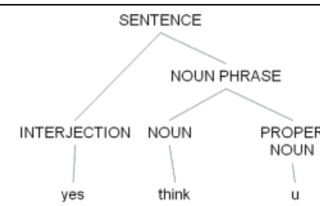
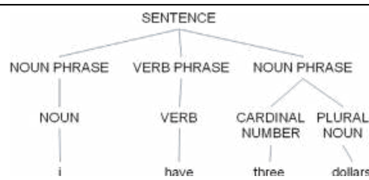
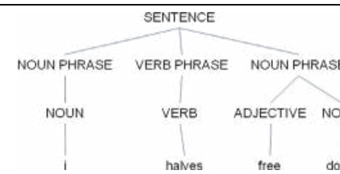
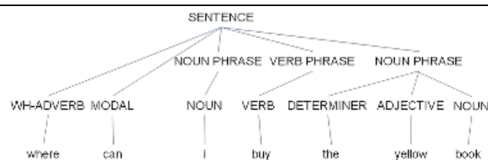
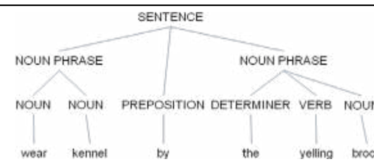
User Utterance	Chunking Result	Hypothetical ASR Result	Chunking Result
"Yes, thank you."	<VERB PHRASE (yes PARTICLE) (thank VERB)> <NOUN PHRASE (you PROPER NOUN)>	"Yes think U"	(yes INTERJECTION) <NOUN PHRASE (think NOUN) (u PROPER NOUN)>
"I have three dollars."	<NOUN PHRASE (i NOUN)> <VERB PHRASE (have VERB)> <NOUN PHRASE (three CARDINAL NUMBER) (dollars PLURAL NOUN)>	"I halves free dollar"	<NOUN PHRASE (i NOUN)> <VERB PHRASE (halves VERB)> <NOUN PHRASE (free ADJECTIVE) (dollar NOUN)>
"Where can I buy the yellow book?"	(where WH-ADVERB) (can MODAL) <NOUN PHRASE (I NOUN)> <VERB PHRASE (buy VERB)> <NOUN PHRASE (the DETERMINER) (yellow ADJECTIVE) (book NOUN)>	"Wear kennel by the yelling brook"	<NOUN PHRASE (wear NOUN) (kennel NOUN)> (by PREPOSITION) <NOUN PHRASE (the DETERMINER) (yelling VERB) (brook NOUN)>

The third example exemplifies the idea that longer user utterances will yield more ASR errors, and thus, amplifying the final chunking result differences. The next section will describe how this effect is compounded when structural information is processed through NLP parse trees.

Parse Tree Generation

A full semantic analysis of the user input would map the utterance into an NLP parse tree, containing all POS tags, phrase chunks and chunk nestings. This style of input processing requires the most computational power, and often results in longer execution times. Hence, parse trees are better used as pre-execution routines. Chunk nestings give sentence structure data to augment the chunking and tagging information. Again, ineffective ASR results can greatly affect the parsing of a user utterance. In fact, the lack of highly accurate ASR facilities makes parse tree generation a particularly inefficient method for spoken dialog manager applications. Table 4 shows examples of parse tree processing. It can be argued from these instances that differences in the ASR results and the user utterance will result in differences in their resulting parse tree structures.

Table 4: Parse tree processing examples

User Utterance	Parse Tree Result	Hypothetical ASR Result	Parse Tree Result
"Yes, thank you."		"Yes think U"	
"I have three dollars."		"I halves free dollar"	
"Where can I buy the yellow book?"		"Wear kennel by the yelling brook"	

Statistical Categorization

Statistical categorization input processing has been implemented by dialog managers that strive for context-free operation. Gandhe et al (2009) use ML for their tactical questioning training ECA, Hassan. The idea behind this ML-based method is to prepare the user utterance such that it fits a lexical relationship derived from a training session with a massive corpus. This style of interpretation is often used in conjunction with an input-output pairing discourse method, where the user's words analyzed to determine the most statistically-relevant response reaction, as dictated by the training data.

Statistical categorization eliminates the need for any manual manipulation of large amounts of corpus data. The disadvantage of this method, however, reflects the same issues with any ML-based solution, where the machine's full automation of speech action selection may result in an unnatural or awkward sequence of agent responses.

Expectation-based Matching

On the other side of the spectrum from ML, knowledge-based systems have provided solutions for machine autonomy using manual data modeling, largely overseen by a domain expert. As opposed to statistical categorization, a knowledge-based approach for input processing is exhibited in expectation-based matching. In this method, a pre-determined set of information is provided within the dialog manager to help with interpreting the user utterance. These matching templates may come in the form of keywords, slots or patterns.

Keyword-matching

Matching keywords is the simplest of all of the expectation-based input processing styles. Given a list of expected keywords and the word-by-word parsing of the user utterance, the dialog manager can easily discern which parts of the input string are usable. This method can easily be implemented for option menu-enabled systems. A drawback for keyword-matching is the need to manually program the words that must be identified. Additionally, different keywords can only apply to certain situations, so an entire keywords state infrastructure needs to be defined when using this input processing method. Babu et al’s message-taking virtual receptionist, Marve (2006), Sidner’s collaborative interface engine, Collagen (2002) and Wobcke et al’s Smart Personal Assistant (2005) uses keyword detection for interpreting user utterances. Table 5 gives hypothetical keyword-matching examples.

Table 5: Keyword-matching input processing examples

User Utterance	Keyword-matching Processing Result	Hypothetical ASR Result	Keyword-matching Processing Result
“Yes, thank you.”	{yes}	“Yes think U”	{yes, think}
“I have three dollars.”	{three, dollars}	“I halves free dollar”	{halves, free, dollar}
“Where can I buy the yellow book?”	{where, buy, yellow, book}	“Wear kennel by the yelling brook”	{wear, kennel, yelling, brook}

The first two examples show that even an inaccurate ASR can retain some of the same keywords that the user has uttered. The last example reflects the same phenomenon that plagues every input processing method, where a differing ASR result will amount to a differing input processing result.

<u>Slot-Identification</u>

Slot-identification refers to the parsing of user input for certain words that would fill a database query slot. This method would be similar to keyword-matching if not for the fact that a database querying system is driving the expected utterances. The development of slot-identification systems can prove tedious, as the expected slot categories must be explicitly specified ahead of time through manual means. Travel information deployment systems often use slot-identification techniques, as seen in PURE (Agarwal, 1997), MALIN (Flycht-Eriksson and Jönsson, 2000), LMSI Arise (Lamel et al, 1998), and DACST-AST (Niesler and Roux, 2001). Business applications have also embraced this style of input processing, as featured in inventory queries (Owda et al, 2007), e-mail management (Williams, 1996), process models (Lemon and Liu, 2006), and financial services (Hardy et al, 2003). Table 6 shows some hypothetical slot-identification processing scenarios. As with previous examples, the error-prone ASR results can substantially affect a database query.

Table 6: Slot-identification input processing examples

User Utterance	Slot-Identification Processing Result	Hypothetical ASR Result	Slot-Identification Processing Result
"Yes, thank you."	{yes} ➔ Agree = "yes"	"Yes think U"	{yes, think} ➔ Agree = "yes" AND NextAction = "think"
"I have three dollars."	{three, dollars} ➔ Value = 3 AND Currency = "dollars"	"I halves free dollar"	{halves, free, dollar} ➔ Item = "halves" AND Value = 0 AND Currency = "dollars"
"Where can I buy the yellow book?"	{where, buy, yellow, book} ➔ Answer = "location" AND Action = "buy" AND Color = "yellow" AND Item = "book"	"Wear kennel by the yelling brook"	{wear, kennel, yelling, brook} ➔ Action = "wear" OR Action = "yell" AND Item = "kennel"

Pattern-matching

The original ELIZA (Weizenbaum, 1966) chatbot utilized a pattern-matching style of user input interpretation. In this method, the user input is filtered using a regular-expression-like pattern library. The patterns are hierarchically ordered, such that more specific filters take precedence over their more general peers. This method of utterance processing is best used when a clean user input string is retrieved. The fine-grained requirement of proper sentence formation for adequate pattern-matching means poorly performing ASR systems would be highly detrimental for this style of input interpretation.

ALICE, the modernized version of ELIZA, continues the pattern-matching chatbot tradition with its pattern categories: atomic, default and recursive. (Atwell and Shawar, 2007)

Patterns take on a literal structural form, detecting a certain combination of words, spaces and wildcards (represented by asterisks). Atomic patterns do not use wildcards, acting as a keyword finder. An atomic pattern could be represented as *<pattern>Yellow book<pattern>*, where the only pattern match for this category is the string "Yellow book." Default categories represent less specific patterns, usually consisting of a keyword followed by a wildcard. An example of a default pattern could be *<pattern>Yellow *<pattern>*, where both "Yellow book" and "Yellow rubber duck" could be possible pattern matches. Recursive categories use extra reduction rules to process the user utterance, such as re-phrasing, separating, or synonym-matching. The reduced version of the input is then run through the pattern-matching process. A recursive pattern could match the utterance "Yellow novel" to *<pattern>Yellow book<pattern>* if a synonym-based reduction rule *<pattern>novel<pattern>* is available to transform "novel" to "book."

Pattern-matching has exhibited its strength as an easy method for performing input-output response style conversation by simply creating a category rule base and letting the ALICE engine run its course. The disadvantage, however, is that the depth of follow-up conversation is lacking and the sheer volume of rules needed is vast. Hence, the ALICE framework suffers from ELIZA's original problems of offering a shallow chat exchange and requiring a sizable response modeling effort. Nevertheless, many conversation agents have opted for ELIZA-style input processing, such as MAGA the museum guide (Augello et al, 2006), Valerie the robo-receptionist (Gockley et al, 2005), VIRMA the insurance risk analyzer (L'Abbate et al, 2005), TARA the terrorism information repository (Schumaker et al, 2007), and COGITO the E-commerce guide (Thiel and Stein, 2000).

Input Processing Method Approach

For this dissertation, the input processing approach combines keyword-matching with an NLP treatment. The idea is to gain a sense of context with each user utterance. Full understanding of an input string may be considered a fruitless effort, since it cannot be guaranteed that an ASR can provide perfect or even near-perfect accuracy. For computational efficiency, only a portion of the user input should be examined. A simple NLP chunking of the utterance gives the dialog manager just enough information to develop the contextual frame it needs. What results is a filtering of user input that is boiled down to the lone keyphrases that can be used for matching purposes. Table 7 gives examples of this proposed chunk-based keyphrase-matching input processing.

Table 7: Input processing examples

User Utterance	Input Processing Result	Hypothetical ASR Result	Input Processing Result
"Yes, thank you."	{yes}	"Yes think U"	{yes, think}
"I have three dollars."	{three dollars}	"I halves free dollar"	{halves, free dollar}
"Where can I buy the yellow book?"	{buy, yellow book}	"Wear kennel by the yelling brook"	{wear kennel, yelling brook}

These instances exemplify the idea that identifying specific noun phrases (i. e., free dollar, yelling brook) and verb phrases (i. e., wear kennel) allows for an extra layer of contextual cues. This keyphrase recognition can instantly give the conversation agent clues on whether the ASR has completely missed the user's utterance and immediately take action, such as re-posing questions. A technically detailed explanation of this method is provided in the next chapter. The following section covers the opposite side of the input/output processing with discussion on the different methods of maintaining output response knowledge.

Knowledge Manager

On the other side of the user input is the agent output. These responses may be canned text strings, either manually modeled or fabricated from ML techniques, or they may be dynamically produced from a database query, or they could be directly fetched from a Web source. The speech action repository that a dialog manager uses to create responses exists in the knowledge manager. Depending on the application of the agent, the knowledge management style can vary its data-keeping in terms of accessibility, structure, and construction. The following types of knowledge management for a dialog manager are: scripts, action-reaction pairs, relational databases, and corpus-based sources. This section describes each of these knowledge repositories.

Scripts

Scripting deals with the direct authoring of speech actions. This style of knowledge pairs easily with finite state-based discourse methods. Given a certain state of a situation, the agent's exact speech actions are provided in a pre-meditated set of scripted statements. This was exemplified in the ATC work of Schaefer (2001) and Bickmore and Cassell's small talk-based real estate agent Rea (2000). Both of these systems had clearly defined finite states of conversational discourse.

For quick prototyping work, scripting can be a simple method of getting an agent to react to a user utterance. Most of these efforts, however, cannot afford to have huge volumes of pre-defined scripts, as the time needed to develop these dialogs can be quite sizeable. Some prototype ECA work, such as Cassell et al's children's playmate Sam (2000), Massaro et al's facial interface Baldi (2001) and Lee et al's robotic penguin Mel (2005), relied on scripts to simply give some semblance of a response to its users. These systems were more focused on non-discourse related aspects of ECA research

Action-Reaction Pairs

Action-reaction pairs refer to an input-driven response system, where a user action immediately triggers the firing of an agent output. This knowledge management style exists as an extensive repository of these pairings, essentially a text-based hash table. Three methods of producing these lookup entries are discussed in the remainder of this section: rule base modeling, question-response pair generation and Wizard-of-Oz experimentation.

Rule Base Modeling

An explicit rule base reflects the action-reaction style of agent behavior, such those seen in expert systems. (Gonzalez and Dankel, 1993) Creating this collection of rules is often a manual effort. Rule bases are sufficient for conversation agent purposes when the subject matter is well-defined and tightly constrained. Kopp et al's agent, Max (2005) uses a rule base for guiding visitors around a museum. Since Max's knowledge is limited to the domain of the museum's information, a thorough modeling of this data through building a rule base is a reachable goal.

When open responses are expected, however, the rule base creator may be overwhelmed with handling this wider range of inputs.

Many of the descendents of Weizenbaum's pattern-matching ELIZA (1966), such as those built upon the AIML framework (Wallace, 2002), can attest to using the rule base method. For these systems, a hierarchy instilled within the pattern-matching rules is provided. This allows for the case in which multiple, simultaneous rule firings occur. As with Kopp et al's work (2005), many of these ELIZA-like systems have a very narrow realm of expertise, such as travel information (Allen et al, 2001), health assistance (Turuenen et al, 2008), automotive electronics (Baca et al, 2003), and appliance control (Harris and Rosenfeld, 2004).

L'Abbate et al's AIML-based insurance risk analyzer, VIRMA (2005) has expanded the rule base method to include Web services to augment the system knowledge. These services use a case-based reasoning algorithm to produce an agent response when the local AIML rules cannot fire. A similar effort in Augello et al's museum guide agent, MAGA (2006) uses the Cyc (Lenat, 1995) ontology to enhance its pattern rule base with a common sense database.

Question-Response Pair Generation

Automatic generation of action-reaction pairs represents the opposite end of the spectrum to the hand modeling of a rule base. Vrajitoru and Ratkiewicz's genetic algorithm (GA) efforts (2004) sought to achieve this autonomy of response generation. ML-based efforts, however, have been the weapon of choice for this area of research. A repository of question-response pairs lends itself to the statistical categorization method of input processing. With this set of data pairings, statistical profiling of a question will provide a mathematically-derived response match. As with any ML effort, pre-processing of a source corpus must be completed to fabricate each question-response pair. Huang et al (2007) used an online forum as a source corpus. Colby's PARRY (1973) incorporated excerpts from autobiographic essays from paranoia patients. Jabberwocky (Voth, 2006) sourced its responses from a series of past conversations.

The downside of creating a question-response repository through automatic means is that the quality of responses may be lacking if no control measures are taken. The effectiveness of the ML process is only as good as its sourced training data. It may be the case that human intervention may be needed to manually massage the resulting question-response pairings for quality control.

Wizard-of-Oz Experimentation

A Wizard-of-Oz (WoZ) response corpus refers to the use of human-to-human experimental data to determine the most common interactions that could occur in a human-to-machine exchange. WoZ data effectively predicts what a human might say if s/he encounters a conversation agent. For earlier systems, such as SCHISMA (Andernach et al 1995), a prescribed answer is hand-tailored for each of these predicted utterances. In more recent efforts, an ML-based matching system is used. This can be seen in Artstein et al's Sergeant Star (2009), Gandhe et al's Hassan (2009), and Kenny et al's Justina (2009).

WoZ experiments are preferable for taking open tasks down to a more constrained environment. Human responses can be more predictable when there is a large enough WoZ data set. The disadvantages of this style of knowledge management, however, is the fact that not *all* user responses can be accounted for and that conducting the WoZ experiments requires an enormous time commitment.

Relational Databases

Data-driven dialog managers use relational databases to provide the knowledge back-end for conversational content. Relational database knowledge management in dialog managers essentially reflects the idea of using speech-based commands to access an organized information source. These interactions can be described as a back-and-forth exchange between a computer and a human, with the machine trying to figure out its user's data needs. This type of system is effective for looking up itemized data entries, such as library books (Cenek, 2001) (Ahad et al, 2007), commercial inventory (Thiel and Stein, 2000) (Owda et al, 2007), and transportation routes (Lamel et al, 1998) (Litman et al, 1998) (Johnsen et al, 2000). Goh et al's crisis communication system (2006) and Stede and Schlangen's chatbot (2004) both expanded their knowledge base facilities by including multiple databases to cover a wider range of domains.

Slot-filling dialog managers are almost always paired with relational databases, as they utilize discourse models based around building queries for fetching data. The actual response output for a slot-filling system tends to converge toward a follow-up questioning session as needed query points are discovered. The fact that a relational database back-end does not contain actual response information remains a glaring limitation of such knowledge management for a dialog manager. What happens is that the front-end discourse model is burdened with surrounding a back-end database query with natural language to drive the HCI experience. This style of conversation resembles that of an expert system attempting to pin down its final answer, rather than that of human-to-human banter.

Corpus-based Sources

The corpus-based method of knowledge management for dialog managers simply takes declarative statements from a body of text and repeats it as a response to the user. The challenging task here lies in the proper identification of *what* information the user is seeking. When using corpus-based sources, two problems can arise given a user utterance: 1) the user request is not answered correctly, and 2) the user request is answered correctly, but rife with verbosity or extraneous information. These are the same problems that QA systems face. The difference between a traditional QA infrastructure and an informational assistive conversational agent, however, is the layering of dialog constructs in the chatbot's discourse that makes its HCI experience more like a human-to-human exchange. On the other hand, a QA system's dialog demeanor resembles that of a search engine with some incorporation of natural language.

Kurohash and Higasa's Virtual Help Desk (2000) explicitly used a set of dictionary entries to contain all of their agent's speech actions. Inui et al (2003) developed an informational chatbot whose response base was rooted entirely in a single corpus. They emphasize that their method excels in language-independence and domain independence, which would also hold true for Kurohash and Higasa's work. The common thread between these agents is that each used a keyword-matching system to determine which portion of the corpus to deploy to the user. From these examples, it is apparent that corpus-based knowledge management can be used as a direct source of dialog responses. The advantage of this method lies in the idea that an established information source can be used immediately without added preparation effort or hand-modeling strife.

The Lifenaut project (Mayer, 2009) explicitly asks users to populate a corpus with rote memory data points. From this data, they claim their work can re-construct a full persona for use in an ECA. While Lifenaut is still in its infancy, the thought is that eventual technologies could use each of these unique personality corpuses to animate a digital version of people. Hence, Mayer embraces the idea that corpus-based sources are a viable method of knowledge management.

Knowledge Manager Approach

From this survey of knowledge management systems, it can be seen that the input processing methods from the previous section often work well with certain types of knowledge organizations. This can be seen in the following pairings: pattern-matching with rule bases, expectation-matching with relational databases, and statistical categorization with question-response pairs. For this dissertation, the proposed input processing method of NLP-based keyphrase-matching will be paired with a contextually-organized corpus-based knowledge manager. The contextualization of the corpus data simply refers to the idea that related information is grouped together under a single heading, or context name. Moreover, related contexts may be grouped under yet another context name. The resulting effect is a contextually-layered data corpus. Deeper detail of this knowledge management method is addressed in the next chapter. The following section discusses the different mechanisms that can be implemented to deploy the information within the knowledge manager to serve as speech actions.

Agent Discourse Mechanism

Agent discourse refers to the method in which a dialog manager conducts its conversational actions. This can be referred to as the *speech action engine*. This component drives the agent's conversational actions in response to user utterances. In the previous sections on input processing and knowledge management, some of these speech action mechanisms were alluded to. In this section, three categories of discourse models are discussed in detail: rule-based, frame-based, and agent-based.

Rule-based

Rule-based discourse simply requires a valid input to execute a corresponding output, as dictated by a rule base. This reflects the action-reaction style of behavior modeling. Since the rule base exists as a huge collection of single actions, rule firing conflicts can arise. These can be managed by prioritizing the rules, as done in ELIZA (Weizenbaum, 1966), whose simultaneous firings are handled by creating a hierarchy of input pattern specificity. Figure 4 depicts the input-output flow of rule-based conversational behavior.

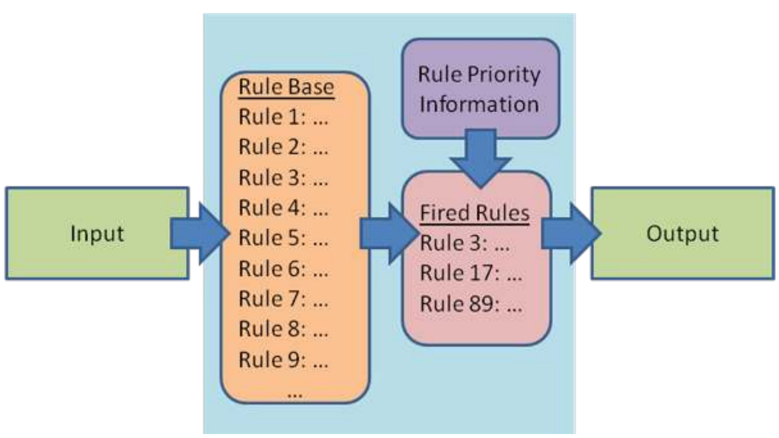


Figure 4: Rule-based discourse

Basic rule-based behavior has been widely used in building conversation agent discourses. All projects that utilize an ELIZA (Weizenbaum, 1966) derivative incorporate rules to produce speech actions. This research includes such applications as ECA prototyping (Thórisson, 1999) (Gockley et al, 2005), robotic humanoid platforms (Hoshino et al, 2005), language tutoring (Jia, 2003), QA (Rosset et al, 2006) (Quarteroni and Manandhar, 2007), and PDA-based guides (Santangelo et al, 2006).

There are two shortcomings of this basic rule-based discourse: 1) the rule base requires a tedious hand-coded effort to create, and 2) there is only a single layer of behavior. Specialized rule-based discourse models have been developed to overcome these issues. Statistical correlation methods address the manual modeling through ML-based automation. Finite-state machines have been developed to complexify the behavior of rule-based discourse agents. These are discussed next.

Statistical Correlation

Statistical correlation refers to the use of ML to drive speech action execution. Given a user input, statistical categorization uses a mathematically-derived relationship to pair the utterance with its output response. These question-response pairs are pre-processed from a training set, where the end result of this effort is an autonomously-derived rule base. The training sets for these ML routines can come from WoZ experimentation data (Artstein et al, 2009) (Kenny et al, 2009), or existing textual corpora (Colby, 1973) (Voth, 2006).

A drawback of this style of rule-based discourse is the burden of the machine to create natural language response from purely mathematical relationships. The ML process uses context-free calculations, meaning that the resulting output could end up with some potentially unnatural or awkward question-response pairings. Human intervention may be required to check the quality of the ML results.

Finite State Machine

A more complex version of the rule-based discourse method, the finite state machine (FSM) style models the agent's behaviors as a set of states. User input triggers the transitioning from one state to another. A specific behavior is executed for each state and input pair. This adds a dimension of deeper understanding, as the agent appears to have both cognition of the user's input, by processing the input utterance, as per usual, as well as a sense of directed purpose, through its internal state management.

FSM modeling has seen much action in the conversation agent design community, as it is a surefire method to create an autonomous entity with sufficient believable characteristics. Several researchers have sought this technology for developing agents, including Ahad et al's emotion-based Neva (2007), Augello et al's museum guide MAGA (2006), Babu et al's virtual receptionist Marve (2006), Bohus and Rudnicky's aircraft maintenance tool LARRI (2002), Microsoft's Peedy the Parrot (Ball, 1999), Lee et al's animatronic penguin Mel (2005), Bernsen et al's edutainment character H.C. Andersen (2004) and Niesler and Roux's hotel room booking agent DACST-AST (2001).

Two challenging aspects of FSM-based behavior modeling are: 1) out-of-bounds or irregular inputs must be accounted for at each state, especially when dealing with open dialog, and 2) hand-modeling of FSM's is a tedious and time-consuming effort. The first issue deals with how to handle user utterances that are either poorly received from the ASR system or are completely out of context with the current operational state. One way out would be to have the agent re-pose the question to possibly get the user to repeat their answer, in case the ASR picks up a valid input. This can prove to be ineffective if the agent must resort to this technique frequently, causing the entire conversation to be full of "Can you repeat that?" questions from the agent. Figure 5 gives a general diagram for an FSM-based discourse model.

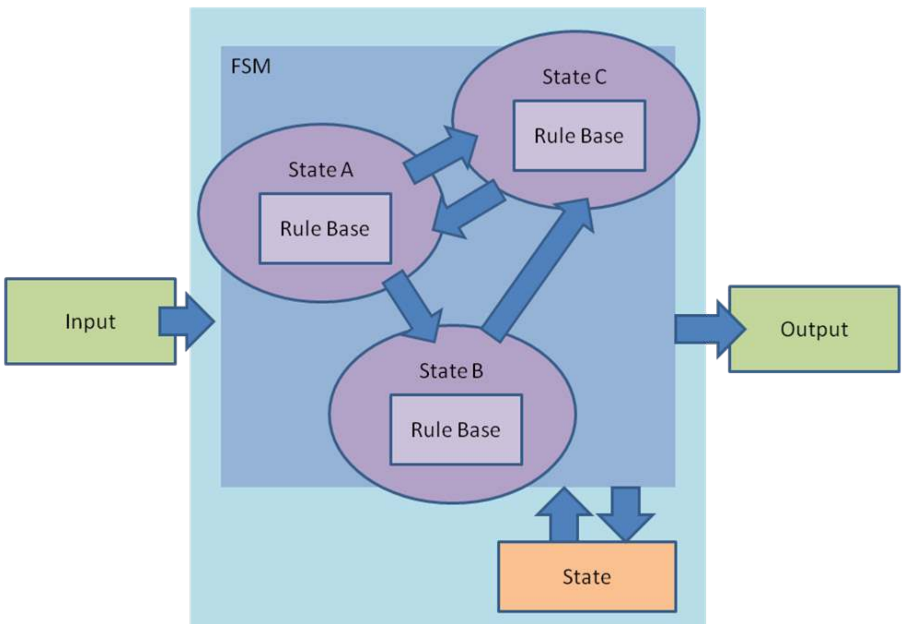


Figure 5: FSM-based discourse

As mentioned before, modeling FSM's for dialog behaviors is a burdensome effort in that the state transition graph and each individual state behavior for an agent must be developed by hand. Within each state behavior, there is possibly a set of sub-behaviors to define. Among these sub-behaviors, a knowledge management system, typically served by separate rule bases or pre-defined scripts, is defined. Needless to say, this front-end effort for discourse modeling demands an expensive cost in manpower. Nevertheless, FSM-based discourse can prove quite effective as an end result if each of the conversational behaviors required by the system is thoroughly defined.

Frame-based

Frame-based discourse lends itself to slot-filling behaviors, where the machine ends up playing a guessing game of what data the user would like to access. This discourse model can also be described as information state update, as mentioned in Lemon and Liu's work. (2006) The agent crafts a series of follow-up questions to ask in order to complete a proper query into a database. This persistent questioning behavior is the heart of frame-based behavior. Similar information-seeking conversational habits can be seen in expert systems and case-based reasoning software. This style of discourse takes aim solely at the user's initiative for attaining information. The agent simply exists to retrieve data from its knowledge base using a carefully planned exchange of words. A similar effect can be found in GUI-based "Wizard" interactions. In these exchanges, the machine queries the user for certain data points in order to finish a prescribed task. These data points serve as the conversational goals, although a traditional two-sided conversation is not actually being executed.

As mentioned earlier in this chapter, researchers have produced many frame-based models. Much of this work is geared toward immediate use in production systems, as slot-filling works well for data query-based applications. These include travel information (Larsson and Traum, 2000) (Skantze et al, 2006), ticket booking (Andernach et al, 1995), library cataloging (Cenek, 2001) and inventory checking (Owda et al, 2007).

This purely assistive style of discourse exemplifies the main weakness of frame-based systems – that only the user's needs are important. The effect of this downside is that the machine tends to sound like a data servant rather than a conversational peer. The constant pursuit of filling in query arguments through follow-up questions causes the overall interaction to feel like an interrogation session. What is missing from this picture is a mixed-initiative kind of discourse, where even the agent itself has its own agenda of where the conversation should go. The next section discusses the agent-based discourse model, which incorporates this concept.

Agent-based

An agent-based discourse method can be seen as an agent-driven version of the FSM model. The main conversational drive is not the input itself, but rather how the user utterance affects the agent's own goals. This style of behavior can be classified under the BDI model.

A key characteristic of agent-based discourse is the use of mixed-initiative, where both the user and the machine can contribute speech actions that drive each party toward their own respective dialog goals. A conversation can be modeled as a set of verbal actions between two (or more) participants to move toward an end goal state. This particular state may or may not be pre-determined, as some conversations are often open-ended. Regardless, the idea is that a conversation has a start state, and the actions between the two interlocutors' leads to the final goal state.

What makes agent-based discourse modeling especially compelling is this delegating of the chatbot with its own agenda for not only helping its user, but also how the user can help the chatbot. This two-sided model of conversation gives way for a more natural exchange of words, and it also gives the user the sense that s/he is talking to another being that also has a vested interest in the interaction. Some examples of agent-based discourse models include Max the museum guide (Kopp et al, 2005), Eugene the cuttlefish (Wallis, 2005), Smart Personal Assistant (Wobcke et al, 2005), and VIRMA the insurance risk analyst (L'Abbate et al, 2005)

The above section runs through the main styles of discourse models in which a conversation agent can undertake. Rule-based agents were characterized as simple action-reaction machines that could be improved by using ML-based methods to automate the rule production and by using FSM designs to add depth to the conversational behavior set. Frame-based discourse sought the use of a database back-end to drive agent responses. The main idea behind this mechanism is to have the user provide enough description of her/his information-seeking needs such that a proper database query can be executed. This method is based solely on user-initiative, which tends to give a lopsided conversational exchange as the agent ends up spending much of its asking follow-up questions. In contrast to these user-centric systems, agent-based conversational models were presented, whose mixed-initiative style of discourse allows for a more balanced, two-sided dialog.

Discourse Model Approach

In this dissertation, a CxBR discourse model is presented. This design works in conjunction with the NLP-based keyphrase-matching input processing and the contextually-organized corpus-based knowledge management system. The key attributes considered when developing this discourse model were:

- Tolerance to error-prone ASR inputs: Robustness to ASR failure is addressed by driving the input processing system toward identifying keyphrases extracted from the data in the knowledge manager. This extraction is performed using NLP-based POS tagging and phrase chunking techniques. The contextually-driven discourse engine prepares the dialog manager as to which keyphrases should be expected.
- Open dialog adaptability: Open dialog is handled by the combined efforts of all three dialog manager sub-systems. The input processor's keyphrase detection is readily available to extract any important concept-bearing words from the user utterance. The knowledge manager's contextually-layered organization allows for focused, bottom-up domain matching. The context-based discourse engine provides a flexible conversation model that can both provide depth for known user assistive goals, as well as recover from detected unknown domain concepts.
- Domain independence for immediate flexibility of agent expertise: Domain independence is maintained by separating the knowledge management data from the rest of the dialog manager. The data kept in the knowledge manager should be interchangeable in that the input processor and the discourse mechanism will not need modification if the topical information source is altered. This allows for a quick turnover rate for developing expert agents for different domains or for augmenting existing topical information.
- Minimal conversational awkwardness: Minimizing conversational awkwardness is primarily the effort of the context-based discourse model. This sub-system maintains a mixed-initiative conversation policy to convey to the user that the agent itself has its own goal-directed agenda. Additionally, the discourse model contains specific

conversational interaction behavior rules to keep up the illusion of naturalness and minimize awkwardness. These rules pertain to certain social graces of dialog, such as reasonable repetitiveness, timing expectations, and barge-in reactions.

- Maximum fulfillment of assistive goals: Assistive goal fulfillment is also the work of the discourse engine. It is the role of the dialog manager to not only appear natural, but also assist the user in accomplishing her/his tasks that need to be addressed. The discourse model's CxBR paradigm lends itself to providing a system of both detecting and servicing user goals.

The remainder of this chapter is devoted to describing the entire CxBR-based dialog management system that is the main contribution of this research. Before continuing with this discussion, however, a briefing on the general CxBR architecture is provided. A description of each of the major components involved in the paradigm is presented. To begin, the CxBR architecture consists of *Contexts*, *Context-Transition Logic*, *Missions*, and the *Agent Interface*.

Contexts

The state of an agent's environment, in both internal and external terms, makes up its behavioral *context*. In CxBR, the full set of contexts makes up the agent's entire behavioral repertoire. This implies that every state that is encountered is accounted for with a context. A major context contains all of the functionalities needed to allow the agent to successfully manage the situation associated with the active context. The *active* major context controls the actions of the agent. Only one major context can be active at any one time. A *default* major context is provided when an unknown set of environmental circumstances occurs and the agent does not have explicit direction on how to behave. The default context remains the primary operation mode until a recognizable state is reached, in which case another specific context is *activated*. Additionally, *sub-contexts* may exist within major contexts. These nested components allow an agent to perform smaller behavioral units to accomplish sub-goals within a context's main goal.

Context-Transition Logic

Context-Transition Logic manages the sequential activation of contexts. This selection process is dictated by a set of *transition rules*, which continuously monitor environmental signals to determine if a context switch is in order. In tactical situations, an overriding set of rules, called *universal transition rules*, may intervene at any time to force the agent into another context. These rules often come into play when an agent's self-preservation is in question. A CxBR framework maintains the entire knowledge base for the CxBR agent.

For robustness, the set of all transition rules should suggest a single context activation given any set of environmental states. The practice of automatically selecting a single context from a set of multiple viable context suggestions is known as a *competing context* framework. (Saeki and Gonzalez, 2000) In current CxBR practices, it is more favorable to opt for hardcoded single context activation for the sake of simplicity.

Missions

Incorporating contexts and context-transition logic, a mission represents the highest level of agent description. Missions do not directly affect behavior. A *mission goal* dictates the final objective that the agent is to pursue. Mission *constraints* exist to define the limitations that the agent behavior must obey. *Context topology* refers to the graph of contexts that defines the (limited) feasibility of transitions among the various contexts. The contexts serve as nodes, with context-transitions represented as directional edges. Overall, a mission is defined by the collection of the mission goal, mission constraints, and the context topology.

Agent Interface

The *agent interface* exists as the link to the world that exists outside of the CxBR agent. For this dissertation, the speech-based dialog system serves as the interface between the agent and the external environment. The CxBR agent resides as the conversational entity in this system, whose behavior is defined by the natural language responses given back to the human user. Figure 6 exhibits the block diagram for a generalized CxBR architecture. (Gonzalez et al, 2008)

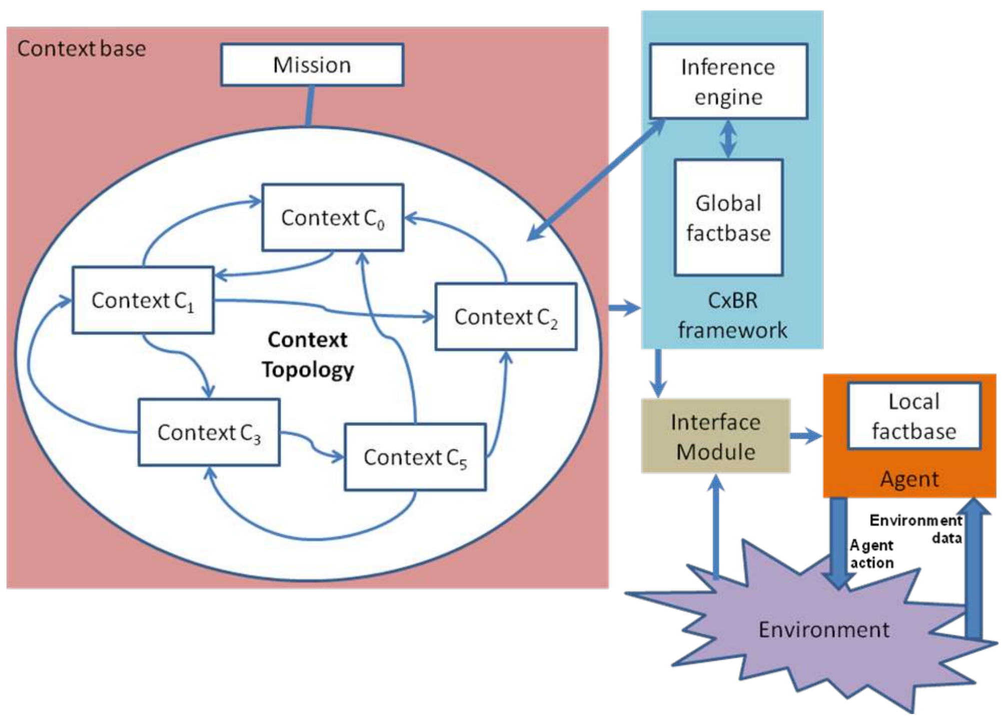


Figure 6: General CxBR architecture block diagram (Gonzalez et al, 2008)

Context-Based Reasoning Dialog Manager

This next section describes the inner-workings of the CxBR dialog manager, which consists of a keyphrase parsing input processor, a corpus-based knowledge manager, and a context-driven discourse model. The underlying theme for this dialog manager is the supervision of conversation *goals*. In essence, a spoken dialog between parties is a sequence of passing goal-oriented statements to one another. (Grice, 1975) The intent is to achieve some form of resolve for each of these exchanges, otherwise viewed as completing goals. For this dissertation, the dialog manager is charged with managing these goal completion tasks.

In terms of CxBR, these conversational goals can be equated to contexts. Hence, detection of goals is performed in the inference engine. Servicing of goals is the work of traversing the context topology until the mission is completed. The rest of this chapter describes the conceptual aspects of goal management. The next section provides a general description of a goal management system. This is followed by a discussion of the role of contexts within goal management. A third section presents the relationship between knowledge and goals.

Goal Management

Goal management in a dialog system comprises the processes that recognize and satisfy the interlocutor's needs as conveyed by her/his utterances. This section explains why goal management is important in a conversation and how it enhances the naturalness of an HCI session.

Within any conversation, regardless of the presence of machine agents, there exists some sense of goal-oriented activity on the part of all the participants. (Grice, 1975) Often, these activities are characterized as some form of knowledge transfer, such as requesting or delivering information from or to an expert. (Isaacs and Clark, 1987) Every participant contributes utterances, or speech acts, to drive the conversation toward purposefulness. In a two-party conversation, both sides go into the conversation with the intention of getting something out of the interaction. The participants begin talking to one another in an initial state, only to end in a different state – a goal state. This model of conversation assumes that its conclusion occurs when both participants are satisfied with how much they have achieved from the session. Hence, under normal conditions, the goals of both speakers are accomplished when their conversation ends.

This same model may be applied to the interaction between an assistive chatbot and a human user. The chatbot's primary goal as an assistive entity is to satisfy the user's needs. The human's goals, on the other hand, are simply the tasks that s/he wants to accomplish from talking to the chatbot. Unbeknownst to the chatbot, the human's goals could be any number of things. The only way the chatbot can determine the user's intentions is to infer them from oral interactions. This is the essence of the *goal management* concept - the idea that an assistive dialog system understands a user's needs through the use of conversation management.

The aim of providing a goal management system is to offer a general approach to creating the effect of a natural, open dialog HCI experience. Open dialog refers to a loose set of conversational input constraints, allowing the agent to handle a wide range of user utterances. Additionally, one or more user goals can exist at any time during an open dialog interaction. This contrasts with the closed, highly-constrained and unnatural multiple choice-style of input expectation found in automated airline booking agents and telephone-based credit card payment systems. Moreover, these types of interactions can only accommodate one user task at a time. The open dialog style allows for a more natural flow of the conversation. To realistically accomplish the illusion of open dialog through goal management, the following assumptions must exist:

- The dialog system is limited to an expert domain, and the user is cognizant of the dialog system's functionality as an expert entity. This constrains the user to a topical *context* with which the chatbot is deeply familiar, without jeopardizing the open dialog style.
- The user's goals are limited to those related to the chatbot's expertise. This assumption dictates that the user understands the agent's limitations as a domain-specific entity.

This section has established the role of goal management in a conversation and it has discussed its use as a means to achieve a more natural feel to human-computer exchanges. Two assumptions were made to frame the goal management problem into a manageable approach. These assumptions pertain to the notion of contexts and the importance of a constrained knowledge base, respectively. The next two sections describe how each of these issues affects goal management in a dialog system.

Contexts In Goal Management

This section discusses the role of contexts in a conversational goal management system. Contexts directly correlate to reaching specific goals, and natural oral interactions often transition through a number of conversational goals. These two ideas suggest the notion that conversational goal management can benefit from a CxBR approach.

The dynamics of conversation has been studied in the psycholinguistics discipline. Urbanová (2001) makes note of the fact that a dialog between two people involves an extra layer of implicit knowledge. Garrod and Pickering (2004) back up this argument, asserting that a monologue requires more explicit explanations than a dialog, where both parties are living, breathing bodies of knowledge. Hence, this agreement of well-versed knowledge does not need to be presented in a conversation, a notion that lends itself to domain-specific interactions. (Clark and Marshall, 1978) Urbanová (2001) makes this claim to show that interpreting another interlocutor's communicative channel does always require verbalized mechanisms, since common implicit knowledge between parties can help to fill in semantic blanks during a conversation.

In this dissertation, goal management is achieved using a context-based approach. Edmondson's (1999) work establishes the role of contextualization in a human-computer interaction. A *context* refers to a particular situation that is dictated by the configuration of internal and external circumstances. (Gonzalez and Ahlers, 1998) (Stensrud et al, 2004) (Gonzalez et al, 2008) For an oral conversation, these circumstances refer to the internal state of the conversation agent and the state of the human user. For every context, there is an associated goal and a group of relevant actions that are executed to achieve this goal. A *goal* is defined as an end state that an agent desires to reach.

It is imperative that a dialog system be able to properly manage conversation goals, as the user can have multiple goals and s/he may introduce new goals at any time. Henceforth, the system must be able to service many goals at one time, as well as be prepared to take on more goals, unannounced. This necessity to be able to jump between different goals in real-time lends itself to the CxBR architecture. CxBR agents provide responses that are directly related to its active context. The fact that contexts correspond to accomplishing particular goals combined with the idea that conversational goals take on a very fluid nature yields the assertion that goal management can be facilitated using CxBR methods.

One aspect of CxBR is its dependence on knowledge to make inferences about the state of the environment, especially when a context to be activated needs to be selected. A proper knowledge management framework must be present for a CxBR agent to appropriately decide its response outputs. This stipulation is no different for the CxBR-based dialog system presented here. The requirement of knowledge for context recognition is extended to the process of goal management. The next section establishes the importance of knowledge in goal management.

Knowledge In Goal Management

A major facet of goal management is goal recognition, which is the process of identifying which conversational goals need to be addressed. In this dissertation, a context-driven inference engine performs this service, as dictated by the CxBR architecture. Thus, a strong knowledge base must be in place for proper goal recognition. The information found in this knowledge base is analogous to the rote knowledge that a human learns and manipulates to make decisions. In this dissertation, three knowledge models are considered: domain-specific knowledge, conversational knowledge, and user-profile knowledge. The approaches to these models are presented in the remainder of this section.

Domain-Specific Knowledge

The scope and depth of the domain-specific knowledge is modeled after that of a traditional expert system (Gonzalez and Dankel, 1993), where a domain specialist meticulously adds information to a machine by hand. In most expert systems, knowledge exists as a set of if-then statements to make decisions. For the purposes of this work, the domain-specific knowledge is organized as a semantic network. Such a paradigm for knowledge representation lends itself to language applications, as seen in the work of KL-ONE (Forrest, 1991) and ConceptNet (Liu and Singh, 2004). In short, the knowledge contained in this source corresponds to the assumption that the dialog system specializes on a particular domain expertise. For the sake of goal management, this knowledge base provides the information needed to recognize domain-specific goals.

Conversational Knowledge

Alongside the expert knowledge, the agent's basic conversational speech actions must be maintained to serve as mediators between rote knowledge deployments. These particular responses can also be tailored to reflect the unique personality of an agent. Examples of conversational knowledge includes such cues as querying ("What would you like to know about?"), greeting ("Good morning"), and clarifying ("Can you repeat that?"). The entirety of this knowledge base encompasses all of the agent quips that do not reflect any expertise, but rather serve as transitioning actions. In terms of goal management, conversational knowledge provides domain-independent dialog mediation elements to aid in goal recognition and goal fulfillment.

User-Profile Knowledge

A third source of knowledge is a user profile database. MacWhinney et al (1982) describe the importance of memory during a conversation. They claim that memory structure is key when dealing with natural dialogs, as it provides an extra layer of interactive immersion. In this dissertation, all that the agent knows about the human with whom it is communicating exists in the user-profile knowledge. Once the user has identified him or herself, the knowledge manager can immediately retrieve her/his individual profile. This is particularly important for the sake of providing an HCI experience that escalates the level of realism and conveys an effect of personalization.

The user-profile knowledge also serves as a repository of individualized account data. The information contained in this source can serve to drive the agent's conversational actions. Specifically, any missing data points within a user's profile can trigger the agent to pose questions to retrieve this knowledge, a technique known as slot-filling (McTear, 2002). Goal management benefits from this knowledge base because it can use slot-filling to service a particular user goal.

Contextualized Knowledge

Contextualized knowledge refers to a cross-section of all three knowledge sources that is relevant for the active context of the conversation. Each piece of information within the knowledge manager is annotated with a context tag. Once the dialog system determines the context of the conversation, knowledge that is labeled with the current context is elicited as valid information for the conversation and funneled into the contextualized knowledge base. Once this subset of information is established, the dialog manager can then work with a manageable portion of the entire knowledge base. This is especially true when performing goal management, which may require memory-intensive processes.

The concept of contextualized knowledge is a novel feature of this dissertation. The idea that only a portion of an agent's entire knowledge is needed at any given time reflects how a human does not require processing every single fact s/he knows to make decisions. A CxBR-based architecture lends itself to this concept, since the determination of an active context, and therefore an active set of contextualized knowledge, is a built-in function of CxBR.

This section established the connection between a robust knowledge backbone and goal management routines. The next section describes a general framework for providing a method of goal management in a dialog system.

Framework For Goal Management

Goal management in a dialog system involves three parts: 1) goal recognition, 2) goal bookkeeping, and 3) context topology. Goal recognition refers to the process of analyzing user input utterances to determine the proper conversational goal that is to be addressed. This is analogous to the context activation process in CxBR methods. Goal bookkeeping deals with organizing the identified goals, and then servicing the recognized goals in the order they are received, using a stack. Context topology refers to the entire set of speech acts of the conversation agent. This structure also includes the transitional actions when moving between contexts when a goal shift is detected. The context topology carries out the responses needed to clear out the goal bookkeeping stack. The next sections further describe each of these goal management parts.

Goal Recognition

Goal recognition is accomplished using linguistic analysis of each user utterance. This is similar to the inference engine found in CxBR systems (Gonzalez and Ahlers, 1998) (Stensrud et al, 2004) (Gonzalez et al, 2008), where the transition rules determine the active context according to the state of the environment. The difference with the goal recognizer, however, is that the context is resolved using keyphrases that are extracted from a parts-of-speech parsing of input responses. With the aid of the knowledge base described previously, the user utterance is interpreted, and the context associated with this understanding is identified and activated.

Goal Bookkeeping

Goal bookkeeping describes the process of servicing every identified goal that is presented to the agent. Immediately after recognizing a goal, it is pushed on to the goal stack. The goal stack is modeled after Branting et al's (2004) discourse goal stack model (DGSM). The original DGSM only supported minor detours from the conversation path. In this dissertation, more complex interruptions may occur, such as switching to entirely different contexts. Thus, the DGSM was modified to be able to handle conversation paths that experience entire paradigm shifts between context changes.

Context Topology

The context topology maintains the interactive structure of the contexts, which carry out the actual, executed actions of the dialog system. This interaction is controlled by the goal bookkeeping component. Upon receiving the activated goal to be addressed from the goal stack, the context topology operates on this information to provide the proper agent response. Each context within the context topology corresponds to a certain conversational task, whether user motivated (external) or agent motivated (internal). Most of these conversational tasks adhere to a specific *user* task goal. These are known as User Goal-Centered Contexts. The remaining conversational tasks constitute the Agent Goal-Driven Contexts. The inclusion of all User Goal-Centered Contexts and Agent Goal-Driven Contexts constitutes the entire ECA context topology.

Agent Goal-Driven Contexts

Agent Goal-Driven Contexts exist to serve the most basic conversation functions. These contexts include: greeting, closing, clarification follow-up, and non-specific initiative. Greeting and closing simply refer to the “Hello” and “Good-bye” tasks. The goal of the clarification follow-up context is to verify the semantic intent of the last user utterance. Non-specific initiative is used when the dialog system recognizes no user task goals, and thus, must contribute some conversational small talk until a user goal is again established. This may be viewed as the default context in CxBR terms. Essentially, the Agent Goal-Driven Contexts exist to fill in the gaps when the user has not taken any initiative to convey what s/he wants to do.

User Goal-Driven Contexts

User Goal-Driven Contexts are used for domain-specific tasks that the user wants to accomplish. Each of these contexts alone can operate independently as a single-goal conversation agent. The two typical user goals are information retrieval and assistive analysis. Information retrieval refers to a user asking the agent to fetch a piece of data from an expert knowledge base. This is similar to the work done by QA agents. (Schumaker et al, 2007) Assistive analysis describes the process where the user gives a series of related data points and the agent must give an expert conclusion or analysis of this information. This type of functionality is typically found in expert systems (Gonzalez and Dankel, 1993) and CCBR agents (Aha et al, 2005). The main point here is that the conversation agent in this dissertation can support a variety of user goals in different modalities, a practice normally performed by separate systems.

Chapter Summary

This chapter described the three primary design decisions in building a dialog manager: input processing method, knowledge management, and discourse model. Different choices for each sub-system were presented, and particular component selections were highlighted and justified for purposes of this dissertation. A system of goal management for dialog systems was presented in light of these design decisions. The next chapter discusses a prototype of a CxBR-based dialog manager with a description of its inner-workings. Figure 7 shows the overall design of the dialog manager.

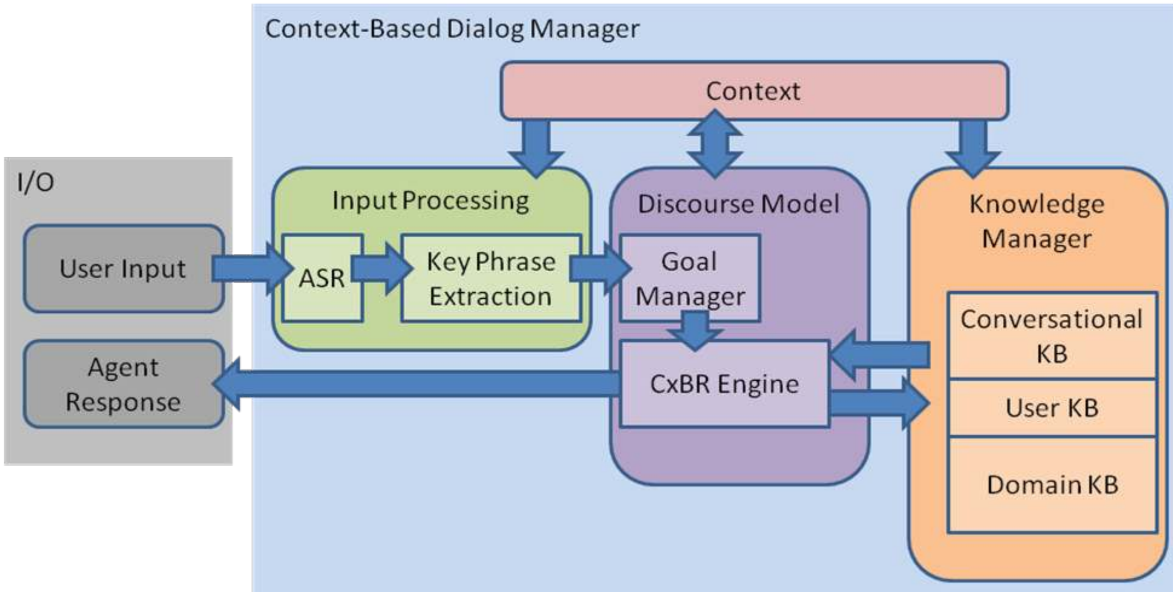


Figure 7: CxBR-based dialog manager design

CHAPTER FIVE: PROTOTYPE DEVELOPMENT

The methods involved in this dissertation which were conceptually described in the previous chapter are incorporated and tested in a prototypical platform known as the CONtext-centric Corpus-based Utterance Robustness (CONCUR) dialog manager. This prototype provides a fully functional dialog-based agent that incorporates the conceptual contributions put forth in this dissertation, demonstrating the viability of CxBR applied to the natural language domain in the role of managing a dialog for an ECA. In this chapter, a framework prototype that embodies the contributed approach is described, providing insight into the development of the prototype, framed in an application-specific manner. Implementation details for the CONCUR system are provided in this chapter

Prototype Overview

The previous chapter discussed the different approaches in building a conversation agent. Under the problem constraints of open dialog, ASR limitations and domain-independent knowledge management, certain design choices were made with regard to this analysis. These decisions were incorporated in CONCUR, which consists of three major components: the Input Processor, the Knowledge Manager, and the CxBR-based Discourse Model. The Input Processor serves as a listening comprehension filter, where utterances are converted to contextually-relevant content to be processed by the agent. The Knowledge Manager acts as an agent's rote memory. Finally, the Discourse Model poses as the brain's cognitive mechanism that purposefully drives a conversation. Figure 8 depicts a top-level block diagram of the overall CONCUR architecture. This is a more detailed version of Figure 7, from the previous chapter, with the specific major components (Input Processor, Knowledge Manager, and Discourse Model) drawn out and high-level information flow annotated. The minor components, the Interpreter and the Generator, act as the agent's "ear" and "voice," respectively. They are simply mechanical interaction devices necessary for a fully functioning virtual agent setup. In this case, the setup is the Project LifeLike Avatar. (DeMara et al, 2008) The next section contains a brief description of this agent platform. The remaining major components are described in further detail in the remainder of this chapter.

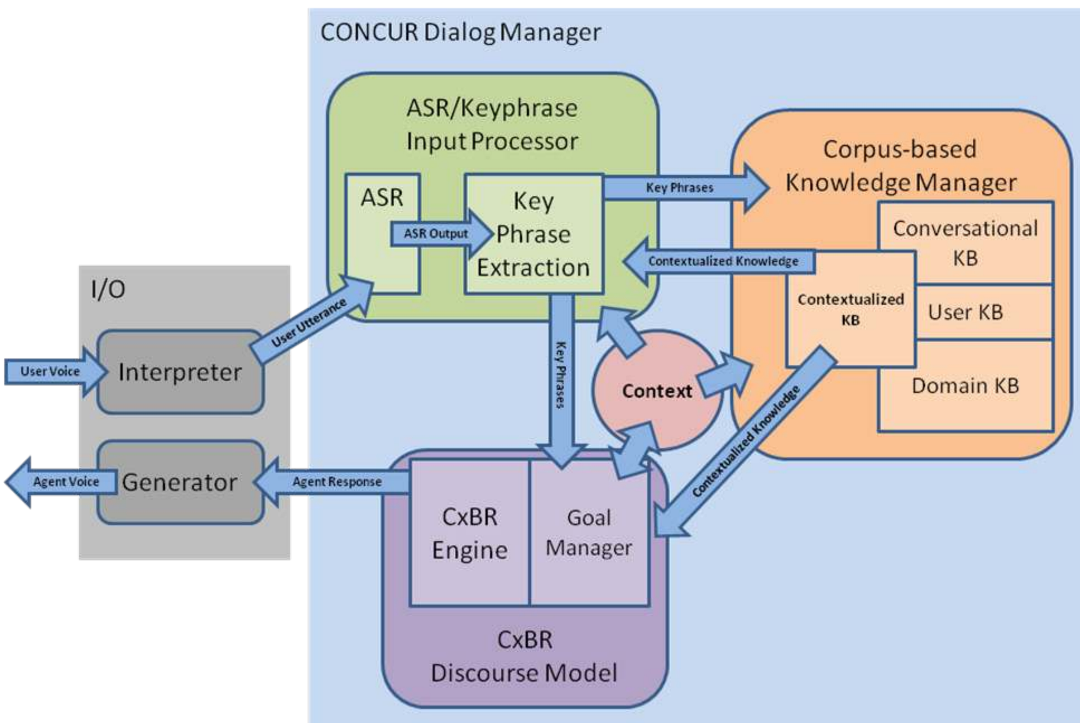


Figure 8: Basic CONCUR block diagram

Project LifeLike Avatar

The Project LifeLike Avatar is a research effort seeking to create a virtual double of a specific human, in this case, the Program Director of the National Science Foundation's (NSF) Industry/University Cooperative Research Center (I/UCRC), Dr. Alex Schwarzkopf. (DeMara et al, 2008) This agent exists as an ECA whose animated upper torso and expressive face exist on a 52” computer monitor. A microphone “ear,” or Interpreter, picks up the user utterances, and a speaker serves as a voice box, or Generator. Three separate pieces of software make up Project LifeLike: the agent externals, the speech recognizer, and the dialog manager.

Agent externals refer to physical sensory interfaces at the user level. These modalities include the 2-dimensional image of the avatar, with 3-dimensional movement capabilities, and the voice generation of text-based messages delivered from the dialog manager speech engine. The speech recognizer begins with the microphone input and concludes with an ASR result string, which is passed onto the CONCUR. Finally, the CONCUR dialog manager serves as the processing facility of the ASR string to produce a response string to the agent externals. Figure 9 gives a Project LifeLike Avatar block diagram. The aim is to use this avatar platform for integration with the CONCUR dialog manager in a proof-of-concept system for demonstration and experimentation purposes in the realm of speech-based ECA design.

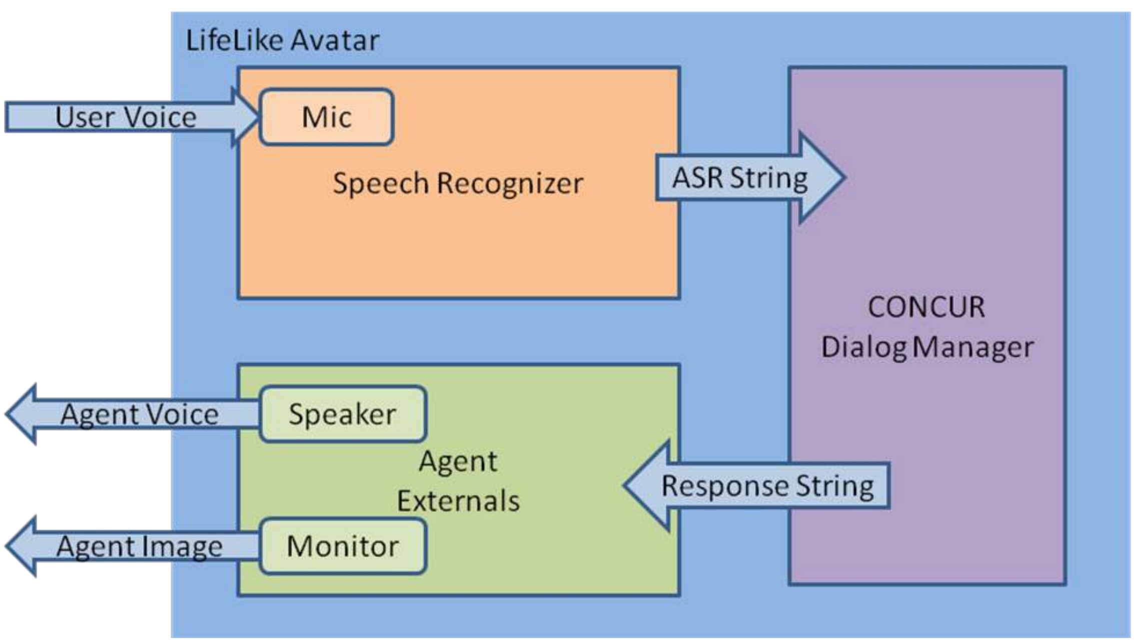


Figure 9: Project LifeLike Avatar block diagram

CONCUR Chatbot

A text-based CONCUR platform for testing purposes was also devised to detach the dialog manager from its physical embodiment (both visually and aurally) for testing purposes. This separation from the Project LifeLike Avatar eliminates any ASR-related input errors, since a speech recognizer is not employed. Developed using the Jabber chat protocol, now known as the Extensible Messaging and Presence Protocol, users could communicate with the CONCUR Chatbot using a Google Chat client. Figure 10 depicts a high-level diagram of this keyboard-only conversation agent. The sections following this figure continue to describe the inner-workings of the CONCUR sub-systems.

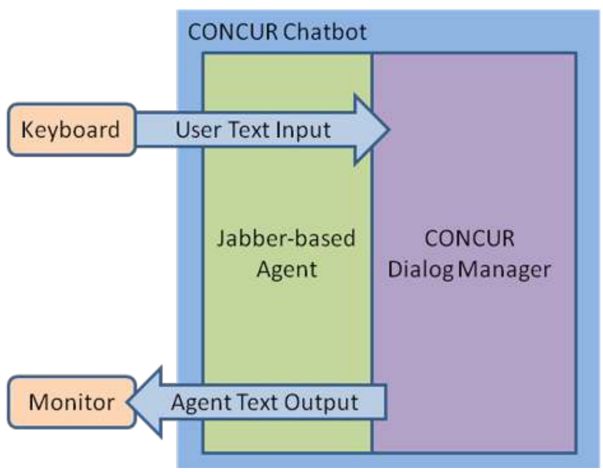


Figure 10: CONCUR Chatbot block diagram

Input Processor

The first component of CONCUR is the Input Processor. This sub-system parses the raw speech-based user input for contextual keyphrases. The resulting utterance picked up by a speech recognizer is chunked into phrases, which also includes a word-for-word POS tagging. These procedures are performed using an NLP toolkit. The utterance phrase chunks are then filtered for noun and verb phrases, discarding the remainder of the sentence. For example, if the speech recognizer detects the words “I am interested in supplemental funding,” the Input Processor would identify the verb phrase “interested,” and the noun phrase “supplemental funding.”

The functional constructs of the Input Processor also contribute to annotating the domain expertise corpus with keyphrase indices. As part of the corpus pre-processing routine, an NLP treatment of the domain corpus is performed, providing an automatically generated keyphrase list for each contextual layer. It must be noted that for the purposes of this dissertation, ASR technology is not part of this investigation, but rather, it is treated as a support utility for the CONCUR infrastructure. One aspect of the ASR facilities, however, is the use of keyphrase constraints to assist in speech recognition disambiguation. The pre-processed set of keyphrases, extracted from the knowledge base corpus, is fed into the speech recognizer. By adding this set of speech recognition constraints, the agent has a better chance at identifying contextually relevant utterances. Figure 11 depicts the Input Processor block diagram.

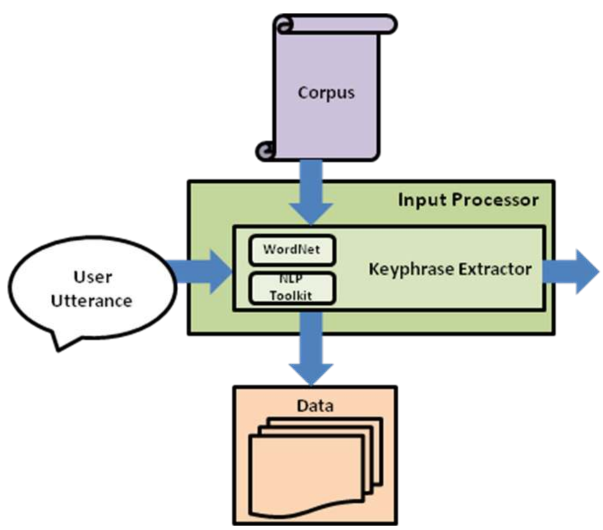


Figure 11: Input Processor block diagram

Table 8 and Table 9 depict the member components and member functions, respectively, of the UserResponse data object. The Input Processor transforms a raw user utterance string into this container. The *FullSentence* variable stores the unaltered input string, while *KeyPhrases* is a list of the verb phrases and noun phrases identified from original utterance. The member function *ComputeKeyPhrases* uses NLP chunking to populate the list of strings in *KeyPhrases*. The *Matches* function compares the current *KeyPhrases* list with that of a *MicroContext* object, to be described in the next section, and returns the number of matching phrase chunks from each list.

Table 8: UserResponse class member components

Component	Data Type	Description
FullSentence	String	Entire string of user utterance
KeyPhrases	List of Strings	List of key phrases chunks

Table 9: UserResponse class member functions

Function	Return Data Type	Description
ComputeKeyPhrases	List of Strings	Perform key phrase chunking to populate KeyPhrases
Matches	Int	Comparator between a UserResponse and a MicroContext

Table 10 and Table 11 describe the *MicroContext* class. The Input Processor uses this data structure to store each individual corpus sentence. *MicroContext* is organized almost identically to the *UserResponse* class. *FullSentence* variable stores the raw corpus string. *KeyPhrases* is a list of the verb phrases and noun phrases identified from this string. The function *ComputeKeyPhrases* populates *KeyPhrases* with a list of NLP chunking strings. The *Name* member component keeps track of the data corpus item's contextual depth through a file system directory-style naming convention. The next section on contextual knowledge management describes this *MicroContext* naming organization.

Table 10: MicroContext class member components

Component	Data Type	Description
FullSentence	String	Entire string of user utterance
Name	String	Context name
KeyPhrases	List of Strings	List of key phrases chunks

Table 11: MicroContext class member functions

Function	Return Data Type	Description
ComputeKeyPhrases	List of Strings	Perform key phrase chunking

The Input Processor's two primary functions are to contextually prepare the knowledge corpus files and to process the user utterance after it has passed through the ASR system. The knowledge file preparation algorithm is shown in Figure 12. The user utterance processing algorithm is displayed in Figure 13.

1. Given corpus file *P* with context set *X*
2. Open *P*
3. Declare KnowledgeBase *B*
4. For each context $c \in X$
    a. For each FullSentence $s \in c$
        i. Perform NLP chunking on *s*, giving keyphrase set *K*
        ii. Store *s*, *K*, *c* in MicroContext *m*
    b. Add *m* to *B*
5. Close *P*

Figure 12: Knowledge File Preparation Algorithm

1. Given user utterance *u*
2. Perform NLP chunking on *u*, giving keyphrase set *K*
3. Store *u*, *K* in UserResponse container *U*

Figure 13: User Utterance Processing Algorithm

Knowledge Manager

A major feature of CONCUR remains its dependency on *contextual relevance*. This concept speaks to the idea that two data points may be within contextual proximity of each other if they share some form of conceptual commonality. Hence, *contextualization* requires a pre-defined set of related data. For dialog-based systems, contextualization exists when groups of words maintain conceptual relationships with each other. These lingual relationships are contained in the Knowledge Manager portion of the CONCUR architecture.

Knowledge bases used by CONCUR all reflect a pre-established contextual relationship mapping. This is done by using a contextual layering system of organizing information, a format similar to that of an outline or an encyclopedia entry. Hence, all responses that will be said by the conversation agent are pre-annotated in the knowledge base with a contextually-driven naming system derived from its outline depth, and each of these agent responses is stored individually as a *MicroContext* data object. As an example, the *Name* component in a *MicroContext* that corresponds to the corpus item, “The deadline for the proposal is July 5,” is populated with the context label “Planning grant proposal.Deadline.” This *Name* denotes the idea that the response statement pertains to the immediate context “Deadline,” and a *super*-context called “Planning grant proposal.” Another context, “Planning grant meeting,” could also fall under this same super-context, yielding the tag “Planning grant proposal.Planning grant meeting.” In this case, both “Deadline” and “Planning grant meeting” are contextually relevant items under the “Planning grant proposal” super-context. Table 12 and Table 13lists the member components and member functions in what is effectively a collection of *MicroContexts*, known as a *KnowledgeBase* data object.

Table 12: KnowledgeBase class member components

Component	Data Type	Description
MicroContexts	List of MicroContexts	All MicroContexts involved in KnowledgeBase
ContextNameList	List of Strings	List of all contexts involved in KnowledgeBase

Table 13: KnowledgeBase class member functions

Function	Return Data Type	Description
AddKnowledgeFromFile	Void	Load corpus file into KnowledgeBase format
ComputeContextNameList	Void	Populate ContextNameList with all context names
GetContextualizedKnowledgeBase	KnowledgeBase	Return KnowledgeBase pertaining to a certain context
IdentifyContext	List of Strings	Determine the active context from a user utterance
GetParent	String	Return name of parent context
GetChildren	List of Strings	Return names of all children contexts
GetSiblings	List of Strings	Return names of all sibling contexts

MicroContexts maintains a list of all individual *MicroContexts* for a single data corpus, and *ContextNameList* is a list of all unique *MicroContext* names, as well as any higher-level super-context names. This list is computed using the member function *ComputeContextNameList*. *AddKnowledgeFromFile* receives a corpus file name and populates a *KnowledgeBase* object with the help of the Input Processor. *GetContextualizedKnowledgeBase* returns a contextualized subset *KnowledgeBase* from the *MicroContexts* list. This function executes the idea that only a fraction of the agent's total knowledge is necessary to perform an immediate knowledge management activity. Contextualized knowledge bases are discussed later in this section. *IdentifyContext* aids in current context identification process found in the Discourse Model's Inference Engine, which will be discusses in later section. This function uses a series of context key phrases matching comparisons, using the *MicroContext* class' *Matches* member function, to provide CONCUR's best estimate of a user's contextual intent as a list of candidate context names. *GetParent*, *GetChildren*, and *GetSiblings* return the context name(s) for a certain context's root (parent), next levels (children), or laterally equivalent levels (siblings). These functions are used when CONCUR attempts to locate any spatially relevant context names during discourse navigation, which will be discussed later. The next section describes the three types of corpora that are transformed into *KnowledgeBase* data objects.

The Knowledge Manager consists of three sources of data: user data, conversational knowledge, and domain-specific knowledge. This section describes each of these knowledge bodies. Figure 14 depicts the Knowledge Manager block diagram.

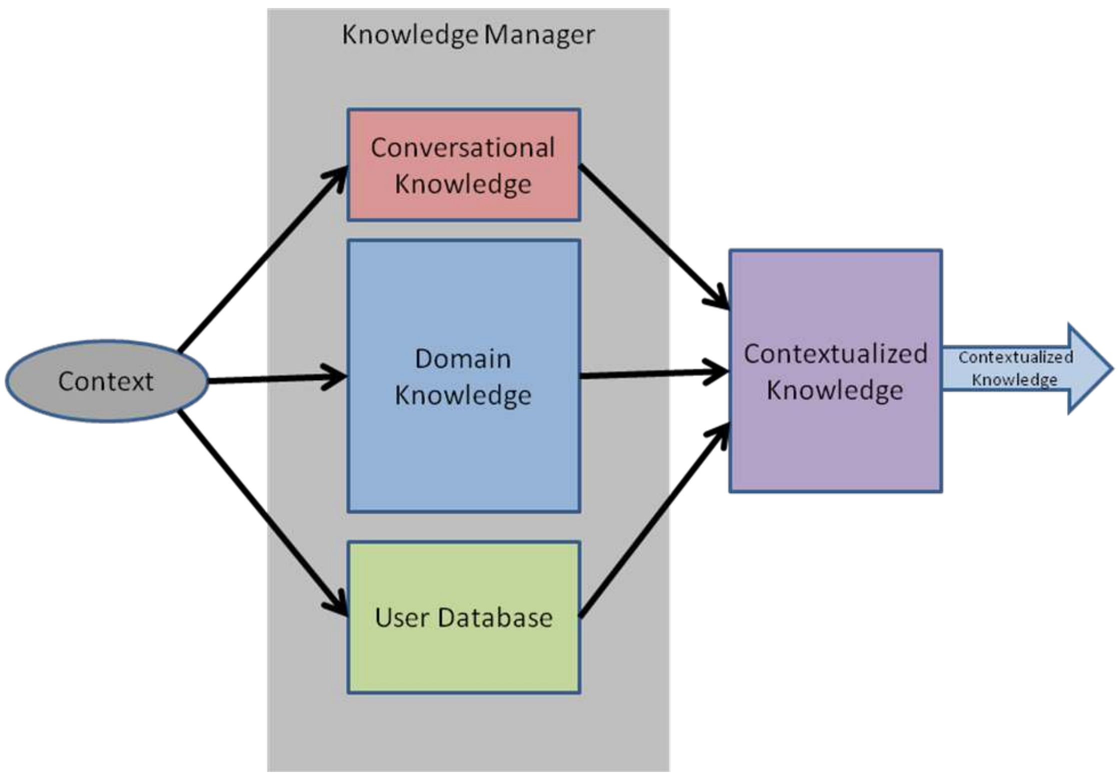


Figure 14: Knowledge Manager block diagram

User Database

The User Database keeps track of individual user profiles, where specific characteristics or traits of every different user are maintained. This data source reflects that of a person's ability to keep track of the many people s/he encounters in her/his lifetime. For CONCUR, this information provides an added layer of personalized context when dealing with specific users. Hence, the User Database contributes to the naturalness of a conversation by incorporating personally relevant information.

Another function of the User Database is its utility as a backend profile information data center. Here, CONCUR provides its usefulness as a productivity tool in which the user can update her/his individual account through a dialog-based input system. The User Database simply serves as the backend data container.

Conversational Knowledge Base

A backbone of conversational knowledge is needed to deploy the unique behaviors of the avatar. This database only deals with the transitional speech actions to be interspersed among the domain knowledge deployments. For example, sentences that are unique to the individual being represented by the avatar such as “Howdy,” “Keep the peace,” or “What else would you like to know?” would be found in the Conversational Knowledge.

Domain-Specific Knowledge Base

The Domain-Specific Knowledge Base provides topical information relevant only to the user goals featured in the dialog system. Such a body of knowledge reflects that of a person’s expertise about a certain topic. For this CONCUR prototype, the expertise of the National Science Foundation (NSF) Industry/University Cooperative Research Center program (I/UCRC) is used.

The Domain-Specific Knowledge Base may be treated as an Expert System-like data repository. In expert systems, hand-tailored definitions, rules and relationships are developed for a specific domain. For CONCUR's NSF I/UCRC domain, the existing AlexDSS Expert System (Sherwell et al, 2005) provides a solid basis for such domain information. The prototype prepares this data in a format that can be used by CONCUR's Domain-Specific Knowledge Base. This format is discussed in greater detail in Chapter Five.

Contextualized Knowledge Base

The main idea behind the knowledge manager is to provide only a subset of the agent's entire knowledge base, as directed by a specified context. Such grouping of knowledge allows for a more easily consumable amount of data for a machine to process and for the user to hear. This serves as a metaphor for human knowledge management. As an example, consider a person that is brushing her/his teeth. S/he has no real need to understand how to drive a car, at that particular moment. For those instances, it is seen that the dental hygiene knowledge contexts and the automobile operation contexts are not useful to the person at the same time. Hence, as humans, a *localization* of a knowledge base is the minimal, and perhaps preferred, amount of knowledge needed at any given time.

For CONCUR, this localization of global knowledge exists within the knowledge manager as the Contextualized Knowledge Base, as seen in the *KnowledgeBase* class' *GetContextualizedKnowledgeBase* member function. This *KnowledgeBase* subset serves as a context-based sampling of information of a certain data source. This is made possible by attributing every single piece of knowledge with a *contextual profile*, as identified in the

MicroContext class' *Name* member component. Basically, these profiles simply tag each data point with a set of conversationally relevant contexts. Hence, during a conversation, the agent utilizes the Contextualized Knowledge Base to solicit all *KnowledgeBase MicroContext* items that are relevant for a certain context. The algorithm in Figure 15 describes the contextualized knowledge retrieval process.

1. Declare contextualized KnowledgeBase Z
2. Given current context x and KnowledgeBase B
    a. For each MicroContext $m \in B$
        i. If context name $c \in m$ exists in x, add m to Z
3. Return contextualized KnowledgeBase Z

Figure 15: Contextualized Knowledge Retrieval Algorithm

CxBR-Based Discourse Model

The 'brains' of CONCUR resides in its CxBR-based Discourse Model. The effectiveness of CONCUR's intelligence lies in the performance of this system. It serves as the final destination for the previously mentioned mechanisms, where they all come together to contribute information for processing within the Discourse Model. As a result of this collaboration, the actual speech acts of the dialog-based agent are executed from the Discourse Model. With knowledge of the current state of the conversation, the Discourse Model pieces together the information of the Input Processor and the Knowledge Manager, also combining its own CxBR devices to provide an appropriate reply to the user.

Two sub-systems comprise the entire Discourse Model. These include the Goal Bookkeeper and the Context Topology. Within the Goal Bookkeeper is a support structure known as the Inference Engine. This component acts as a context activator, whose main purpose is to process the immediate information environment to determine the current contextual state. Such functionality may be regarded as *goal recognition*. The Goal Bookkeeper manages the servicing of different conversational tasks, thus addressing the issue of multiple, asynchronous *goal management*. The Context Topology provides the underlying contextual relationship infrastructure. The Goal Bookkeeper feeds information to and from the Context Topology to determine the most appropriate speech act to provide. Figure 16 displays the overall block diagram of the CxBR-based Discourse Model. The remainder of this section describes each Discourse Model component.

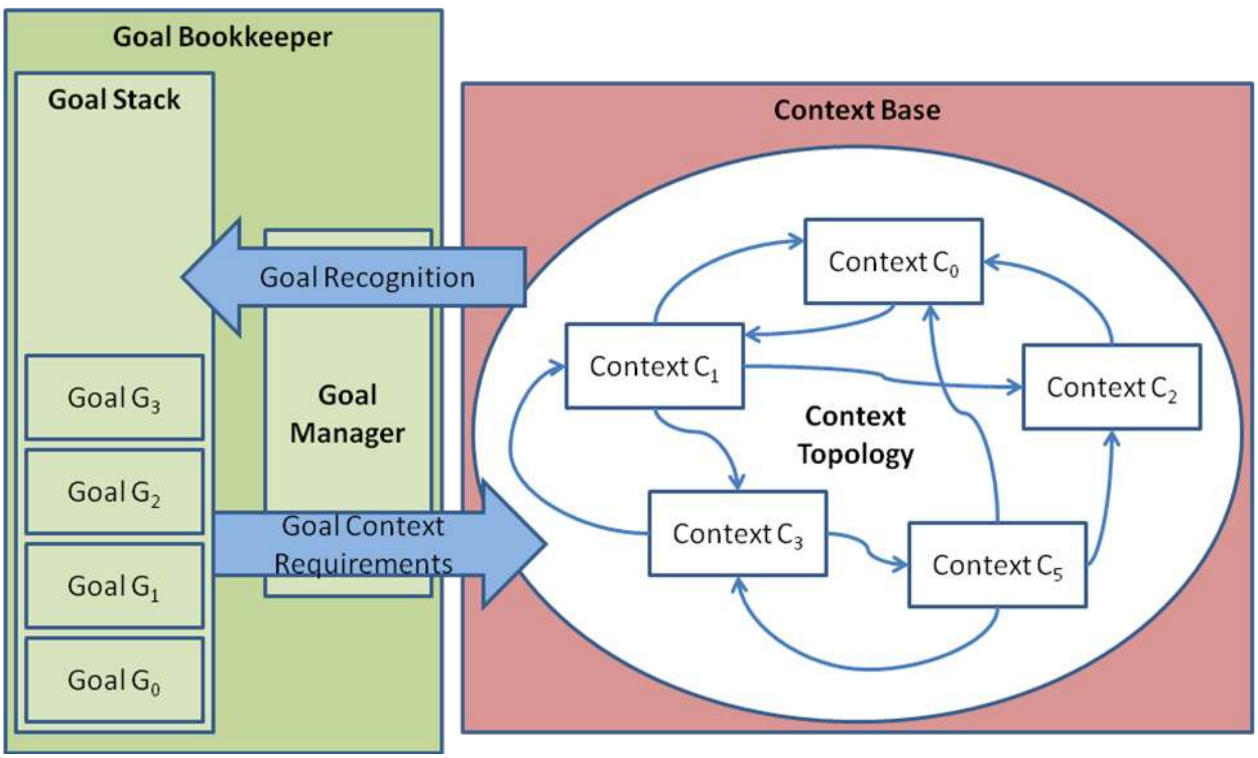


Figure 16: CxBR-based Discourse Model block diagram

Goal Bookkeeper

The Goal Bookkeeper maintains the goal-based activities of conversation agent, and it consists of two parts: the Inference Engine and the Goal Stack. The Inference Engine determines the user intent from her/his utterances, and the Goal Stack manages the lineage of user intents throughout a conversation. Table 14 and Table 15 list the member components and member functions of the *GoalBookkeeper* class.

Table 14: GoalBookkeeper class member components

Component	Data Type	Description
GoalStack	Stack of Strings	Goal stack
ActiveContextName	String	Current context
ResponseHistory	List of Strings	History of agent responses
DomainResponseHistory	String	History of agent domain expertise responses
ContextHistory	List of Strings	History of identified domain expertise contexts

Table 15: GoalBookkeeper class member functions

Function	Return Data Type	Description
PushContext	Void	Pushes a context to the top of the GoalStack
PopContext	String	Pops top string from GoalStack and returns it as the ActiveContextName
ContextGoalMet	Boolean	Checks if a context goal has been fulfilled

The *GoalStack* member is implemented as a stack containing each encountered context name whose goals have not been completed. *PushContext* and *PopContext* push and pop context names from the top of *GoalStack. ActiveContextName* contains the current conversational context, as determined by the Inference Engine. *ContextHistory* records every context name that has appeared on the *GoalStack* at least once. *ResponseHistory* records each agent responses, and *DomainResponseHistory* keeps a list of every agent response deployed from the expert domain knowledge base. The latter history is important in determining whether an information deployment goal has been fulfilled, a function performed by *ContextGoalMet*. The algorithm in

Figure 17 lists the general operation of *ContextGoalMet*. The remainder of this section discusses the Goal Bookkeeper in light of the *GoalBookkeeper* class.

1. Given DomainResponseHistory *h,* contextualized KnowledgeBase Z
   a. For each MicroContext $m \in Z$
      i. If FullSentence $s \in m$ does not exist in z, return false
   b. If all s ∈ m exists in Z, return true

Figure 17: Context Goal Completion Check Algorithm

Inference Engine

The Inference Engine takes the keyphrase version of user input, as seen in the *UserResponse* class' *KeyPhrases* member component, plus the currently active context, stored in the *GoalBookkeper* class' *ActiveContextName*, to determine the next context to add to the Goal Stack. Upon determining the context of a user's input, a contextual comparator resolves whether a context transition is needed. The general idea is that the Inference Engine serves as the goal identifier and the context activator.

The Inference Engine uses a word-by-word analysis of the user input to determine the best context to activate. This analysis is conducted using a standardized NLP toolkit. Knowledge from the domain-specific knowledge base, plus a WordNet-based general ontology is used to perform the contextualization. Pre-processing of the knowledge corpus provides a set of keyphrases that are relevant for each context. User input that aligns to the active context and matches up to the context's keyphrase tags reassures the system that no context transitioning is required. Context transitions occur when the active context does not adhere to the Inference Engine's interpretation of the user input's context. Another scenario may happen where the user input cannot be matched to *any* context. This occurs when the user has presented a statement that either lies outside of the agent's domain boundaries or has been misinterpreted by the system. In both cases, the user is informed of the contextual limitation and the last context that was being serviced is activated.

The Inference Engine is triggered after every user speech act. Its fundamental purpose is to determine the contextual locale of the conversation. Upon receiving this input, the algorithm listed in Figure 18 is executed.

1. Given UserResponse *U,* current context *x*, KnowledgeBase *B*
2. If context $c \in U$ matches *x*, return *x* as current context
3. Else request push of *x* and *c* to goal stack *G* and return *c* as current context

Figure 18: Inference Engine Algorithm

The Inference Engine's context identification process is inspired by past linguistic work pertaining to using contextual information for re-constructing semantic intent from ambiguous or ill-formed text. de Almeida and Libben (2005) used context to help disambiguate semantic meaning of trimorphemic words, such as the adjective 'unlockable.' This word has an ambiguous meaning because it can mean both 'ability to be unlocked' and 'not capable of being locked.' For example, the former interpretation of unlockable could describe a door equipped with a functioning deadbolt, in which case the bolt could be undone to unlock the door. The latter definition reflects a door that does not have any lock at all or contains a broken deadbolt, making it not able to being locked. Both of these cases describe a door that is unlockable. de Almeida and Libben conceded that the surrounding sentence information, or context, provides a decisive force in interpreting which definition of the trimorphemic word to use. Brown and Knight (1990) and Frankish and Turner (2007) conducted similar research in obtaining semantic intent from misspelled words. They demonstrated that text with scrambled letters could be correctly identified given the context of the sentence. For example, the sentence "teh orbwn dgo umpjed orev eth ecnfe," can be interpreted as "the brown dog jumped over the fence." Here, the unraveling of a few of the scrambled words creates a domino effect in processing the remaining misspelled words. The idea is that a partial interpretation of an otherwise semantically ambiguous sentence can lead to context discovery, which serves as a guide to understanding the rest of the textual information. The work in this dissertation operates under similar terms, but at a conversational level, rather than at an individual word level. In particular, the Inference Engine pieces together the user's semantic intent by collecting contextual information through keyphrase detection of partially accurate ASR results.

Goal Stack

The Goal Stack, existing as the *GoalBookkeeper* class' *GoalStack*, directly manages the conversation flow. This mechanism serves as an agent's short-term memory during a conversation. Its job is to ensure that all contexts that are introduced into a dialog exchange are attended to in the order they are brought forth. The Goal Stack performs its context management immediately upon the Inference Engine's selection of the Current Context. The algorithm in Figure 19 delineates the Goal Stack's operation.

1. Given goal stack *G*, accept current context *c* from Inference Engine
2. Check if *c* has been initialized:
    a. If *c* has not been initialized, initialize Context instantiation and push *c* onto *G*, perform necessary Context Transition
    b. If *c* has been initialized, check if Context Transition is needed:
        i. If *c* is not the same as the Previous Context *p*, push *c* onto *G* and initiate a Context Transition
        ii. Else if *c* is the same as the *p*, no further action needed
        iii. Else if *c* is a completed goal, pop *p* from *G* and initiate a Context Transition

Figure 19: Goal Stack Algorithm

Context Topology

Contexts represent the set of behaviors through which the system will respond. It must be noted that these behaviors are a set of pre-programmed speech acts, rather than synthesized sentences, as dynamic response generation remains a difficult problem in and of itself. Two major types of contexts make up these dialog-based behaviors: Agent Goal-Driven Contexts, and User Goal-Driven Contexts. These two sides of conversational contexts reflect Grice's (1975) treatment on goals in dialog. Agent Goal-Driven Contexts pertain to those actions needed for the avatar itself to perform its duties as an interfacing agent. User Goal-Driven Contexts refer to those behaviors needed to help support the user fulfill her/his needs. The co-existence of User Goals and Agent Goals allows for a mixed-initiative style of dialog. The following lists the set of contexts and sub-contexts required for the system, plus a brief description of their general functionalities.

1. Agent Goal-Driven Contexts: Conversational Knowledge actions
    a. Introduction Context
        i. Greeting Sub-Context : Exchange conversational introductions
        ii. Name Sub-Context: Agent gives its name
    b. Interruption Context
        i. Repeat Sub-Context: Reiteration of last statement by Agent
        ii. Reiteration Sub-Context: Agent asks for reiteration of User's previous input
        iii. Ignorance Sub-Context: Agent pleads ignorance of User's intent
        iv. Uncertainty Sub-Context: Agent is uncertain of User's utterance
    c. Restart Context: Agent requests that the conversation be reset

d. Initiative Context
    i. Cold Query Sub-Context: Agent asks for a general topic to discuss
    ii. Warm Query Sub-Context: Agent asks for a contextually relevant topic to discuss
    iii. Context Transition Sub-Context: Context change has been detected by Agent
    iv. Continue Sub-Context: Current context continuation
e. Closing Sub-Context
    i. Farewell Sub-Context: Exchange farewells before ending conversation
    ii. Quit Sub-Context: User requests a pre-mature exit of conversation
    iii. Done Sub-Context: Agent requests a pre-mature exit of conversation

2. User Goal-Driven Contexts: Domain Knowledge actions
    a. Agent Output Clarification Context: Re-deliver Agent's previous output
    b. Overview of Planning Grant Context
        i. About the Planning Grant Sub-Context
        ii. Deadlines Sub-Context
        iii. Considerations for Writing a Planning Grant Proposal Sub-Context
        iv. Rejected IUCRC Proposal Sub-Context
        v. Joining an Existing Center Sub-Context
        vi. Letter of Intent Sub-Context
    c. Planning Grant Paper Context
        i. Planning Grant Proposal Sub-Context
        ii. Title Sub-Context

iii. Project Summary Section Sub-Context

iv. Objective Section Sub-Context

v. Project Description Section Sub-Context

vi. Supplementary Documents Sub-Context

vii. Marketing Plan Sub-Context

viii. Staff Plan Sub-Context

ix. Membership Agreement Sub-Context

x. Draft Agenda Sub-Context

xi. Letter of Interest Sub-Context

xii. Budget Sub-Context

d. Planning Grant Meeting Context

i. About the Planning Grant Meeting Sub-Context

ii. Project Presentations Sub-Context

iii. Executive Summary Sub-Context

iv. Avoid Presentation Pitfalls Sub-Context

v. L. I. F. E. Forms Sub-Context

vi. Industry Needs and Expectation Workshop Sub-Context

vii. NSF Closed Session Sub-Context

The general algorithmic operation of the Context Topology is listed in Figure 20.

1. Given current context c, and DomainResponseHistory h
2. Retrieve contextualized KnowledgeBase Z for c
3. Declare list of response strings R
4. For each FullSentence $s \in Z$ such that $s \notin h$, add s to R
5. Return R

Figure 20: Context Topology Algorithm

Overall CONCUR Operation

All three dialog manager components (Input Processor, Knowledge Manager, and CxBR Discourse Model) cooperate to formulate a natural language speech action mechanism. This cooperation results in the entirety of the agent's cognitive activity. Figure 21 gives the overall operational algorithm of CONCUR. The boldface phrases are references to previously mentioned component algorithms.

1. Pre-process corpus files for Conversational, Domain and User-Profile knowledge into separate KnowledgeBase containers using **Knowledge File Preparation**
2. Push Initiative Context to **Goal Stack** *G*
3. Push Introduction Context to *G*
4. Main loop:
   a. Pop Current Context *c* from *G*
      i. Execute Closing Context if End conversation parameters are satisfied
      ii. Else if *c* is completed as per **Context Goal Completion Check**, push Initiative.Warm Query Sub-Context to *G*
      iii. Else, use **Context Topology** to service next immediate *c* MicroContext through Initiative.Continue Sub-Context
   b. Process User Input *U* via **User Utterance Processing**
   c. Apply *U* to **Inference Engine**
      i. If *U* is an interruption, service Interruption Context
      ii. Else if *U* is out-of-corpus according to **Contextualized Knowledge Retreival**, push Interruption.Ignorance Sub-Context to *G*
5. End conversation parameters
   a. $c \in$ Closing Context
   b. Time limit for conversation exceeded and Initiative.Cold Query Sub-Context is to be popped off *G*

Figure 21: Overall CONCUR Algorithm

Example Dialog

Table 16 demonstrates the multiple, asynchronous conversational context management that CONCUR provides as a dialog agent. It must be noted that this particular sequence of conversation is not an actual dialog taken from prototype testing, but rather an archetypical, but realistic, exchange with CONCUR. The idea behind this conversation is to represent and exemplify the different types of contextual management features of the system.

Table 16: Example CONCUR conversation

(Greeting Context) **CONCUR Agent:** Hello there! What is your name?
User: My name is Alex.
(Cold Query Search Context) **CONCUR Agent:** Hi Alex. We've met. How can I help you?
User: Did you see the Yankees game last night?
(Ignorance, Re-address Cold Query Context) **CONCUR Agent:** I'm sorry. I'm not sure what you are talking about. What things related to NSF that you would like to discuss?
User: Can you talk about SBIR supplements?
(Context Transition to SBIR Context) **CONCUR Agent:** SBIR supplements for the coming school year are almost due. I can email you the materials for those supplements. What else do you want to know about? *(Context Transition to Warm Query Context)*
User: I'd like to speak with you about my latest REU proposal.
(Context Transition to REU Proposal Context) **CONCUR Agent:** The REU stands for Research Experience for Undergraduates. What else do you want to know?
User: Oh, that reminds me, have can you tell me about MIPR requests?
(User goal interrupt, Context Transitioning to MIPR Context) **CONCUR Agent:** MIPR is an acronym for Military Interdepartmental Purchase Request. Do you want to know more about it?
User: No thanks.
(Context Transition back to REU Proposal Context) **CONCUR Agent:** REU programs typically support ten students. The sites may exist either in the US or abroad. Can I tell you more information?
User: No, that's all I need to know for now.
(Context Transition to User-Goal Search Context) **CONCUR Agent:** Is there anything else I can help with?
User: Nope. Thanks for your time.
(Context Transition to Closing Context) **CONCUR Agent:** OK. Well, I'll see you friend. Keep the peace.

This example begins with a greeting context, undergoes a series of conversational goal switches, and concludes with a closing context. The idea here is to provide exemplary situations that CONCUR is designed to handle as a prototypical system. It is noted that the example dialog is of a domain-specific nature, where the CONCUR agent portrays an expert on the NSF I/UCRC program. Boldface type indicates an interlocutor turn by either the agent or the user. The italicized text denotes the Context activity and the underlined words indicate the contextual cues from the User responses to be processed by CONCUR.

Table 16 captures some of the key functionalities of CONCUR dialog system. Specifically, CONCUR's strength lies in its ability to service a variety of different conversational goals in an asynchronous manner. Its prototype provides a user interaction similar to the dialog exchange depicted in this section.

Overall Design

Figure 22 gives an overarching diagram of the CONCUR design. Each of the main parts (Input Processor, Knowledge Manager, and Discourse Model) is represented with their respective support sub-systems. The inputs to CONCUR are the user utterance and the three data corpus sources (Domain, Conversational, and User Data). The resulting output is the agent response, a text string that is synthesized by the LifeLike Avatar text-to-speech facilities.

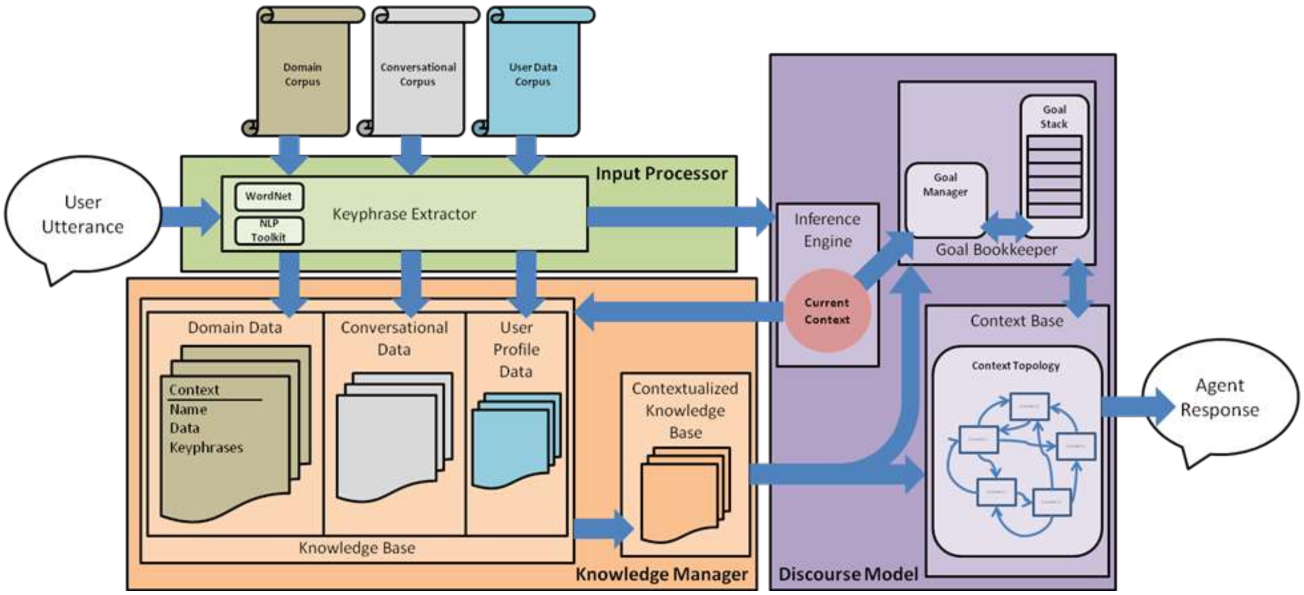


Figure 22: Overall CONCUR block diagram

While the details of each major component have been previously discussed in this chapter, a few of important observations about CONCUR from Figure 22 include:

- The Current Context is a central piece of data within the architecture, as the idea of contextualization in speech behavior modeling is the main focus of this dissertation. After the Inference Engine has determined which context to concentrate on, the Knowledge Manager is immediately charged with collecting a Contextualized Knowledge Base in regards to this context, and the Discourse Model's Goal Bookkeeper is charged with assessing whether the context aligns to the current conversational goal situation.
- The Input Processor's responsibility is two-fold: 1) prepare the data corpus for use in the Knowledge Manager during the pre-conversation phase of CONCUR's operation, and 2) continually process the user's input during a live dialog. In each of these tasks,

a keyphrase extraction technique is executed, with the assistance of an NLP-based POS determiner and a WordNet ontology.

- While the use of context permeates throughout CONCUR, the essence of traditional CxBR infrastructure exists within the Discourse Model. The characteristic context topology and essential inference engine, as dictated by CxBR design (Gonzalez and Ahlers, 1998) (Stensrud et al, 2004) (Gonzalez et al, 2008), reside within this portion of CONCUR's design.

Chapter Summary

This chapter began with general approach of goal management in a dialog system. Included in this chapter were descriptions for the high-level elements needed to accomplish the goal management aspect of a conversation agent. Additionally, a generalized framework tying together these elements was presented.

The technical detail of the CONCUR dialog system was laid out, complete with a prototype to demonstrate its functional strengths. A common theme of each of CONCUR's major components is the use of context-based constraints. This architecture speaks to the idea that CxBR lends itself easily to supporting the asynchronous switching between user goals, as defined by its continuous context-transitioning operation. The remainder of this dissertation speaks to presenting the results of testing the CONCUR prototype.

CHAPTER SIX: EVALUATION AND RESULTS

The purpose of the evaluation process in this dissertation was to collect data supporting the hypothesis that the presented dialog system provides an HCI experience that can be characterized by three features: 1) ASR resilience, 2) knowledge management domain independence, and 3) open dialog discourse. The prototype system, CONCUR, was evaluated using both qualitative and quantitative methods. This chapter begins with an overview of general conversation agent evaluation challenges, followed by a description of the dialog system's primary objectives and a discussion of the metrics associated with these goals. Following that section is an explanation of the evaluation plan to experimentally assess the dialog system using these objectives and metrics. The final section gives the acquired data sets and their associated discussions.

How Others Have Evaluated Conversation Agents

Evaluation of chatbots has always remained a controversial topic, as it is unclear on how to quantitatively describe how *well* a conversation agent performs, or how much *better* one is over another. Furthermore, one of the hurdles in this dissertation remains the definition of *naturalness*, as in how well a chatbot can maintain a natural conversation flow (Edlund et al, 2006). The following section describes some current chatbot evaluation methods (both quantitative and qualitative), as well as a definition for naturalness in relation to HCI applications.

Typically, conversation agents have been evaluated using a subjective method, usually involving a human user questionnaire. Semeraro et al (2003) employ this technique for their bookstore chatbot. In the questionnaire, seven characteristics were appraised: impression, command, effectiveness, navigability, ability to learn, ability to aid, and comprehension. Users would assess their associated satisfaction for each of these metrics, ranging from 'Very Unsatisfied' to 'Very Satisfied.' Semeraro et al recognize the fact that this subjective evaluation does not provide statistically verified conclusiveness, but rather it serves as a general indicator of performance.

Shawar and Atwell (2007) propose a universal chatbot evaluation system. They suggest three metrics, which were applied to an ALICE-based Afrikaans conversation agent. The first metric concerns dialog efficiency, which deals with the four major types of ALICE pattern-matching styles: atomic matching, first word matching, most significant matching, and no matching. Using four dialog trials, ranging from 17 to 51 conversation turns, the agent response match type frequency was recorded. Shawar and Atwell saw that first word and most significant matching methods were the most frequently used styles. The second metric is the dialog quality metric, which qualitatively categorizes, by human judgment, a chatting agent's responses into *reasonable*, *weird but understandable*, and *nonsensical*. The final metric is users' satisfaction, which is also qualitatively measured. Feedback from the chatting software end-users is collected and used to directly evaluate the agent's performance. Despite their efforts to establish a set of generic metrics, Shawar and Atwell discourage the use of a universal evaluation system for conversation agents. Instead, they conclude that the proper assessment of chatbots is the end result in how successfully it accomplishes its intended goals.

Evaluation of maintaining *naturalness* in a conversation similarly suffers from the same inherent problems of the general chatbot evaluation system. Again, subjectivity plays a large role in assessing the naturalness of a conversation. Rzepka et al (2005) used a 1-to-10 scale for two metrics: a "naturalness degree," and a "will of continuing a conversation degree." Thirteen human judges used these measures to evaluate a Rzepka et al's conversation agent's utterances. While their assessment system did not identify a concrete baseline for *universal* naturalness, they were able to make *relative* measurements of naturalness between different dialogue management approaches, such as comparing an ELIZA-based manager with a WWW-based commonsense retrieval system.

Chatbot evaluation remains an open problem, especially because of its dependence on subjective assessment. Researchers have used questionnaire-based methods to provide general insight on the effectiveness of their conversation agents. Similarly, measuring conversational *naturalness* also relies on user subjectivity. The major pitfall of these evaluation methods is their lack of quantitative universality, as no set of chatbot performance metrics has yet been defined. Nevertheless, current research has found success in using these techniques to make relative comparisons between conversation agents. Conversation agent evaluation, with emphasis on naturalness, plays a substantial role in appraising the performance of the work in this dissertation.

Objectives

An assistive dialog system proves its effectiveness under the light of two primary objectives: 1) dialog performance, and 2) task success. Dybkjær and Bernsen (2001) refer to these as "naturalness of dialogue structure" and "task and domain coverage," respectively. Each of these aims reflects different aspects of a human-computer conversation. Dialog performance relates to the *experience* of the interaction, while task success is concerned with the *utility* of the dialog exchange. Basically, these two objectives separately assess the effectiveness of the *means* (dialog performance) and the *ends* (task success).

The main goal of this dissertation is to achieve task success and dialog performance levels that are: 1) effective in terms of functional competence, and 2) acceptable in terms of minimal *awkwardness*. The following section describes the metrics chosen to measure task success and dialog performance.

Evaluation Process

The evaluation process featured in this dissertation is derived from the PARAdigm for DIalogue System Evaluation (PARADISE). (Walker et al, 1997) A multimodal version of this system exists in PROMISE (Berringer et al, 2002), but this work will reference PARADISE for simplicity's sake. Sanders and Scholtz (2000) affirm that ECA and chatbot goals for interaction are essentially the same. Figure 23 depicts the structure of the objectives and their corresponding metrics within PARADISE.

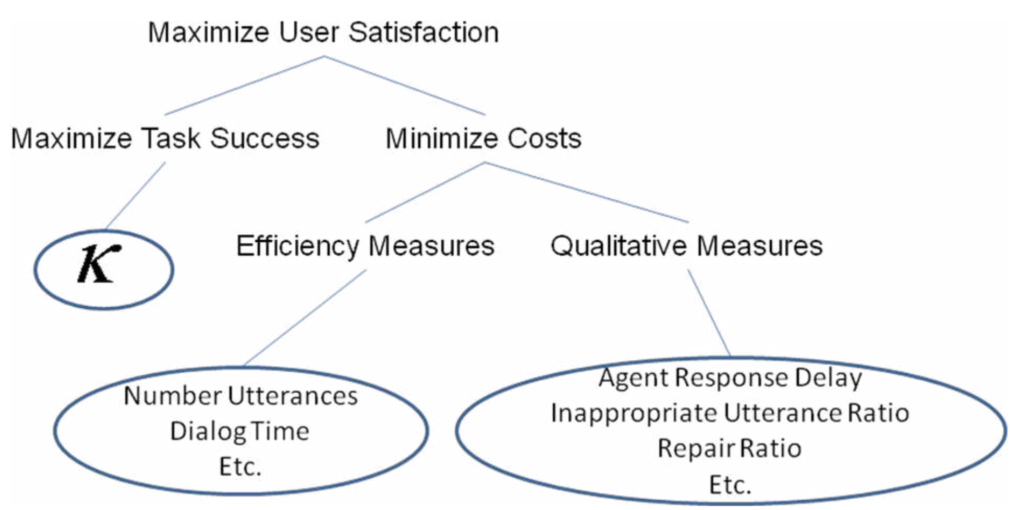


Figure 23: PARADISE objectives and metrics (Walker et al, 1997).

In this diagram, the master objective is *user satisfaction*, which is comprised of *task success* and *dialog costs*. Walker et al (1997) further break down the dialog costs to efficiency measures and qualitative measures. These PARADISE-based objectives directly reflect the task success and dialog performance objectives mentioned in the previous section. The next sections discuss the metrics involved in task success and dialog costs.

Task Success

The tasks involved with a conversational dialog system are of a multiple-goal nature. Thus, for any exchange, all of these goals must be recognized and satisfactorily serviced for the entire task to be considered successful. Conversations are modeled as a set of attribute-value pairs. Every user goal (and sub-goal) corresponds to an *attribute*, and the dialog agent's response to those goals represents a *value*.

As in PARADISE (Walker et al, 1997), an attribute-value matrix (AVM) is created for both the expected response and the actual agent response in a conversation. A confusion matrix is produced to identify the discrepancies between the expected and actual attribute-value pairings. Table 17 gives an example Attribute-Value confusion matrix for a travel schedule system, with Departure and Arrival attribute-value pairings. (Walker et al, 1997)

Table 17: Example Attribute-Value confusion matrix (Walker et al, 1997)

	KEY													
	Depart-City				**Arrival-City**				**Depart-Range**		**Depart-Time**			
DATA	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14
v1	16		1		4				3	2				
v2	1	20	1			3								
v3	5	1	9	4	2		4	2						
v4	1	2	6	6			2	3						
v5	4				15				2	3				
v6	1	6				19								
v7			5	2	1	1	15	4						
v8		1	3	3	1	2	9	11						
v9	2				2				39	10				
v10									6	35				
v11											20	5	5	4
v12												10	5	5
v13											5	5	10	5
v14												5	5	11
SUM	30	30	25	15	25	25	30	20	50	50	25	25	25	25

The rows represent the actual values recorded, and the columns reflect the expected values. In this matrix, there are four possible values for the departure city task. The departure city *v1* was correctly identified 16 out of 30 times, while *v2* was agreed upon 20 out of 30 times. This type of accuracy data may be extrapolated from the Attribute-Value confusion matrix. From these data, task success, κ is computed as the percentage of ‘right’ responses given by the agent.

<u>Performance Function</u>

Walker et al’s PARADISE uses a performance function to evaluate the *total* effectiveness of a dialog system in relation to its task success, κ, and its dialog costs, c_i:

$$Performance = (\alpha \times \mathrm{N}(\kappa)) - \sum_{i=1}^{n} w_i \times \mathrm{N}(c_i)\,. \tag{1}$$

In this relationship, κ is weighted by α, and each w_i is a weight on c_i. The weight assignments are established arbitrarily by the evaluation system developer. The function N is a Z-score normalization process used to balance out the effects of κ and c_i on the overall system performance. It is important to note that the weight assignments and the N function appear to serve as subjective factors rather than representations of consistently quantifiable information or empirical data.

This performance function purportedly allows for a normalized method of comparing two different dialog systems using the same conversational task goals. While the notion of an aggregate index to represent overall agent performance would provide an excellent method to compare different ECAs and chatbots using a single number, the arbitrary nature of the weight assignment variables in the performance function does not serve as a fair comparison if two different agents use different sets of w_i weights. In this dissertation, the performance function was not used for this reason. Instead, individual quantitative categories for different systems were compared amongst each another to ensure level grounding for each agent's performance assessments.

Dialog Costs

Dialog performance is defined as a function of two types of dialog costs: efficiency and quality. Efficiency costs refer to the resource consumption needed to accomplish a single task or sub-task. These attributes can be measured in a solely quantitative manner. Qualitative costs measure the actual conversational content. These metrics may be recorded quantitatively or qualitatively. For qualitative assessments, users are given a Likert scale-based questionnaire following their interactions, providing feedback on the dialog system's naturalness, friendliness, et cetera. Walker et al (1997), Stibler and Denny (2001), Charfuelán et al (2002), and Hassel and Hagen (2005) provide some examples on suitable dialog costs. Table 18 delineates the relevant cost metrics for this dissertation.

Table 18: Dialog cost metrics

Metric	Type	Data Collection Method
Total elapsed time	Efficiency	Quantitative Analysis
Total number of user turns	Efficiency	Quantitative Analysis
Total number of system turns	Efficiency	Quantitative Analysis
Total elapsed time per turn	Efficiency	Quantitative Analysis
User words per turn	Efficiency	Quantitative Analysis
System words per turn	Efficiency	Quantitative Analysis
Word-Error Rate	Efficiency	Quantitative Analysis
Total number of out-of-corpus misunderstandings	Quality	Quantitative Analysis
Total number of general misunderstandings	Quality	Quantitative Analysis
Total number of inappropriate responses	Quality	Quantitative Analysis
Total number of user goals	Quality	Quantitative Analysis
Total number of user goals fulfilled	Quality	Quantitative Analysis
Awkwardness rate	Quality	Quantitative Analysis
Conceptual accuracy	Quality	Quantitative Analysis
Conversational accuracy	Quality	Quantitative Analysis
Usefulness	Quality	Questionnaire
Naturalness	Quality	Questionnaire

Evaluation Metrics

The primary objective of the work in this dissertation is to provide a balanced sense of dialog performance, or *naturalness*, and task success, or *usefulness*, during a human-computer interaction. This effect can be best evaluated not by a purely quantitative treatment, but rather, through a method employing relative comparisons of both quantitative metrics and qualitative assessments. In this section, the various metrics needed for these comparisons is defined, each of which can be observed using either quantitative or qualitative measures. The most important outcomes are the assessments of whether the agent is performing in a human-like fashion and if the agent can perform in a useful manner.

Efficiency Metrics

Efficiency Metrics pertain to those interaction traits that can be empirically observed, with no need for qualitative interjection. For the most part, the prototype software internally monitors these metrics. The ASR-related metric, Word-Error Rate, was measured by comparing the textual chat log from the agent with an audio recording transcript of the exchange. The following lists the Efficiency Metrics collected, with their complete definitions:

- **Total elapsed time**: Start-to-finish time of interaction
- **Total number of user turns**: Number of times user has an utterance
- **Total number of system turns**: Number of times system has an utterance
- **Total elapsed time per turn**: Average time per turn
- **User words per turn**: Average number of words the user says per turn
- **System words per turn**: Average number of words the agent says per turn
- **Word-Error Rate**: Number of words added, deleted, and replaced divided by the expected number of words

Quality Metrics

A second set of results, called the Quality Metrics, was observed using both Quantitative Analysis and Questionnaire-based data. The quantitatively measured metrics are collected after each user interaction, where the transcript of the trial is analyzed. This analysis was executed using a manual quality inspection done by a single human judge (in this case, myself), but results in a set of qualitative findings. The Quantitative Analysis metrics included the following:

- **Total number of out-of-corpus misunderstandings**: Number of times system begs ignorance for a user request involving information not present in the corpus
- **Total number of general misunderstandings**: Number of times system begs ignorance that is not an out-of-corpus misunderstanding
- **Total number of inappropriate responses**: Number of times system gives a nonsensical response
- **Total number of user goals**: Number of conceptual goals a user brings forth
- **Total number of user goals fulfilled**: Number of times the system successfully fulfills user goals
- **Out-of-corpus misunderstanding rate**: Percentage of system turns that resulted in an out-of-corpus misunderstanding
- **General misunderstanding rate**: Percentage of system turns that resulted in a general misunderstanding
- **Error rate**: Percentage of system turns that resulted in an inappropriate response
- **Awkwardness rate**: Percentage of system turns that resulted in a general misunderstanding or an inappropriate response

- **Goal completion accuracy**: Ratio of user goals fulfilled to total user goals
- **Conversational accuracy**: Percentage of non-awkward responses from the agent

The manual quality inspection consisted of reviewing the conversation transcript and determining if an agent turn resulted in a general misunderstanding, out-of-corpus misunderstanding, or an error. For user turns, the utterances classified as information requests, or goals, were identified. Assessing the completion of these goals was performed by analyzing the quality of the system responses immediately following these requests.

At the conclusion of each test subject's interaction, the user was given an exit survey. This questionnaire directly addressed the remaining quality metrics that are impossible to observe without the user's personal input. The bulleted list below describes the metrics with their corresponding survey statements, which are answered using a Likert scale response system. APPENDIX A displays the survey instrument employed for each user trial.

- **Naturalness**:
  - If I told someone the character in this tool was real they would believe me.
  - The character on the screen seemed smart.
  - I felt like I was having a conversation with a real person.
  - This did not feel like a real interaction with another person.
- **Usefulness**:
  - I would be more productive if I had this system in my place of work.
  - The tool provided me with the information I was looking for.
  - I found this to be a useful way to get information.
  - This tool made it harder to get information than talking to a person or using a website.

- This does not seem like a reliable way to retrieve information from a database.

Each survey question is a statement in which the user provides her/his level of agreement, where a '1' rating is 'I disagree' and a '7' corresponds to 'I agree.' The "Naturalness" statements aim to determine whether the user was able to experience a natural or human-like conversational exchange, while the "Usefulness" statements check if the agent was able to perform as a capable information deployment tool.

Because of the way they are worded, the last statement from the Naturalness group and the last two statements from the Usefulness group are negatively presented. This means that a score of 7 is the worst score that can be assigned. Hence, when analyzing their results for an aggregate 'Naturalness' or 'Usefulness' score, the assessments from these survey statements must be translated in a positive manner, such that a score of 1 becomes a score of 7, 2 becomes 6, 3 becomes 5, and 4 remains the same. This allows the aggregate indicators of Naturalness and Usefulness to undergo a positive normalization in the presence of a negatively presented scaling system for the final three survey statements.

Description of Data Sets

During the evaluation process, users were subjected to test trials under four different assistive dialog systems. A data set associated with each of these systems was collected to provide a comparison study between different conversation agent setups, with special attention to each of CONCUR's core characteristics: 1) resilience against ASR-related errors, 2) support for open dialog, and 3) capability to maintain effectiveness over different expert domains. Table 19 provides a summary table of the data set collection setup. From this table, it is observed that each data set has a different configuration of dialog system, interface type, input method, speech action engine, and expertise knowledge domains. Additionally, the final two rows present the number of user trials and assessment surveys that were collected for each data set.

Table 19: Dialog System data set collection setup

	Data Set			
	1	**2**	**3**	**4**
Dialog System	AlexDSS	CONCUR	CONCUR	CONCUR
Interface Type	ECA	ECA	Chatbot	Chatbot
Input Method	Speech	Speech	Text	Text
Speech Action Engine	Menu-driven	CxBR	CxBR	CxBR
Domain Corpus	NSF I/UCRC	NSF I/UCRC	NSF I/UCRC	Current Events
Number of Trials	20	30	30	20
User Surveys Collected	30	30	n/a	20

Data Set User Trial Procedure

In this section, the general user trial process for each data set is explained. Each trial involved a consenting adult, as per the guidelines of the Institutional Review Board. APPENDIX B presents the IRB approval letter allowing the human user trials. For each user trial, two pieces of data needed to be attained: a completed user interaction perception survey, and a verbatim transcript of the trial conversation. The transcript for the speech-based systems included a voice recording transcription paired with the ASR output for each user response. The premise behind collecting the survey and the transcript was to build four groups of representative data points pertaining to each of the different conversation agent configurations.

Acquiring Data Sets 1 and 2 began with asking test subjects to interact with the LifeLike Avatar (DeMara et al, 2008) at a desk in an office cubicle. Users wore a headset or spoke into a desktop microphone to present their voice input to the computer. A set of personal speakers projected the agent's response. A voice recording of each trial was made, data that would be eventually manually transcribed into a text file. Users were asked to speak in a natural manner, as if they were conversing with another human, and they were prompted to treat the conversation in an information-seeking mindset, similar to a student speaking with her/his teacher.

Data Sets 3 and 4 were acquired using the CONCUR Chatbot. For Data Set 3, user input from the Data Set 2 voice transcripts were manually entered into the text-based agent, resulting in a simulated chat session log. Since no additional users were involved for these trials, no survey results were collected for Data Set 3. Gathering transcripts for Data Set 4 involved asking Google Chat users to conduct an online chat with the CONCUR agent. Similar to the first two data set trials, test subjects were asked to communicate naturally, being mindful of correct spelling, and they were also told to treat the exchange as an information-seeking conversation. A verbatim log of each trial was retained for quantitative analysis, and the user filled out a system quality questionnaire at the conclusion of the trial (for Data Sets 1, 2 and 4). Upon collecting the chat logs, voice transcriptions, internal system data and exit surveys, quantitative analysis of these items was done to complete the data set acquisition.

The user base for each interactive trial was selected under the assumption that cultural bias should not be a major factor when compiling results. Overlapping of data set users did occur, but the time separation between each interaction trial was several months. This minimized the effect of any unfair bias for agent assessments from the same user. Data acquisition trials were conducted involving English speakers with an approximate 2-to-1 ratio of male to female subjects. Although the trials were originally designed to be used by adults, the subjects in this study held at least two years of college education. Nevertheless, the general guideline was to test any person with minimal competency in interacting with a chatbot and/or a speech-based ECA.

For each speech-based trial from Data Sets 1 and 2, four demographic categories were recorded for each user: gender, education level, non-native spoken accent presence, and expertise familiarity. Data Set 3's demographics were identical to those of Data Set 2 because they shared the same input data. The Data Set 4 demographics consisted of only gender and education level because of the lack of speech-based input and the publicly accessible nature of its Current Events domain corpus. The next sections describe the data sets in greater detail, each with a demographic break down of its users.

Data Set 1: Speech-based NSF I/UCRC AlexDSS ECA

For Data Set 1, the AlexDSS Expert System (Sherwell et al, 2005) knowledge engine was directly implemented as the discourse mechanism of the LifeLike Avatar. The purpose of this data set was to provide the primary comparative speech-based ECA for the CONCUR platform featured in Data Set 2. With its origins as a Web-based expert system, AlexDSS' method of discourse resembles that of a menu-driven automated phone operator (Karis and Dobroth, 1991) (Gorin et al, 1997) (Béchet et al, 2004) (Gustafson et al, 2008), using a highly constrained style of user input expectation. Its dialog management was built using a manually modeled method, a time-consuming process when compared to the corpus-based knowledge management of CONCUR.

Upon completion of Data Set 1, 30 trials were performed, but only 20 transcripts from those user trials were recorded. All of the trials resulted in a completed survey. Eighteen (18) users tested with the AlexDSS ECA at the January 2009 NSF I/UCRC Director's Meeting in Washington, D. C. From this event, 18 surveys and eight chat transcripts were attained. Twelve

(12) additional tests for Data Set 1 were performed at the Intelligent Systems Lab (ISL) at the University of Central Florida (UCF) in Orlando, Florida during May of 2010. This set of trials resulted in twelve more transcripts and twelve additional surveys.

Of the 20 recorded transcripts, 14 male and six female subjects were used. Five of the subjects had non-native speaking accents, and the remaining 15 held no noticeable accent. Eight trials were conducted using persons already familiar with the NSF I/UCRC infrastructure, and the other 12 people were not previously aware of the program. Four trials involved undergraduate students, and the other 16 people held at least a college Bachelor's degree. Eight of those subjects were graduate students working on either a Master's degree or a Doctoral degree. Seven out of the 20 trials involved people holding a Doctoral degree. Table 20 breaks down the Data Set 1 user demographics for the AlexDSS ECA interaction data trials.

Table 20: Data Set 1 user demographics

Demographic Category	**Demographic Sub-category**	**Number of participants (Percentage of all participants)**
Gender	*Female*	6 (30%)
	Male	14 (70%)
Education Level	*Some undergraduate school*	4 (20%)
	Bachelor's degree	1 (5%)
	Some graduate school	8 (40%)
	Doctoral degree	7 (35%)
NSF I/UCRC Familiarity	*None*	8 (40%)
	Some	12 (60%)
Non-native Speaking Accent	*None*	15 (75%)
	Some	5 (25%)

Data Set 1 Results

The results from Data Set 1 acquisition trials are listed in the next four tables. Table 21 begins with the user survey results, which consisted of a collection of 30 questionnaires.

Table 21: Data Set 1 Survey results

Trial	Statement 1: If I told someone the character in this tool was real they would believe me.	Statement 2: I would be more productive if I had this system in my place of work.	Statement 3: The character on the screen seemed smart.	Statement 4: I felt like I was having a conversation with a real person.	Statement 5: The tool provided me with the information I was looking for.	Statement 6: I found this to be a useful way to get information.	Statement 7: This tool made it harder to get information than talking to a person or using a website.	Statement 8: This does not seem like a reliable way to retrieve information from a database.	Statement 9: This did not feel like a real interaction with another person.
1	3	6	6	7	5	6	4	2	5
2	6	5	7	6	7	6	3	2	2
3	6	7	7	6	5	6	5	1	2
4	3	4	6	3	7	7	5	1	5
5	6	4	5	2	3	4	5	2	3
6	7	2	5	5	5	6	3	5	2
7	3	4	6	4	1	4	7	4	4
8	1	1	3	4	1	1	7	6	4
9	4	6	6	2	6	5	3	4	5
10	1	5	2	2	1	4	6	6	6
11	1	3	3	4	5	5	4	5	2
12	5	3	2	6	4	1	5	5	1
13	6	4	5	5	4	4	6	2	2
14	5	6	7	7	6	7	3	2	2
15	2	4	6	6	6	4	4	2	2
16	1	5	6	4	4	5	2	2	3
17	2	2	5	3	2	5	5	4	5
18	3	3	2	2	2	3	6	3	5
19	2	4	4	3	6	5	4	4	5
20	3	6	6	7	5	6	4	2	5
21	2	4	4	2	6	6	4	4	6
22	1	3	2	1	1	7	2	1	7
23	3	4	5	5	7	7	3	1	5
24	1	2	3	2	4	5	4	6	7
25	3	2	5	4	5	4	4	4	3
26	5	6	6	4	6	5	4	3	4
27	4	5	4	5	3	5	4	3	4
28	1	4	5	3	7	7	3	2	4
29	3	5	5	4	6	5	3	3	4
30	3	4	4	5	7	7	5	4	5
Average	**3.20**	**4.10**	**4.73**	**4.10**	**4.57**	**5.07**	**4.23**	**3.17**	**3.97**

According to this table, the strength of the Data Set 1 system was the ECA's ability to provide useful information in a reliable, as verified by the results of Statements 3, 5, 6, and 8. From Statement 1, it was assessed that users were skeptical of the "realness" of the agent. The remainder of the survey results yielded average reactions from the user base. Overall, the average response for each of the questions hovered around a score of 3 to 5.

Table 22: Data Set 1 Efficiency results

Trial	Total Elapsed Time (min)	Number of User Turns	Number of System Turns	Elapsed Time Per Turn (s)	User Words Per Turn	System Words Per Turn	WER
1	10:04	29	30	4.79	7.38	37.97	40.93%
2	4:56	20	21	4.17	2.95	26.14	40.00%
3	2:43	9	10	4.19	1.78	29.90	31.25%
4	2:49	11	12	3.66	1.55	36.08	17.65%
5	4:52	16	17	4.78	4.56	30.59	64.38%
6	7:06	22	23	4.85	3.77	38.26	46.99%
7	1:53	7	8	3.65	2.14	27.50	73.33%
8	2:49	11	12	3.44	4.09	30.67	76.19%
9	3:39	16	17	3.43	1.81	27.71	51.72%
10	1:54	9	10	4.06	2.56	19.80	112.50%
11	2:19	10	11	4.09	1.70	23.45	47.06%
12	3:08	10	11	4.58	2.70	36.45	46.43%
13	2:47	11	12	4.18	2.82	24.33	53.33%
14	3:29	15	16	4.11	4.93	24.00	57.14%
15	2:04	9	10	4.44	1.78	18.10	112.50%
16	2:51	11	12	4.39	1.55	27.00	58.82%
17	4:33	18	19	4.14	2.89	26.84	62.75%
18	2:28	10	11	3.78	1.80	29.55	72.22%
19	2:22	11	12	4.06	1.18	20.92	115.38%
20	3:30	12	13	4.12	2.75	36.00	36.36%
Average	**3:36**	**13.35**	**14.35**	**4.15**	**2.83**	**28.56**	**60.85%**

Table 22 shows Data Set 1's efficiency metrics results. Twenty (20) trials were performed using the the speech-based NSF I/UCRC AlexDSS agent. On average, each exchange lasted just under 4 minutes, typically consisting of 13 user turns and 14 agent turns. Each turn averaged about four seconds in duration. ASR facilities operated with a WER of about 61%. In other words, the agent was capable of catching less than half of what the users were trying to say.

Table 23: Data Set 1 Quality results

Trial	Number of System Turns	Out-Of-Corpus Misunderstandings	General Misunderstandings	Errors	User Goals	Goals Fulfilled
1	30	0	3	8	13	8
2	21	0	3	5	6	1
3	10	0	0	0	3	3
4	12	0	0	0	5	4
5	17	1	1	3	3	2
6	23	0	2	1	5	4
7	8	0	0	0	2	2
8	12	0	0	3	4	2
9	17	0	1	4	5	3
10	10	0	2	0	1	0
11	11	0	0	0	1	1
12	11	0	1	1	3	2
13	12	0	2	1	4	3
14	16	0	4	1	6	4
15	10	0	3	0	2	0
16	12	0	1	1	2	2
17	19	0	2	1	8	5
18	11	0	1	0	3	3
19	12	0	2	1	3	0
20	13	0	0	1	5	4
Average	**14.35**	**0.05**	**1.40**	**1.55**	**4.20**	**2.65**

Table 23 displays Data Set 1's quality results. From its 20 recorded transcripts, the AlexDSS ECA averaged approximately 14 system responses per exchange. During these turns, it averaged 0.05 out-of-corpus misunderstandings, 1.40 general misunderstandings and 1.55 errors per exchange. Out-of-corpus misunderstandings occur when the agent must plead ignorance if the user's information request does not exist in the domain corpus. A general misunderstanding transpires when the agent must plead ignorance for any other reason. Errors arise when an agent response is out of place or simply wrong. The AlexDSS dialog manager managed to keep these metrics low because of the direct utterance input expectation afforded by its menu-driven discourse style. An average of 4.20 goals, or user information requests was encountered in each conversation, and the agent was able to accomplish 2.65 of these goals.

Table 24: Data Set 1 Quantitative Analysis results

Trial	Out-Of-Corpus Misunderstanding Rate	General Misunderstanding Rate	Misunderstanding Rate	Error Rate	Awkwardness Rate	Goal Completion Accuracy	Conversational Accuracy
1	0.00%	10.00%	10.00%	26.67%	36.67%	61.54%	63.33%
2	0.00%	14.29%	14.29%	23.81%	38.10%	16.67%	61.90%
3	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
4	0.00%	0.00%	0.00%	0.00%	0.00%	80.00%	100.00%
5	5.88%	5.88%	11.76%	17.65%	23.53%	66.67%	76.47%
6	0.00%	8.70%	8.70%	4.35%	13.04%	80.00%	86.96%
7	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
8	0.00%	0.00%	0.00%	25.00%	25.00%	50.00%	75.00%
9	0.00%	5.88%	5.88%	23.53%	29.41%	60.00%	70.59%
10	0.00%	20.00%	20.00%	0.00%	20.00%	0.00%	80.00%
11	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
12	0.00%	9.09%	9.09%	9.09%	18.18%	66.67%	81.82%
13	0.00%	16.67%	16.67%	8.33%	25.00%	75.00%	75.00%
14	0.00%	25.00%	25.00%	6.25%	31.25%	66.67%	68.75%
15	0.00%	30.00%	30.00%	0.00%	30.00%	0.00%	70.00%
16	0.00%	8.33%	8.33%	8.33%	16.67%	100.00%	83.33%
17	0.00%	10.53%	10.53%	5.26%	15.79%	62.50%	84.21%
18	0.00%	9.09%	9.09%	0.00%	9.09%	100.00%	90.91%
19	0.00%	16.67%	16.67%	8.33%	25.00%	0.00%	75.00%
20	0.00%	0.00%	0.00%	7.69%	7.69%	80.00%	92.31%
Average	**0.29%**	**9.51%**	**9.80%**	**8.71%**	**18.22%**	**63.29%**	**81.78%**

Table 24 exhibits the results for the AlexDSS ECA quantitative analysis. An average misunderstanding rate of 9.80% was observed (0.29% for out-of-corpus misunderstandings, 9.51% for general misunderstandings) with an error rate of 8.71%. Misunderstanding rate consists of the ratio of the sum of general and out-of-corpus misunderstandings to the number of total system responses. Error rate is the ratio of the number of errors to the number of agent turns. User goals were completed at an average rate of 63.29%, and the agent yielded an 81.78% Conversational Accuracy. Conversational Accuracy reflects the percentage of agent responses that are not considered awkward. An awkward response is one that results in a general misunderstanding or an error.

Data Set 2: Speech-based NSF I/UCRC CONCUR ECA

The Data Set 2 trials involved a speech-based CONCUR dialog manager using the NSF I/UCRC domain and a fully operational LifeLike Avatar. This ECA represented the performance of speech-based systems that specialize in context-sensitive dialog management, which is the primary impetus of this dissertation. Insisting upon open dialog inputs when soliciting the user for responses, this CONCUR agent contrasted with the menu-driven AlexDSS system from Data Set 1. The NSF I/UCRC corpus mentioned in the previous chapter was loaded into this agent. The complete expert knowledge is listed in APPENDIX C. This corpus was built by collecting pertinent data about the NSF I/UCRC program (Gray and Walters, 1998) (Sherwell et al, 2005) and compiling it into an encyclopedia-style entry. A subset of this corpus is depicted in Figure 24.

//Planning Grant Paper

::Planning Grant Proposal

The planning grant proposal is required to acquire the 10000 dollar planning grant award. See the current solicitation for guidelines to preparing a planning grant proposal.

::Title

The title for a planning grant must be headed as "Planning Grant IUCRC for AREA" where area is the research area for which the center is being proposed.

::Project Summary Section

The project summary is a one page description of the industry, research focus, and university capabilities. This section needs to explicitly state the Intellectual Merit and Broader Impact of the Proposal. See the current solicitation for guidelines to preparing a Project Summary.

Figure 24: Subset of NSF I/UCRC Corpus

For Data Set 2, 30 trials were performed, recorded and resulted in a completed survey. Fourteen (14) users tested the CONCUR ECA at the January 2010 NSF I/UCRC Director's Meeting in Washington, D. C., in which 14 surveys and 14 chat transcripts were acquired. Sixteen (16) additional tests for Data Set 2 were performed at the ISL at UCF in February of 2010. From this effort, 16 more transcripts and surveys were added to the data set.

Twenty-one (21) male and nine female subjects were used, and six of the 30 trials involved people with non-native speaking accents. Fourteen (14) subjects were persons already familiar with the NSF I/UCRC infrastructure, while the remaining 16 people were not previously aware of the program. Seven users were undergraduate students, and the remaining 23 had at least a college Bachelor's degree. Of these remaining people, 10 were working on a graduate degree and 11 held a Doctoral degree. Table 25 breaks down Data Set 2's user demographics.

Table 25: Data Set 2 user demographics

Demographic Category	**Demographic Sub-category**	**Number of participants (Percentage of all participants)**
Gender	*Female*	9 (30%)
	Male	21 (70%)
Education Level	*Some undergraduate school*	7 (23%)
	Bachelor's degree	2 (7%)
	Some graduate school	10 (33%)
	Doctoral degree	11 (37%)
NSF I/UCRC Familiarity	*None*	14 (47%)
	Some	16 (53%)
Non-native Speaking Accent	*None*	24 (80%)
	Some	6 (20%)

<u>Data Set 2 Results</u>

An example conversation from this data set is exhibited in Figure 25. Within this figure, the capitalized letters following some of the user responses are the actual utterances that were transcribed from a voice recording, while the lower-cased text represents the actual ASR output. This particular example lasted just over a minute, and it had a WER of 29% and scored a 100% in both Conversational Accuracy and Goal Completion Accuracy. The complete set of 30 transcripts for Data Set 2 can be found in APPENDIX D.

(01/14/10 18:36:56) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(01/14/10 18:37:02) User to Avatar: harm run the legal (HI I'M RODNEY KWICKO)

(01/14/10 18:37:05) Avatar to User: It's good to see you. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What do you want to talk about?

(01/14/10 18:37:28) User to Avatar: of a letter of interest

(01/14/10 18:37:28) Avatar to User: A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. Do you have questions about that information?

(01/14/10 18:37:47) User to Avatar: no

(01/14/10 18:37:47) Avatar to User: How else can I help you out?

(01/14/10 18:37:53) User to Avatar: know what about the marketing plan (UM, WHAT ABOUT THE MARKETING PLAN)

(01/14/10 18:37:53) Avatar to User: The marketing plan explains how the center shall make itself attractive to potential industrial members. See the current solicitation for guidelines to preparing the marketing plan. Do you have questions about that information?

(01/14/10 18:38:11) User to Avatar: no

(01/14/10 18:38:11) Avatar to User: I guess we're done now. Keep the peace.

Figure 25: Example conversation from speech-based NSF I/UCRC CONCUR

The results from the Data Set 2 trials are listed in the next four tables. Table 26 begins with the user survey results.

Table 26: Data Set 2 Survey results

Trial	Statement 1: If I told someone the character in this tool was real they would believe me.	Statement 2: I would be more productive if I had this system in my place of work.	Statement 3: The character on the screen seemed smart.	Statement 4: I felt like I was having a conversation with a real person.	Statement 5: The tool provided me with the information I was looking for.	Statement 6: I found this to be a useful way to get information.	Statement 7: This tool made it harder to get information than talking to a person or using a website.	Statement 8: This does not seem like a reliable way to retrieve information from a database.	Statement 9: This did not feel like a real interaction with another person.
1	6	7	7	6	7	7	1	1	2
2	7	6	5	5	7	7	5	6	5
3	5	4	6	6	6	6	4	3	3
4	6	2	3	3	4	5	5	3	5
5	2	4	7	1	4	7	6	4	7
6	7	4	5	3	7	5	5	3	4
7	5	1	4	2	1	1	7	6	3
8	3	3	6	5	4	7	1	1	3
9	5	6	5	3	6	6	4	3	5
10	5	5	4	5	5	6	3	3	3
11	4	2	4	5	5	4	5	5	4
12	1	2	4	2	3	2	6	7	7
13	3	6	7	5	7	7	4	2	2
14	5	4	6	5	2	3	4	6	5
15	5	4	5	3	3	3	6	5	5
16	6	3	5	4	5	6	4	3	5
17	5	2	5	5	3	4	6	2	2
18	1	4	6	5	4	6	2	2	3
19	4	4	5	3	4	7	7	4	5
20	2	4	5	3	5	5	4	3	5
21	3	4	6	3	5	2	6	3	6
22	2	7	4	4	6	7	6	1	4
23	4	3	6	4	7	7	4	5	4
24	2	3	2	2	6	5	4	4	6
25	3	4	5	5	4	5	6	2	4
26	4	5	7	4	7	6	2	2	4
27	4	3	5	4	7	7	3	2	5
28	5	7	5	4	6	7	1	2	2
29	5	6	4	5	6	6	4	4	4
30	3	1	1	1	1	7	5	6	7
Average	**4.00**	**4.97**	**3.83**	**4.90**	**5.43**	**4.33**	**3.43**	**4.30**	**4.00**

Thirty (30) surveys were collected to judge the speech-based NSF I/UCRC CONCUR ECA. Overall, users were satisfied with the information deployment duties of the agent, as verified by the results from Statements 2 and 5. Statement 4 scored an average of 4.90 pertaining to if the user was conversing with another person. Statement 3 was assessed a 3.83, judging on whether the agent seemed intelligent.

Table 27: Data Set 2 Efficiency results

Trial	Total Elapsed Time (min)	Number of User Turns	Number of System Turns	Elapsed Time Per Turn (s)	User Words Per Turn	System Words Per Turn	WER
1	2:01	7	8	5.74	2.71	28.75	21.05%
2	2:43	11	12	4.95	3.27	23.25	33.33%
3	5:12	17	18	6.12	5.12	30.06	42.17%
4	6:00	18	19	6.66	5.22	33.26	69.15%
5	4:48	14	15	6.85	3.93	36.47	54.55%
6	2:22	7	8	6.74	6.29	30.75	50.00%
7	2:27	7	8	6.98	7.71	27.63	51.85%
8	2:46	9	10	6.16	7.78	23.80	72.86%
9	3:20	9	10	7.42	2.78	36.40	80.00%
10	1:15	5	6	4.98	3.40	23.50	29.41%
11	1:39	6	7	5.49	6.00	28.00	72.22%
12	2:35	9	10	5.76	6.22	25.10	67.86%
13	8:30	27	28	6.30	4.56	30.79	44.72%
14	2:02	6	7	6.78	8.00	25.86	56.25%
15	3:03	9	10	6.77	5.44	32.80	55.10%
16	2:37	8	9	6.55	5.13	29.33	48.78%
17	3:42	11	12	6.71	4.91	33.33	85.19%
18	2:28	9	10	5.49	3.89	29.50	71.43%
19	3:28	12	13	5.77	5.33	26.77	39.06%
20	2:56	11	12	5.32	4.55	29.33	68.00%
21	3:50	12	13	6.39	8.08	26.85	47.42%
22	2:41	9	10	5.98	4.33	29.90	79.49%
23	4:14	16	17	5.30	4.06	27.18	69.23%
24	2:31	11	12	4.59	3.73	24.75	60.98%
25	3:36	14	15	5.15	3.14	25.60	72.73%
26	3:18	8	9	8.24	3.00	39.33	33.33%
27	3:27	11	12	6.27	2.18	28.33	33.33%
28	4:30	13	14	6.93	6.46	29.79	65.48%
29	4:44	15	16	6.31	4.20	35.31	96.83%
30	1:19	6	7	4.42	6.67	20.86	82.50%
Average	**3:21**	**10.90**	**11.90**	**6.10**	**4.94**	**29.09**	**58.48%**

Table 27 exhibits the efficiency metrics results from the 30 trials involving the speech-based CONCUR agent. The average experience lasted about three and a half minutes, consisting of nearly 11 user turns and approximately 12 agent turns. Each turn lasted about six seconds. The ASR operated with a WER of nearly 59%.

Table 28: Data Set 2 Quality results

Trial	Number of System Turns	Out-Of-Corpus Misunderstandings	General Misunderstandings	Errors	User Goals	Goals Fulfilled
1	8	0	2	0	4	4
2	12	0	3	2	5	4
3	18	0	3	4	8	5
4	19	0	2	6	12	8
5	15	1	3	4	9	5
6	8	0	1	3	3	1
7	8	1	2	3	5	0
8	10	1	0	5	5	1
9	10	0	2	3	3	2
10	6	0	0	0	2	2
11	7	3	0	1	3	1
12	10	0	2	2	5	4
13	28	2	6	7	18	9
14	7	0	0	3	4	2
15	10	3	1	0	3	2
16	9	0	1	0	5	5
17	12	0	4	3	4	2
18	10	1	3	1	2	1
19	13	1	1	0	7	6
20	12	0	3	1	6	6
21	13	0	4	2	9	5
22	10	0	1	4	5	1
23	17	2	3	2	10	6
24	12	1	0	6	5	1
25	15	1	2	5	8	2
26	9	0	0	1	5	5
27	12	1	0	2	6	5
28	14	0	3	4	8	4
29	16	0	2	8	4	4
30	7	3	0	0	2	1
Average	**11.90**	**0.70**	**1.80**	**2.73**	**5.83**	**3.47**

Table 28 shows the quality metrics for Data Set 2. The CONCUR agent averaged about 12 system responses per conversation, with an average of 0.70 out-of-corpus misunderstandings, 1.80 general misunderstandings and 2.73 errors per exchange. Users presented 5.83 goals in each dialog encounter, where the agent was able to effectively service 3.47 of these requests.

Table 29: Data Set 2 Quantitative Analysis results

Trial	Out-Of-Corpus Misunderstanding Rate	General Misunderstanding Rate	Misunderstanding Rate	Error Rate	Awkwardness Rate	Goal Completion Accuracy	Conversational Accuracy
1	0.00%	25.00%	25.00%	0.00%	25.00%	100.00%	75.00%
2	0.00%	25.00%	25.00%	16.67%	41.67%	80.00%	58.33%
3	0.00%	16.67%	16.67%	22.22%	38.89%	62.50%	61.11%
4	0.00%	10.53%	10.53%	31.58%	42.11%	66.67%	57.89%
5	6.67%	20.00%	26.67%	26.67%	46.67%	55.56%	53.33%
6	0.00%	12.50%	12.50%	37.50%	50.00%	33.33%	50.00%
7	12.50%	25.00%	37.50%	37.50%	62.50%	0.00%	37.50%
8	10.00%	0.00%	10.00%	50.00%	50.00%	20.00%	50.00%
9	0.00%	20.00%	20.00%	30.00%	50.00%	66.67%	50.00%
10	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
11	42.86%	0.00%	42.86%	14.29%	14.29%	33.33%	85.71%
12	0.00%	20.00%	20.00%	20.00%	40.00%	80.00%	60.00%
13	7.14%	21.43%	28.57%	25.00%	46.43%	50.00%	53.57%
14	0.00%	0.00%	0.00%	42.86%	42.86%	50.00%	57.14%
15	30.00%	10.00%	40.00%	0.00%	10.00%	66.67%	90.00%
16	0.00%	11.11%	11.11%	0.00%	11.11%	100.00%	88.89%
17	0.00%	40.00%	40.00%	25.00%	58.33%	50.00%	41.67%
18	7.69%	23.08%	30.77%	10.00%	40.00%	50.00%	60.00%
19	7.69%	7.69%	15.38%	0.00%	7.69%	85.71%	92.31%
20	0.00%	25.00%	25.00%	8.33%	33.33%	100.00%	66.67%
21	0.00%	30.77%	30.77%	15.38%	46.15%	55.56%	53.85%
22	0.00%	10.00%	10.00%	40.00%	50.00%	20.00%	50.00%
23	11.76%	17.65%	29.41%	11.76%	29.41%	60.00%	70.59%
24	8.33%	0.00%	8.33%	50.00%	50.00%	20.00%	50.00%
25	6.67%	13.33%	20.00%	33.33%	46.67%	25.00%	53.33%
26	0.00%	0.00%	0.00%	11.11%	11.11%	100.00%	88.89%
27	8.33%	0.00%	8.33%	16.67%	16.67%	83.33%	83.33%
28	0.00%	21.43%	21.43%	28.57%	50.00%	50.00%	50.00%
29	0.00%	28.57%	28.57%	50.00%	62.50%	100.00%	37.50%
30	24.94%	0.00%	24.94%	0.00%	0.00%	50.00%	100.00%
Average	**6.15%**	**14.49%**	**20.64%**	**21.81%**	**35.78%**	**60.48%**	**63.93%**

Table 29 shows the quantitative analysis results from Data Set 2. Misunderstandings occurred 20.64% of the time (6.15% for out-of-corpus misunderstandings, 14.49% for general misunderstandings), and errors occurred 21.81% of the time. An average of 60.48% of the user goals were completed with a 63.93% conversational accuracy rate.

Data Set 3: Text-based I/UCRC CONCUR Chatbot

The third group of data set trials used a text-based CONCUR dialog manager to simulate ideal ASR conditions. This agent interface exists to contrast with the LifeLike Avatar ECA. User inputs taken from Data Set 2's transcripts were entered into a text-only version of CONCUR. While it was not possible to *exactly* replicate the transcripts, the user responses that pertained to individual user information requests were retained. The responses from the agent were recorded to reflect a version of CONCUR that does not have ASR-related input errors.

Since each trial in this data set uses user input from Data Set 2, the demographic make-up of the subjects is identical to that of the last set of trials. To review these demographics, 30 trials were performed with 21 male subjects and nine female subjects. Fourteen (14) trials involved people with prior knowledge of the NSF I/UCRC program, and other 16 participants were not aware of it before the trial. Seven trials involved undergraduate students, and the other 23 people held at least a college Bachelor's degree. Eleven (11) subjects had a Doctoral degree. Refer back to Table 25 to review the user demographics for Data Set 2, which are shared by Data Set 3.

Data Set 3 Results

The results from the third data set trials are presented in the next three tables. Table 30 begins with a listing of the efficiency metrics results.

Table 30: Data Set 3 Efficiency results

Trial	Total Elapsed Time (min)	Number of User Turns	Number of System Turns	Elapsed Time Per Turn (s)	User Words Per Turn	System Words Per Turn	WER
1	1:35	7	8	4.51	2.71	28.50	0.00%
2	2:26	10	11	4.87	3.00	24.55	0.00%
3	3:44	13	14	6.06	5.77	24.21	0.00%
4	3:00	15	16	4.00	5.87	33.38	0.00%
5	3:18	14	15	4.70	3.93	29.73	0.00%
6	1:44	7	8	4.93	6.29	28.88	0.00%
7	10:14	12	13	17.06	7.58	30.85	0.00%
8	2:20	10	11	4.65	7.40	21.45	0.00%
9	6:01	7	8	17.21	3.43	31.88	0.00%
10	1:52	5	6	7.44	3.40	24.00	0.00%
11	2:08	9	10	4.73	4.33	30.40	0.00%
12	3:15	9	10	7.23	6.22	27.50	0.00%
13	4:14	23	24	3.67	4.61	31.17	0.00%
14	7:50	6	7	26.12	8.00	18.57	0.00%
15	1:29	9	10	3.29	5.44	31.40	0.00%
16	2:02	8	9	5.09	5.13	31.56	0.00%
17	0:51	5	6	3.38	6.20	21.33	0.00%
18	1:26	9	10	3.17	3.89	29.60	0.00%
19	1:56	11	12	3.52	5.09	26.00	0.00%
20	2:56	11	12	5.32	5.00	29.75	0.00%
21	2:28	11	12	4.48	8.64	26.58	0.00%
22	2:05	8	9	5.23	4.38	30.22	0.00%
23	2:18	13	14	3.55	4.54	26.50	0.00%
24	1:30	10	11	3.01	3.90	32.82	0.00%
25	1:28	11	12	2.66	3.27	29.75	0.00%
26	1:01	8	9	2.54	3.00	36.44	0.00%
27	1:28	8	9	3.65	2.00	25.22	0.00%
28	2:29	11	12	4.52	7.18	35.75	0.00%
29	4:44	15	16	6.31	3.67	29.81	0.00%
30	2:31	8	9	6.31	6.63	18.89	0.00%
Average	**2:52**	**10.10**	**11.10**	**6.11**	**5.02**	**28.22**	**0.00%**

Thirty (30) trials were performed with the text-based NSF I/UCRC CONCUR Chatbot. Conversations lengths averaged approximately three minutes, consisting of about 10 user turns and 11 agent turns. The average turn duration was a little over six seconds. Text-based methods were utilized, therefore yielding a WER of 0%.

Table 31: Data Set 3 Quality results

Trial	Number of System Turns	Out-Of-Corpus Misunderstandings	General Misunderstandings	Errors	User Goals	Goals Fulfilled
1	8	0	2	2	0	4
2	11	0	0	0	3	5
3	14	0	0	0	2	8
4	16	0	2	2	3	12
5	15	1	2	3	2	9
6	8	0	1	1	3	3
7	13	0	0	0	10	5
8	11	0	2	2	2	5
9	8	0	0	0	3	3
10	6	0	0	0	0	2
11	10	0	0	0	4	3
12	10	0	1	1	1	5
13	24	4	2	6	5	18
14	7	1	1	2	1	4
15	10	4	0	4	1	3
16	9	0	0	0	1	5
17	6	0	3	3	0	1
18	10	5	0	5	1	2
19	12	1	0	1	0	7
20	12	1	0	1	3	6
21	12	0	3	3	3	9
22	9	0	1	1	0	5
23	14	0	1	1	3	10
24	11	0	1	1	3	6
25	12	0	2	2	0	8
26	9	0	0	0	0	6
27	9	1	0	1	0	5
28	12	0	0	0	2	9
29	16	1	1	2	4	5
30	9	5	0	5	0	2
Average	**11.10**	**0.80**	**0.83**	**2.00**	**5.83**	**3.93**

The Data Set 3 quality results, as seen in Table 31, show that the NSF I/UCRC CONCUR Chatbot averaged almost 11 system responses per exchange. The agent averaged 0.80 out-of-corpus misunderstandings, 0.83 general misunderstandings and 2.00 errors per trial. Users requested an average of 5.83 goals, and the agent was capable of accomplishing 3.93 of these goals.

Table 32: Data Set 3 Quantitative Analysis results

Trial	Out-Of-Corpus Misunderstanding Rate	General Misunderstanding Rate	Misunderstanding Rate	Error Rate	Awkwardness Rate	Goal Completion Accuracy	Conversational Accuracy
1	0.00%	25.00%	25.00%	0.00%	25.00%	100.00%	75.00%
2	0.00%	0.00%	0.00%	27.27%	27.27%	80.00%	72.73%
3	0.00%	0.00%	0.00%	15.38%	15.38%	62.50%	84.62%
4	0.00%	12.50%	12.50%	18.75%	31.25%	66.67%	68.75%
5	6.67%	13.33%	20.00%	13.33%	26.67%	77.78%	73.33%
6	0.00%	12.50%	12.50%	37.50%	50.00%	33.33%	50.00%
7	0.00%	0.00%	0.00%	76.92%	76.92%	0.00%	23.08%
8	0.00%	18.18%	18.18%	18.18%	36.36%	40.00%	63.64%
9	0.00%	0.00%	0.00%	37.50%	37.50%	66.67%	62.50%
10	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
11	0.00%	0.00%	0.00%	40.00%	40.00%	33.33%	60.00%
12	0.00%	10.00%	10.00%	10.00%	20.00%	60.00%	80.00%
13	16.67%	8.33%	25.00%	20.83%	29.17%	50.00%	70.83%
14	14.29%	14.29%	28.57%	14.29%	28.57%	50.00%	71.43%
15	40.00%	0.00%	40.00%	10.00%	10.00%	66.67%	90.00%
16	0.00%	0.00%	0.00%	11.11%	11.11%	100.00%	88.89%
17	0.00%	30.00%	30.00%	0.00%	50.00%	100.00%	50.00%
18	41.67%	0.00%	41.67%	10.00%	10.00%	50.00%	90.00%
19	8.33%	0.00%	8.33%	0.00%	0.00%	71.43%	100.00%
20	8.33%	0.00%	8.33%	25.00%	25.00%	100.00%	75.00%
21	0.00%	25.00%	25.00%	25.00%	50.00%	44.44%	50.00%
22	0.00%	11.11%	11.11%	0.00%	11.11%	100.00%	88.89%
23	0.00%	7.14%	7.14%	21.43%	28.57%	70.00%	71.43%
24	0.00%	9.09%	9.09%	27.27%	36.36%	100.00%	63.64%
25	0.00%	16.67%	16.67%	0.00%	16.67%	75.00%	83.33%
26	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
27	11.11%	0.00%	11.11%	0.00%	0.00%	80.00%	100.00%
28	0.00%	0.00%	0.00%	16.67%	16.67%	66.67%	83.33%
29	11.11%	11.11%	22.22%	25.00%	31.25%	60.00%	68.75%
30	45.18%	0.00%	45.18%	0.00%	0.00%	50.00%	100.00%
Average	**6.77%**	**7.48%**	**14.25%**	**16.68%**	**24.66%**	**68.48%**	**75.34%**

Table 32 shows the quantitative analysis results for the 30 trials that were collected for Data Set 3's CONCUR Chatbot. An average misunderstanding rate of 14.25% was observed (6.77% for out-of-corpus misunderstandings, 7.48% for general misunderstandings) with an error rate of 16.68%. An average goal completion rate of 68.48% was recorded with a 75.34% conversational accuracy rate.

Data Set 4: Text-based Current Events CONCUR Chatbot

The fourth data set was collected to exercise the performance of the CONCUR system with a new domain expertise, specifically, a Current Events domain. The impetus behind this data set was to provide insight on CONCUR's domain independence. Data Set 4's corpus encapsulated information that was more general and more publicly known than the NSF I/UCRC information. To eliminate ASR-related errors, a textual user input system was implemented for this text-based agent. A Jabber-based chat server was implemented for use in the Google Chat environment. The survey instrument was administered online immediately after a conversation trial concluded. The use of an internet-based platform allowed for less dependence on physical location for the sampling of test subjects. No physical embodiment of the agent was used, such as that used in the LifeLike ECA. Figure 26 shows the CONCUR Chatbot user interface.

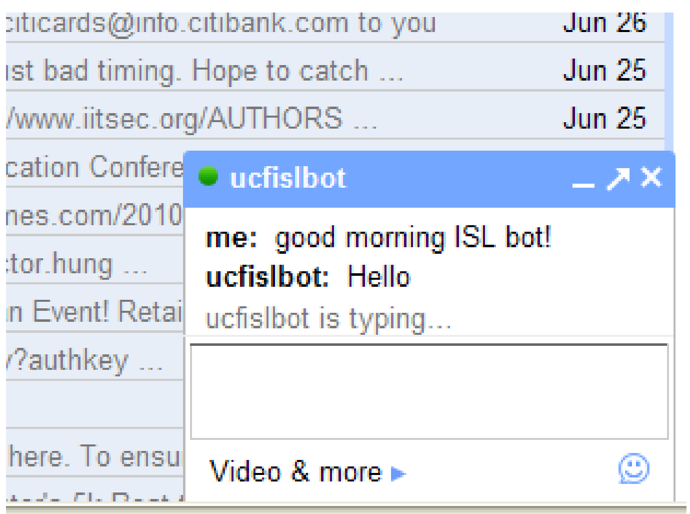


Figure 26: CONCUR Chatbot user interface

A snippet of the Current Events corpus is listed in Figure 27. This knowledge was put together to portray a less specific niche of expertise, not as closely knit as the NSF I/UCRC corpus. The entire expertise corpus for Data Set 3 is listed in APPENDIX E and contains 11 articles involving U. S. and international news, sports and health. Preparing this new corpus was the result of a three-day effort in gathering news reports and formatting the file for use in CONCUR. This preparation method was identical to that used to create the NSF I/UCRC corpus.

//Sports

::Tiger Woods

Tiger Woods' ailing neck isn't bad enough to make him hesitate about scheduling future tournaments. Woods officially entered the July 15-18 British Open today. That of course was widely anticipated to happen at some point, given that this year's Brit is on the Old Course at St. Andrews, where Woods had dominating victories in 2000 and 2005.

::Tony Romo

Cowboys QB Tony Romo missed an opportunity to qualify for a PGA Tour event on Monday in order to be at practice with his Dallas teammates. Romo was scheduled for a 10:57 a.m. tee time in the qualifying for the Byron Nelson Championship, set for this Thursday-Sunday in Irving, Texas. But the time conflicted with a Cowboys OTA session, which is voluntary. Romo chose the Cowboys practice over the golf event. Romo had said last week he hoped to get an afternoon tee time so he could meet both commitments. But organizers of the golf competition were unable to accommodate him. Romo made clear last week he would not be tempted to skip the Cowboys practice.

//Current News

::BP oil leak

After more than three weeks of trying to stop a gushing oil leak in the Gulf of Mexico, BP engineers have achieved some success using a mile-long pipe to capture some of the oil and divert it to a drill ship on the surface some 5,000 feet above the wellhead, company officials said Monday. After two false starts, engineers successfully inserted a narrow tube into the damaged pipe from which most of the oil is leaking. Doug Suttles, BP's chief operating officer, who appeared on several network morning shows Monday, said that the mile-long, 4-inch-wide tube was capturing a little more than 1,000 barrels of oil a day from the blown well and its 21-inch-wide riser pipe, and funneling the oil into the tanker ship.

Figure 27: Subset of Current Events corpus (Cherner, 2010) (Leahy, 2010) (Dewan, 2010)

Data Set 4 consisted of 20 user trials, all resulting in recorded data and completed surveys. Each test subject was a Google Chat user selected from a single existing personal contact list or a student at the UCF ISL. The individuals on the contact list represented a variety of professions, including attorneys, engineers, homemakers, and business analysts. While the chatbot server resided in the ISL at UCF, test subjects were located in various parts of the United States, including California, Virginia, North Carolina, Alabama, South Carolina, Massachusetts, and Florida.

Fourteen (14) male and six female subjects were used. All users had some semblance of knowledge regarding current events. Three undergraduate students were involved, while the remaining 17 people held at least a college Bachelor's degree. Two of the 20 trials involved people holding a Doctoral degree. Spoken accent demographics were not reported because of the non-speech nature of the text-based chatbot. Table 33 gives the Data Set 4 user demographics.

Table 33: Data Set 4 user demographics

Demographic Category	Demographic Sub-category	Number of participants (Percentage of all participants)
Gender	*Female*	6 (30%)
	Male	14 (70%)
Education Level	*Some undergraduate school*	3 (15%)
	Bachelor's degree	6 (30%)
	Some graduate school	9 (45%)
	Doctoral degree	2 (10%)

Data Set 4 Results

The results from the Data Set 4 trials are listed in the next four tables. Table 34 begins by displaying the user survey results.

Table 34: Data Set 4 Survey results

Trial	Statement 1: If I told someone the character in this tool was real they would believe me.	Statement 2: I would be more productive if I had this system in my place of work.	Statement 3: The character on the screen seemed smart.	Statement 4: I felt like I was having a conversation with a real person.	Statement 5: The tool provided me with the information I was looking for.	Statement 6: I found this to be a useful way to get information.	Statement 7: This tool made it harder to get information than talking to a person or using a website.	Statement 8: This does not seem like a reliable way to retrieve information from a database.	Statement 9: This did not feel like a real interaction with another person.
1	4	1	4	3	7	4	4	7	6
2	5	2	5	5	2	2	6	4	6
3	4	3	4	3	4	3	5	5	5
4	1	2	1	1	6	5	4	4	7
5	1	1	1	1	2	3	7	5	7
6	2	2	2	2	1	2	7	5	6
7	1	2	3	1	2	1	7	5	6
8	1	5	2	2	6	4	3	6	5
9	1	1	1	1	1	1	7	7	7
10	1	1	3	1	3	2	3	3	7
11	1	4	1	1	7	5	3	7	6
12	1	1	1	1	2	4	3	3	7
13	1	4	5	1	4	6	4	6	7
14	1	2	1	1	4	1	7	4	7
15	2	3	5	4	5	5	4	3	4
16	1	2	3	4	5	6	5	5	6
17	5	4	6	4	6	6	4	2	4
18	2	1	2	3	3	2	5	4	5
19	4	4	6	4	6	6	3	3	5
20	5	4	4	4	6	6	5	3	6
Average	**2.20**	**2.45**	**3.00**	**2.35**	**4.10**	**3.70**	**4.80**	**4.55**	**5.95**

Twenty (20) surveys were conducted for Data Set 4. From this data, it is apparent that users were not convinced of the realness of the CONCUR Chatbot, as seen in the results for Statements 1, 4, and 9. The perceived usefulness of the agent was lacking as well, as evidenced in Statements 2, 6, 7, and 8. These questions resulted in average normalized assessment scores hovering around 2.5. The most positive survey result was exhibited in Statement 5, where users assessed a mild response of 4.10 to CONCUR's ability as an information source. Overall, the average positive responses for this data set fell in the 2 to 4 range.

Table 35: Data Set 4 Efficiency results

Trial	Total Elapsed Time (min)	Number of User Turns	Number of System Turns	Elapsed Time Per Turn (s)	User Words Per Turn	System Words Per Turn	WER
1	4:50	10	11	9.65	4.50	35.55	0.00%
2	1:06	4	5	5.52	1.75	28.00	0.00%
3	9:29	19	20	9.98	2.95	46.85	0.00%
4	3:18	14	15	4.72	2.07	43.67	0.00%
5	5:00	6	7	16.68	7.67	28.29	0.00%
6	2:48	5	6	11.23	5.40	24.50	0.00%
7	3:06	12	13	5.17	4.75	37.08	0.00%
8	6:00	15	16	8.00	2.73	38.63	0.00%
9	0:55	4	5	4.58	4.25	25.80	0.00%
10	4:28	5	6	17.83	3.00	44.17	0.00%
11	3:21	11	12	6.09	5.45	42.92	0.00%
12	2:05	6	7	6.95	5.17	27.57	0.00%
13	5:50	12	13	9.73	4.25	45.31	0.00%
14	2:22	6	7	7.91	4.17	44.86	0.00%
15	10:46	16	17	13.56	5.06	40.00	0.00%
16	1:42	3	4	11.32	5.67	21.00	0.00%
17	4:19	7	8	12.34	3.57	30.50	0.00%
18	1:21	7	8	3.84	3.14	37.75	0.00%
19	6:39	12	13	11.08	2.67	42.23	0.00%
20	1:40	3	4	11.14	6.33	31.00	0.00%
Average	**4:03**	**8.85**	**9.85**	**9.37**	**4.23**	**35.78**	**0.00%**

Table 35 exhibits efficiency results for the 20 trials that were executed with the Current Events CONCUR Chatbot. Conversation duration averaged approximately four minutes, with about 9 user turns and 10 agent turns per exchange. Each turn averaged almost ten seconds. Since text-based chatting was involved, there was a WER of 0%. Users typed close to four words per turn, while the chatbot delivered an average of about 36 words per response. Unlike the visually-assisted ECA-based systems in Data Sets 1 and 2, the chatbot version of CONCUR was built to textually deliver a set of suggested topic options for the user to select, thus causing an inflation in the system word count.

Table 36: Data Set 4 Quality results

Trial	Number of System Turns	Out-Of-Corpus Misunderstandings	General Misunderstandings	Errors	User Goals	Goals Fulfilled
1	11	2	0	0	7	5
2	5	2	0	0	1	0
3	20	6	0	0	17	11
4	15	1	0	1	11	9
5	7	1	0	1	3	1
6	6	2	0	0	2	0
7	13	3	0	7	9	4
8	16	1	0	5	12	9
9	5	2	0	0	2	0
10	6	2	0	1	1	0
11	12	3	0	1	7	4
12	7	1	0	2	4	1
13	13	4	0	3	10	5
14	7	1	0	2	4	1
15	17	2	0	5	12	8
16	4	0	0	0	1	1
17	8	0	0	0	5	5
18	8	0	0	2	5	3
19	13	1	0	5	7	4
20	4	0	0	1	2	1
Average	**9.85**	**1.70**	**0.00**	**1.80**	**6.10**	**3.60**

According to Table 36's Data Set 4 quality results, the Current Events CONCUR Chatbot averaged 1.70 out-of-corpus misunderstandings, zero general misunderstandings and 1.80 errors per trial. The lack of general misunderstandings is the result of using a text-based system, as no ambiguous requests resulting from flawed ASR were encountered. Hence, any misunderstandings could be attributed to the agent simply not being able to answer the user within its own expertise bounds. An average of 6.10 goals was presented in each exchange, and the chatbot was able to service 3.60 of these goals.

Table 37: Data Set 4 Quantitative Analysis results

Trial	Out-Of-Corpus Misunderstanding Rate	General Misunderstanding Rate	Misunderstanding Rate	Error Rate	Awkwardness Rate	Goal Completion Accuracy	Conversational Accuracy
1	18.18%	0.00%	18.18%	0.00%	0.00%	71.43%	100.00%
2	40.00%	0.00%	40.00%	0.00%	0.00%	0.00%	100.00%
3	30.00%	0.00%	30.00%	0.00%	0.00%	64.71%	100.00%
4	6.67%	0.00%	6.67%	6.67%	6.67%	81.82%	93.33%
5	14.29%	0.00%	14.29%	14.29%	14.29%	33.33%	85.71%
6	33.33%	0.00%	33.33%	0.00%	0.00%	0.00%	100.00%
7	23.08%	0.00%	23.08%	53.85%	53.85%	44.44%	46.15%
8	6.25%	0.00%	6.25%	31.25%	31.25%	75.00%	68.75%
9	40.00%	0.00%	40.00%	0.00%	0.00%	0.00%	100.00%
10	33.33%	0.00%	33.33%	16.67%	16.67%	0.00%	83.33%
11	25.00%	0.00%	25.00%	8.33%	8.33%	57.14%	91.67%
12	14.29%	0.00%	14.29%	28.57%	28.57%	25.00%	71.43%
13	30.77%	0.00%	30.77%	23.08%	23.08%	50.00%	76.92%
14	14.29%	0.00%	14.29%	28.57%	28.57%	25.00%	71.43%
15	11.76%	0.00%	11.76%	29.41%	29.41%	66.67%	70.59%
16	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
17	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	100.00%
18	0.00%	0.00%	0.00%	25.00%	25.00%	60.00%	75.00%
19	7.69%	0.00%	7.69%	38.46%	38.46%	57.14%	61.54%
20	0.00%	0.00%	0.00%	25.00%	25.00%	50.00%	75.00%
Average	**17.45%**	**0.00%**	**17.45%**	**16.46%**	**16.46%**	**48.08%**	**83.54%**

Table 37 shows the quantitative analysis results for Data Set 4. Misunderstandings occurred at a rate of 17.45% (17.45% for out-of-corpus misunderstandings, 0% for general misunderstandings) with an error rate of 16.46%. The agent completed user goals at an average rate of 48.08% with an 83.54% conversational accuracy rate.

Aggregate Results

This section compiles all of the average data set results and presents them in a comparative format. A later section will give a more in-depth evaluation analysis of the data set comparisons.

Survey Results

Table 38 displays the aggregate survey results for the questionnaire-enabled Data Sets 1, 2 and 4. Each column is labeled with an individual survey statement, with the type of quality metric (Naturalness or Usefulness) specified after the statement.

Table 38: Aggregate survey results

Data Set	Statement 1: If I told someone the character in this tool was real they would believe me. (Naturalness)	Statement 2: I would be more productive if I had this system in my place of work. (Usefulness)	Statement 3: The character on the screen seemed smart. (Naturalness)	Statement 4: I felt like I was having a conversation with a real person. (Naturalness)	Statement 5: The tool provided me with the information I was looking for. (Usefulness)	Statement 6: I found this to be a useful way to get information. (Usefulness)	Statement 7: This tool made it harder to get information than talking to a person or using a website. (Usefulness)	Statement 8: This does not seem like a reliable way to retrieve information from a database. (Usefulness)	Statement 9: This did not feel like a real interaction with another person. (Naturalness)
1	3.20	4.10	4.73	4.10	4.57	5.07	4.23	3.17	3.97
2	4.07	4.00	4.97	3.83	4.90	5.43	4.33	3.43	4.30
4	2.20	2.45	3.00	2.35	4.10	3.70	4.80	4.55	5.95

Statements 7 through 9 reflect negatively worded assessments toward the agent. This data was normalized by representing the results of the last three statements as positively worded assessments through a reversal of their scores on the Likert scale (1 becomes 7, 2 becomes 6, 3 becomes 5, 4 remains the same). Table 39 gives the revised version of the survey results.

Table 39: Normalized survey results

Data Set	Statement 1	Statement 2	Statement 3	Statement 4	Statement 5	Statement 6	Statement 7	Statement 8	Statement 9
1	3.20	4.10	4.73	4.10	4.57	5.07	3.77	4.83	4.03
2	4.07	4.00	4.97	3.83	4.90	5.43	3.67	4.57	3.70
4	2.20	2.45	3.00	2.35	4.10	3.70	3.20	3.45	2.05

Table 40 depicts the Naturalness and Usefulness results from averaging the normalized results from Statements 1, 3, 4, and 9 for Naturalness, and averaging the normalized scores from Statements 2, 5, 6, 7 and 8 for Usefulness.

Table 40: Survey results for Naturalness and Usefulness

Data Set	Naturalness	Usefulness
1	4.02	4.47
2	4.14	4.51
4	2.40	3.38

From this data, it can be assessed that the agents in Data Set 1 and Data Set 2, the LifeLike Avatar-based ECAs, both obtained slightly positive responses from users in being both natural and useful. A fairly poor rating of Naturalness was given to the lone text-based agent from Data Set 4, while also achieving a slightly negative assessment of its Usefulness.

Efficiency Metrics Results

Table 41 shows the aggregate efficiency metrics collected from the four data sets. These metrics deal with the measurable, non-quality judgments recorded by each agent.

Table 41: Efficiency metrics

Data Set	Total Elapsed Time (min)	Number of User Turns	Number of System Turns	Elapsed Time Per Turn (s)	User Words Per Turn	Agent Words Per Turn	WER
1	3:36	13.35	14.35	4.15	2.83	28.56	60.85%
2	3:20	10.90	11.90	6.10	4.94	29.09	58.48%
3	2:52	10.10	11.10	6.11	5.02	28.22	0.00%
4	4:03	8.85	9.85	9.37	4.23	35.78	0.00%

The WER results report how well the ASR performed for each agent. Note that Data Sets 3 and 4 did not use speech-based input, so they yielded perfect recognition accuracy. These data reveal that each data set agent conversation was relatively similar in total elapsed times, ranging from nearly three minutes to just over four minutes. This indicates the agents from each data set maintained a fairly consistent three to four minute conversation. The AlexDSS agent in Data Set 1 resulted in a slightly higher average turn count for both the user and the agent over the rest of the field. This is most likely due to the scripted discourse manner in AlexDSS that forces users to completely exhaust a particular topic path to its end. Data Set 4's text-based CONCUR saw a longer amount of time between turns. The text-based nature of this data set probably contributed to the lack of urgency by the user to respond between system responses. Both speech-based agents in Data Set 1 and 2 were virtually equal in recognizing user utterances at a 60% WER. This levels the playing field for any ASR-related metric comparison, since the agents from both data sets suffer from virtually identical misrecognition rates.

Quantitative Analysis Metrics Results

Table 42 gives the aggregate results of the Quantitative Analysis of the Quality metrics. In these metrics, each chat transcript was manually inspected for misunderstandings, erroneous agent responses, and context goal satisfaction. The final two columns, Goal Completion Accuracy and Conversational Accuracy, give a quantitative indicator of each agent's usefulness and naturalness, respectively.

Table 42: Quantitative analysis of quality metrics

Data Set	Out-Of-Corpus Misunderstanding Rate	General Misunderstanding Rate	Misunderstanding Rate	Error Rate	Awkwardness Rate	Goal Completion Accuracy	Conversational Accuracy
1	0.29%	9.51%	9.80%	8.71%	18.22%	63.29%	81.78%
2	6.15%	14.49%	20.64%	21.81%	35.78%	60.48%	63.93%
3	6.77%	7.48%	14.25%	16.68%	24.66%	68.48%	75.34%
4	17.45%	0.00%	17.45%	16.46%	16.46%	48.08%	83.54%

This table recounts how well each agent can handle the user input in terms of minimal conversational awkwardness and maximized assistive utility. The Out-Of-Corpus Misunderstanding Rate assesses the percentage of time the agent must spend to react to a user requesting information that cannot be found in its knowledge base. In these results, it is shown that Data Set 4's chatbot experienced a substantial number of out-of-corpus misunderstandings, while the AlexDSS agent in Data Set 1 saw very little. The explanation of this phenomenon is the simple fact that AlexDSS' highly constrained input expectations from its menu-driven discourse serves as a preventative measure for out-of-corpus information requests. The CONCUR agent, on the other hand, maintains a higher amount of input flexibility, causing users to ask more questions that could potentially be out of the knowledge domain range.

General Misunderstanding Rate addresses the percentage of turns in which the agent is presented with situations that it could not handle, most often because of garbled ASR inputs or erratic user speech, such as stalling. The conversation agent in Data Set 4 did not have to deal with these issues, hence its lack of general misunderstandings. The Data Set 3 CONCUR Chatbot also lacked ASR-related errors, but it still fell victim to user input errors because of its use of Data Set 2 inputs.

Error Rate describes the percentage of turns the agent returns a nonsensical response. The CONCUR agents all had similar Error Rates, while the AlexDSS ECA in Data Set 1 was the least error-prone because of its menu-driven nature. The dialog openness of the CONCUR system plays a part in causing erroneous system responses because the presence of specific QA information requests. This factor deals with the idea that users want very specific answers to questions, and it is discussed in further depth later in this chapter.

In terms of usefulness, the Goal Completion Accuracy metric indicates how effective an agent can service users' information requests. While all of the NSF I/UCRC corpus-based agents (Data Sets 1, 2 and 3) were able to complete over 60% of their users' goals, the Current Events CONCUR Chatbot in Data Set 4 was just under 50% for Goal Completion Accuracy. A later portion of this chapter discusses the cause of this phenomenon.

Awkwardness Rate and Conversational Accuracy give a quantitative indication on the naturalness of the agent's dialog. Essentially, the Conversational Accuracy tells what percentage of the time the conversation agent gave an answer that can be perceived as natural. The Awkwardness Rate is simply the percentage of unnatural responses. While each agent was able to demonstrate better than 60% Conversational Accuracy, the CONCUR ECA in Data Set 2 was far less conversationally accurate than the agents in Data Sets 1, 2 and 4. As with the discrepancy in Goal Completion Accuracy, this Conversational Accuracy observation is addressed later on in this chapter.

This previous section gave an objective presentation of the data sets. The next section delivers an analysis of the results in the light of the three premises of this dissertation: overcoming ASR limitations, knowledge management domain independence, and dialog domain openness. Additionally, a survey-based analysis is presented to establish the current expected state of user perception for ECA technology.

Evaluation Analysis

Evaluating CONCUR consisted of four major exploratory themes, each with its own set of specific questions. The first theme deals with establishing a survey-based baseline of the dialog manager's naturalness and usefulness. The second theme deals with CONCUR's resilience to ASR-related errors. The third investigation discusses its domain-independent knowledge capabilities. The fourth issue involves CONCUR's open dialog discourse characteristics. Table 43 gives the list of specific questions involved for each of the aforementioned evaluation themes.

Table 43: Evaluation Analysis question list

Evaluation Theme		**Question**	
A	Survey-based Analysis	1	What are the expectations of naturalness and usefulness for CONCUR?
		2	Did users rate the CONCUR ECA differently from the AlexDSS ECA?
		3	Did users rate the CONCUR ECA differently from the CONCUR Chatbot?
B	ASR Resilience	1	Can a speech-based CONCUR ECA's Goal Completion Accuracy measure up to other conversation agents?
		2	Can a speech-based CONCUR ECA's Goal Completion Accuracy measure up to the AlexDSS ECA?
		3	Does improving WER affect CONCUR's Goal Completion Accuracy?
		4	Can a speech-based CONCUR ECA's Conversational Accuracy measure up to other conversation agents?
		5	Can a speech-based CONCUR Avatar's Conversational Accuracy measure up to the AlexDSS ECA?
		6	Does improving WER affect CONCUR's Conversational Accuracy?
C	Domain Independence	1	Can CONCUR provide a quick method of providing agent knowledge?
		2	Can CONCUR maintain Conversational Accuracy after changing to a different corpus?
		3	Can CONCUR maintain Goal Completion Accuracy after changing to a different corpus?
D	Open Dialog	1	Does CONCUR allow users more response flexibility than AlexDSS?
		2	Does CONCUR allow users more response flexibility than a generic question-answer agent?
		3	Are users more verbose with CONCUR than with AlexDSS?

The original impetus of the data sets in this dissertation was to weave a story about building a dialog manager that could overcome ASR limitations, provide domain-independent knowledge management, and support open dialog. Comparisons between these data sets helped to answer the questions found in Table 43, giving insight on how CONCUR was able to handle ASR errors, domain-independence and open dialog. This section sheds insight on CONCUR's usefulness and naturalness in respect to these investigations.

Evaluation Theme A: Survey-based Analysis

The Survey-based Analysis deals with processing the user questionnaire results into a set of conclusions regarding the perception of CONCUR's effectiveness in the opinion of human test subjects. These conclusions are based on comparisons between survey-based agent assessments. This section answers the following questions pertaining to analyzing the user survey data:

1. What are the expectations of naturalness and usefulness for CONCUR?
2. Did users rate the CONCUR ECA differently from the AlexDSS ECA?
3. Did users rate the CONCUR ECA differently from the CONCUR Chatbot?

Question A.1: What are the expectations of naturalness and usefulness for CONCUR?

The first evaluation question deals with setting a baseline of user perception. The current state of conversation agent technology has not fully amounted to the idealistic notions of completely human-like dialog exchanges. By comparing the survey data from human users of modern conversation agents, the level of expectation in which a contemporary research effort should live up to can be established. This is the impetus for the first question pertaining to the human-based assessment of CONCUR. Table 44 compares the user assessments between the CONCUR ECA, and two contemporary ECA efforts, Amani and Hassan. (Gandhe et al, 2009)

Table 44: ECA user assessment comparison

	Data Set 2: CONCUR ECA	Amani (Gandhe et al, 2009)	Hassan (Gandhe et al, 2009)
Naturalness User Rating	4.14	3.09	3.55
Usefulness User Rating	4.51	3.24	4.00

To put CONCUR's performance in perspective, these two recent dialog agents from the research literature were selected for comparison. In this table, the Data Set 2 user rating statistics in Naturalness and Usefulness surpass those of its ECA peers, Amani and Hassan. (Gandhe et al, 2009) While the instrument used to determine Amani and Hassan's Naturalness and Usefulness were not identical to that of CONCUR's, the general method of using a Likert scoring assessment based on a 1 to 7 scale was used for all three agents. Specifically, Amani and Hassan's Naturalness assessment statement was "Taken as a whole, Amani/Hassan was a human-like conversation partner," and their Usefulness assessment statement was "In general, Amani/Hassan responded appropriately to what I was saying." (Gandhe et al, 2009) The questionnaire results saw that the CONCUR ECA averaged approximately a full point higher in both categories over Amani and about a half point higher than Hassan.

This analysis dealt with CONCUR's human assessment comparison with contemporary ECA work. The next question in this survey-based evaluation compares the user perceptions of the CONCUR ECA with that of a similar effort, the AlexDSS ECA.

<u>Question A.2: Did users rate the CONCUR ECA differently from the AlexDSS ECA?</u>

This question compares user assessments of naturalness and usefulness between two similar assistive conversation agents, the CONCUR ECA and the AlexDSS ECA. While each system shares the same avatar-based user interface presentation and the same domain expertise, they differ in their conversational discourse styles. CONCUR uses a context-based approach, while AlexDSS employs a menu-driven discourse manner. Table 45 depicts the comparison of survey-based assessments in naturalness and usefulness between the AlexDSS ECA and the CONCUR ECA.

Table 45: AlexDSS and CONCUR user assessment comparison

Agent	Data Set 1: AlexDSS ECA	Data Set 2: CONCUR ECA
Naturalness User Rating	4.02	4.14
Usefulness User Rating	4.47	4.51

According to this table, both avatar-based systems in the speech-based data sets established similar scores in Naturalness and Usefulness. While Data Set 1 recorded lower average ratings than Data Set 2, the differences were deemed statistically insignificant using a two-tailed Welch's t-test with a confidence interval of 95%. The p-values for comparing the Naturalness and Usefulness ratings were 0.687 and 0.886, respectively.

This question compared two avatar-based systems with differing discourse styles. The result of this analysis was that both the CONCUR ECA and the AlexDSS ECA were assessed similarly for Naturalness and Usefulness by its user base. The next question offers insight on how the text-based CONCUR agent was assessed over the speech-based ECA version.

Question A.3: Did users rate the CONCUR ECA differently from the CONCUR Chatbot?

This question compares the user perception between a full-blown ECA environment and a keyboard-based chatbot. Table 46 shows the differences in user assessment results between the ECA-based agents and the CONCUR Chatbot.

Table 46: ECA and Chatbot user assessment comparison

Survey Statement	Data Set 1: AlexDSS ECA	Data Set 2: CONCUR ECA	Data Set 4: CONCUR Chatbot
Statement 1: If I told someone the character in this tool was real they would believe me.	3.20	4.07	2.20
Statement 2: I would be more productive if I had this system in my place of work.	4.10	4.00	2.45
Statement 3: The character on the screen seemed smart.	4.73	4.97	3.00
Statement 4: I felt like I was having a conversation with a real person.	4.10	3.83	2.35
Statement 5: The tool provided me with the information I was looking for.	4.57	4.90	4.10
Statement 6: I found this to be a useful way to get information.	5.07	5.43	3.70
Statement 7: This tool made it harder to get information than talking to a person or using a website.	4.23	4.33	4.80
Statement 8: This does not seem like a reliable way to retrieve information from a database.	3.17	3.43	4.55
Statement 9: This did not feel like a real interaction with another person.	3.97	4.30	5.95
Usefulness	**4.02**	**4.14**	**2.40**
Naturalness	**4.47**	**4.51**	**3.38**

According to this table, the chatbot attained less positive assessments for all but two survey statements. For the fifth and seventh statements, the CONCUR Chatbot was assessed similarly to that of the ECA-based systems. The Usefulness and Naturalness ratings for the Data Set 4 chatbot were consistently lower than its avatar-based counterparts. In general, however, users heavily preferred the ECA environment over the chatbot interface.

Analysis Summary

The survey data dictated that users perceived the CONCUR ECA and the AlexDSS ECA to be more natural and more useful than a collection of comparable ECA systems, such as Amani and Hassan. (Gandhe et al, 2009) Additionally, the data exhibited the preference of an ECA-based interface over a text-based chatbot. The most pertinent finding from the survey data, however, is the fact that test subjects rated both the CONCUR ECA and the AlexDSS ECA equally in terms of Naturalness and Usefulness. This conclusion establishes a sense of equal footing between these two systems as they undergo further comparison in the next section, which pertains to ASR Resilience.

Evaluation Theme B: ASR Resilience

The evaluation theme of ASR Resilience describes how well CONCUR can operate as an assistive conversation agent in the presence of speech recognition inaccuracies. The impetus for this assessment is to establish the fact that CONCUR provides effective dialog management in terms of usefulness and naturalness. Conclusions regarding ASR Resilience were made using data set comparisons between different agents. The following question set was used to perform this evaluation:

1. Can a speech-based CONCUR ECA's Goal Completion Accuracy measure up to other conversation agents?
2. Can a speech-based CONCUR ECA's Goal Completion Accuracy measure up to the AlexDSS ECA?
3. Does improving WER affect CONCUR's Goal Completion Accuracy?

4. Can a speech-based CONCUR ECA's Conversational Accuracy measure up to other conversation agents?
5. Can a speech-based CONCUR Avatar's Conversational Accuracy measure up to the AlexDSS ECA?
6. Does improving WER affect CONCUR's Conversational Accuracy?

Question B.1: Can a speech-based CONCUR ECA's Goal Completion Accuracy measure up to other conversation agents?

This question serves to establish the baseline for Goal Completion Accuracy in the field of speech-based ECA field. Because of the scarcity of quantitative data related to conversation agent utility, very few research efforts have made available such information. Table 47 compares the Goal Completion Accuracy of the Data Set 2 NSF I/UCRC CONCUR ECA with that of an existing ECA project that listed its average goal completion competency, the Virtual Kyoto agent. (Misu and Kawahara, 2007)

Table 47: Goal Completion Accuracy analysis between CONCUR and Virtual Kyoto

	Data Set 2: CONCUR ECA	**Virtual Kyoto (Misu and Kawahara, 2007)**
Average WER	58.48%	29.40%
Goal Completion Accuracy	60.48%	61.40%

According to this table, the Goal Completion Accuracy of the NSF I/UCRC CONCUR ECA is similar to that of the Virtual Kyoto agent (Misu and Kawahara, 2007), despite the presence of twice as many word errors. This comparison supports the idea that CONCUR can provide a relatively adequate speech-based solution for user goal completion, even under the condition of a high WER. The next question deals with the CONCUR ECA's ability to match Goal Completion Accuracies with its AlexDSS counterpart.

Question B.2: Can a speech-based CONCUR ECA's Goal Completion Accuracy measure up to the AlexDSS ECA?

To assess CONCUR's ability to handle user goal completion in lieu of ASR limitations, the Goal Completion Accuracy results from Data Set 1 were compared against those from Data Set 2. This analysis establishes the goal completion tendencies between the speech-based versions of AlexDSS and CONCUR. Table 48 gives the data points related to Goal Completion Accuracy between the AlexDSS ECA and the CONCUR ECA.

Table 48: Goal Completion Accuracy analysis between CONCUR and AlexDSS

		Data Set 1: AlexDSS ECA	Data Set 2: CONCUR ECA
Efficiency Metrics	**WER**	60.85%	58.48%
Quantitative Analysis	**Out-of-Corpus Misunderstanding Rate**	0.29%	6.15%
	Goal Completion Accuracy	63.29%	60.48%

From this table, it was concluded that a speech-based CONCUR ECA's Goal Completion Accuracy measures up to AlexDSS ECA with similarly high WER. The differences between their Goal Completion Accuracies and WERs were statistically insignificant, whose two-tailed Welch's t-tests, with 95% confidence intervals, yielded p-values of 0.654 and 0.734, respectively. To be noted from the table, however, is the large difference in Out-Of-Corpus Misunderstanding Rates. This can be attributed to the menu-driven input of the AlexDSS agent, whose discourse speech expectations rarely allow out-of-corpus flexibility. The next question in this ASR Resilience Analysis deals with the effect of WER on CONCUR's Goal Completion Accuracy.

<u>Question B.3: Does improving WER affect CONCUR's Goal Completion Accuracy?</u>

It was speculated that CONCUR's Goal Completion Accuracy could be enhanced by improving the WER. This conjecture was investigated by comparing the results of Data Set 2 with Data Set 3, where the same user information requests were applied to a high WER system and a low WER system. Table 49 compares the data points related to Goal Completion Accuracy between the CONCUR ECA and the CONCUR Chatbot.

Table 49: CONCUR Goal Completion Accuracy analysis

		Data Set 2: CONCUR ECA	Data Set 3: CONCUR Chatbot
Efficiency Metrics	**WER**	58.48%	0.00%
Quantitative Analysis	**Out-of-Corpus Misunderstanding Rate**	6.15%	6.77%
	Goal Completion Accuracy	60.48%	68.48%

The difference in Goal Completion Accuracy between the speech-based CONCUR and the text-based CONCUR is not statistically significant, according to a p-value of 0.253 under a two-tailed Welch's t-test with a 95% confidence interval. Furthermore, a similar t-test shows that the Out-Of-Corpus Misunderstanding Rate difference is also statistically insignificant, yielding a p-value of 0.838. This table concludes that improved WER does not increase CONCUR's goal completion accuracy. The reason for this phenomenon is that the number of out-of-corpus requests did not change, meaning that no new user goals were identified or corrected with the better recognition rate.

So far, this section has detailed CONCUR's relationship between WER and Goal Completion Accuracy. This adheres to the idea of ASR Resilience for the sake of conversation agent utility. While CONCUR was able to establish an approximate average of 60% user goal completion, a baseline level also shared by both its comparison agents Virtual Kyoto agent (Misu and Kawahara, 2007) and the AlexDSS ECA, an improvement in WER did not affect its Goal Completion Accuracy. The next set of questions deals with the relationship between WER and Conversational Accuracy.

Question B.4: Can a speech-based CONCUR ECA's Conversational Accuracy measure up to other conversation agents?

This question establishes CONCUR's relative baseline expectation of Conversational Accuracy in the field of ECA design. What Conversational Accuracy measures is the percentage of turns that the system reacts to the user with a non-awkward response. Awkward responses encompass two major categories: errors and general misunderstandings. Erroneous output simply refers to an inappropriate answer to a user's information request. A general misunderstanding is defined as a response in which the agent must make a plea of ignorance for the sake of advancing the conversation and not because it simply does not know the answer. The latter describes an out-of-corpus misunderstanding.

The sparseness of evaluation data in the conversation agent realm has made it a difficult task to locate Conversational Accuracy performance metrics for comparison purposes. Despite its existence as a text-based system, the TARA (Schumaker et al, 2007) chatbot was able to provide such information. Table 50 compares average WER and average Conversational Accuracy between CONCUR and TARA.

Table 50: Conversational Accuracy analysis between CONCUR and TARA

	Data Set 2: CONCUR ECA	TARA (Schumaker et al, 2007)
Average WER	58.48%	0.00%
Conversational Accuracy	63.93%	62.06%

From this table, it was concluded that the text-based TARA (Schumaker et al, 2007) produced a similar Conversational Accuracy rating to the speech-based CONCUR agent. While TARA (Schumaker et al, 2007) may have used different methods to measure its Conversational Accuracy, the general idea of determining how often an agent replies with an appropriate response is captured in this comparison. Specifically, in evaluating TARA, a 68.4% and a 39.6% "Average Response Appropriateness" was assessed for its 888 non-domain conversation and 250 domain-based conversation actions, respectively. (Schumaker et al, 2007) The mean of these "Average Response Appropriateness" factors resulted in a 62.06% Conversational Accuracy.

Having established its relative place in the Conversational Accuracy category of conversation agent design, the next question deals with comparing CONCUR with its AlexDSS counterpart in this particular metric.

<u>Question B.5: Can a speech-based CONCUR ECA's Conversational Accuracy measure up to the AlexDSS ECA?</u>

This assessment investigates the differences in Conversational Accuracy between the CONCUR ECA and the AlexDSS ECA. Such an evaluation demonstrates the CONCUR ECA's performance in minimizing response awkwardness in relation to a comparable entity. Table 51 gives the Conversational Accuracy analysis between the two speech-based agents.

Table 51: Conversational Accuracy analysis between CONCUR and AlexDSS

		Data Set 1: AlexDSS ECA	Data Set 2: CONCUR ECA
Efficiency Metrics	**WER**	60.85%	58.48%
Quantitative Analysis	**General Misunderstanding Rate**	9.51%	14.49%
	Error Rate	8.71%	21.81%
	Conversational Accuracy	81.78%	63.93%

From this table, it was determined that a speech-based CONCUR ECA's Conversational Accuracy does not measure up to an AlexDSS ECA with a similarly high WER. Specifically, the AlexDSS ECA is nearly 20% more conversationally accurate than the CONCUR ECA under the same approximate 60% WER. Further investigation into response awkwardness calls for an exploration of its two causes: general misunderstandings and errors.

A preliminary breakdown of the errors and general misunderstandings between the two systems shows that CONCUR produces more of each type of agent response awkwardness, thus impacting the overall Conversational Accuracy. A two-tailed Welch's t-test with a 95% confidence interval results in a p-value of 0.085 between the two General Misunderstanding Rate averages. This indicates that the difference in general misunderstandings is statistically insignificant. A similar t-test for Error Rate concluded with a p-value of 0.0007, meaning the difference in errors was indeed significant. This discrepancy in error count was attributed to the added specific QA requests experienced by CONCUR, a trend not commonly encountered by menu-driven discourse models, such AlexDSS.

A specific QA request refers to a user's desire to find out a particular piece of data, rather than a broad discussion of a general topic. An example of a specific QA request would be "When is the deadline for the application?" A general information request under the same context would be "Please tell me about the deadline." An agent response of "There are two deadlines every year," would satisfy the latter information request but would result in an error for the former QA inquiry. A menu-driven agent, such as AlexDSS, rarely allows for the free response style of QA requesting. Open dialog systems, such as CONCUR, encourage added flexibility in user responses. CONCUR, however, was not designed to perform specific question answering, as high WERs would hinder its ability to recognize the fine-grain inquiry words needed for effective data retrieval. This lack of specific QA constructs ultimately surfaced as erroneous output responses, directly causing a negative impact on Conversational Accuracy.

This analysis question relayed the idea that the speech-based CONCUR ECA's Conversational Accuracy could not match up to that of the AlexDSS ECA. It was speculated that the errors from users' insistence on specific QA requests was the main cause of this difference, where the influx of such requests is the result of CONCUR's dialog openness. Thus, it could be argued that open dialog only leads to more erroneous agent responses, worsening its Conversational Accuracy. The prospect of allowing such rich user input, however, provides promise in adding tremendous naturalness to an ECA experience. The next question attempts to mitigate the error-prone risks of CONCUR's open dialog by investigating whether an improved WER can help improve CONCUR's Conversational Accuracy.

Question B.6: Does improving WER affect CONCUR's Conversational Accuracy?

This question examines the effect of improved WER on CONCUR's Conversational Accuracy. Minimizing the WER from Data Set 2 would give insight on the causes of CONCUR's conversational inaccuracies. This decrease in word misrecognitions was reflected in Data Set 3's CONCUR Chatbot, serving as the baseline for a perfect speech recognition system. The same metrics for the previous question (WER, General Misunderstanding Rate, Error Rate and Conversational Accuracy) were examined for the ECA and chatbot versions of CONCUR, both of which used the same NSF I/UCRC corpus. This analysis is featured in Table 52.

Table 52: CONCUR Conversational Accuracy analysis

		Data Set 2: CONCUR ECA	Data Set 3: CONCUR Chatbot
Efficiency Metrics	**WER**	58.48%	0.00%
Quantitative Analysis	**General Misunderstanding Rate**	14.49%	7.48%
	Error Rate	21.81%	16.68%
	Conversational Accuracy	63.93%	75.31%

From this table, it was concluded that the improvement in WER also increased the Conversational Accuracy of CONCUR. It is evident that a decrease in General Misunderstandings occurred, while the Error Rate did not undergo a statistically significant change, a difference that yielded a p-value of 0.233 from a two-tailed Welch's t-test with 95% confidence interval. Hence, the error count remained the same. It was conjectured that this effect was caused by the same number of specific QA requests found in the Data Set 2 transcripts which carried over to the Data Set 3 trials.

A similar t-test was performed for the General Misunderstanding Rate, resulting in the detection of a statistically significant change with a p-value of 0.01. This drop in general misunderstandings, most likely from a drop in misheard user requests, coincided with an increase in Conversational Accuracy for the chatbot-based CONCUR. It was surmised that the improved WER eliminated any conversational awkwardness that normally occurs with a speech-based system when the agent is trying to negotiate a clarification of the user utterance. This situation is often characterized as a series of "Can you repeat what you just said?" questions from the agent. By eliminating this clarification negotiation through a better WER, the overall Conversational Accuracy increases.

As a conclusion to this analysis, it was found that an improvement in CONCUR's WER resulted to an approximate 10% gain in Conversational Accuracy, a trend strongly tied to a diminished general misunderstanding count.

Analysis Summary

The basic finding here is that CONCUR was able to achieve a similar Goal Completion Accuracy with that of the AlexDSS system while under the same 60% WER conditions from the ASR system. Additionally, CONCUR attained a similar Goal Completion Accuracy with that of Virtual Kyoto (Misu and Kawahara, 2007), at twice the WER. A similar analysis was done for CONCUR's Conversational Accuracy, which saw nearly identical results with that of the TARA (Schumaker et al, 2007) chatbot. Hence, the data set results showed that CONCUR could be used to perform competent speech-based HCI conversation, even when compared to the standards of modern-day ECAs.

In further analysis of CONCUR's Conversational Accuracy, however, it was found that the AlexDSS ECA provided a better solution for minimizing awkward agent responses. Errors related to specific QA requests negatively affected CONCUR's conversational abilities as an open dialog agent. Only when CONCUR existed in text-based format could its Conversational Accuracy be improved. This begs the question of why one would choose the CONCUR dialog manager over the AlexDSS dialog management system.

The main advantages CONCUR has over AlexDSS are two features rooted in its context-centric methods: 1) domain-independent knowledge management, and 2) open dialog conversational discourse model. CONCUR's modularized knowledge management hastens the agent production process and allows for a more flexible platform for ECA development. AlexDSS's expert system knowledge base requires a more time-consuming effort in modeling agent expertise. Additionally CONCUR is based on an open dialog method, while AlexDSS aligns to a more direct, automated phone operator style. This *closed* discourse style is often characterized by expecting a very narrow set of user inputs, causing user utterance inflexibility. CONCUR's more flexible response acceptance provides a more natural conversational experience. The next two evaluation themes discuss CONCUR's capabilities of domain independence and dialog openness.

Evaluation Theme C: Domain Independence

The modularized domain expertise within CONCUR manifests itself as a domain-independent knowledge manager. Domain independence is an important feature for knowledge management because it allows for a quick, interchangeable solution for agent expertise. This evaluation theme utilized the acquired data sets to confirm that this feature is indeed a functional characteristic possessed by the CONCUR dialog manager. A comparison study between different agents was designed to validate this style of knowledge management. The following questions were answered to affirm CONCUR's domain independence:

1. Can CONCUR provide a quick method of providing agent knowledge?
2. Can CONCUR maintain Conversational Accuracy after changing to a different corpus?
3. Can CONCUR maintain Goal Completion Accuracy after changing to a different corpus?

Question C.1: Can CONCUR provide a quick method of providing agent knowledge?

An important aspect of domain independence is the amount of development time for creating a new agent knowledge base. The impetus of this question is to establish CONCUR's relative place in the realm of agent knowledge turnover time. Table 53 compares the development time for creating new domain expertise for a CONCUR agent versus the development time for other dialog systems. The idea here is to provide a quantitative example of how CONCUR's corpus-based knowledge management method is advantageous over existing knowledge acquisition models.

Table 53: Dialog System knowledge development time

Dialog System	Method	Turnover Time
CONCUR	Corpus-based	3 Days
Marve (Babu et al, 2006)	Wizard-of-Oz	18 Days
Amani (Gandhe et al, 2009)	Question-Answer Pairs	Weeks
AlexDSS (Sherwell et al, 2005)	Expert System	Weeks
Sergeant Blackwell (Robinson et al, 2008)	Wizard-of-Oz	7 Months
Sergeant Star (Artstein et al, 2009)	Question-Answer Pairs	1 Year
HMIHY (Béchet et al, 2004)	Hand-modeled	2 Years
Hassan (Gandhe et al, 2009)	Question-Answer Pairs	Years

From this table, it is easy to see that CONCUR's domain-independent knowledge management emphasizes its advantage as a rapid-prototyping tool for ECA dialog creation. The main issue here is that other knowledge management styles require a time-consuming manual modeling of response data. CONCUR, on the other hand, uses a very basic encyclopedia-like corpus structure to provide agent expertise. This allows for a simple exercise of essentially cutting and pasting information into an ECA dialog. Thus, CONCUR's Knowledge Manager enables a shortened knowledge development turnover time as compared to other conversation agent knowledge management systems. The next question validates CONCUR's domain independence capabilities in terms of maintaining Goal Completion Accuracy.

Question C.2: Can CONCUR maintain Goal Completion Accuracy after changing to a different corpus?

This question investigates the Goal Completion abilities of CONCUR for differing domains. The effect of altering domain expertise upon Goal Completion Accuracy was examined by comparing the results from different CONCUR Chatbots under different corpora. Such an analysis gives insight into whether Goal Completion is impacted by different corpus designs, a feature enabled by a domain-independent knowledge manager. Table 54 analyzes the Goal Completion Accuracy for two chatbots with different expertises: the NSF I/UCRC CONCUR Chatbot and the Current Events Chatbot.

Table 54: CONCUR Chatbot Goal Completion Accuracy analysis

		Data Set 3: NSF I/UCRC Chatbot	Data Set 4: Current Events Chatbot
Quantitative Analysis	**Out-Of-Corpus Misunderstanding Rate**	6.77%	17.45%
	Goal Completion Accuracy	68.48%	48.08%

According to this table, it is quite noticeable that an approximate 20% drop in Goal Completion Accuracy was experienced when shifting from the NSF I/UCRC expertise to the Current Events knowledge. Thus, it was observed that CONCUR's goal completion accuracy did not remain consistent after a change to the more generalized Current Events domain corpus.

Also evident from Table 54 is that changing domain expertise nearly tripled the amount out-of-corpus requests, a possible explanation for the decreased goal completion. The finding here is that by opening up the user's perception of what an agent may know through the widening of the domain corpus scope, there is a misconception that the chatbot knows more than what is contained in its knowledge base. This assumption of agent omniscience causes the user to ask for information requests that result in added out-of-corpus misunderstandings, causing inflation in the Out-Of-Corpus Misunderstanding Rate. By augmenting this rate, CONCUR is less capable of completing its user goals, thus resulting in a lower Goal Completion Accuracy.

Thus, it was concluded that the Goal Completion Accuracy for CONCUR is affected by the design of the domain corpus. For a specifically-tailored domain, such as the NSF I/UCRC, the agent was able to maintain a goal completion rate of well over 50%. Switching the expertise to a more generalized knowledge base about Current Events, however, caused CONCUR to complete user goals at a sub-50% rate. While the utility of CONCUR was investigated in this question, the next question deals with whether its conversational abilities are affected by a domain corpus change.

Question C.3: Can CONCUR maintain Conversational Accuracy after changing to a different corpus?

This question addresses the effect of a new domain corpus on Conversational Accuracy. Such an analysis confirms whether domain independence can co-exist with CONCUR's capability to maintain a conversationally adequate response system. This examination of Conversational Accuracy in lieu of a corpus change was performed by comparing a pair of CONCUR Chatbots with different expertise schemes, an NSF I/UCRC domain and a Current Events domain. Table 55 gives an analysis of each agent's Conversational Accuracy.

Table 55: CONCUR Chatbot Conversational Accuracy analysis

		Data Set 3: NSF I/UCRC Chatbot	Data Set 4: Current Events Chatbot
Quantitative Analysis	General Misunderstanding Rate	7.48%	0.00%
	Error Rate	16.68%	16.46%
	Conversational Accuracy	75.34%	83.54%

According to this table, both chatbots produced similar Conversational Accuracy results. With a p-value of 0.099 using a two-tailed Welch's t-test with a 95% confidence interval, the difference between these two rates was deemed statistically insignificant. The general misunderstandings for the NSF I/UCRC CONCUR Chatbot are a result of the speech-based transcripts used for Data Set 3. These misunderstandings associated with misconstrued ASR results do not come into play for Data Set 4's Current Events Chatbot. Moreover, the Error Rates for both systems are nearly similar, a phenomenon caused by the specific QA request effect mentioned earlier in this chapter. From this table, it can be observed that no drastic changes in Conversational Accuracy were experienced as a result of altering the agent's domain expertise. Hence, after changing to a general domain corpus, CONCUR is capable of maintaining its conversational accuracy.

Analysis Summary

The concept to take away from this Domain Independence Analysis is the fact that the CONCUR agent infrastructure can provide a usable and functionally acceptable dialog management experience regardless of any changes to its corpus data. Nevertheless, the quality of the domain corpus does have an impact on the Goal Completion Accuracy of the system. Specifically, CONCUR knowledge engineers should take into consideration the user needs and expectations of the agent before deploying an ECA. In terms of providing a conversationally accurate dialog, a comparison of Data Set 3's results with those of Data Set 4 saw that an equally functional CONCUR-based conversation agent could be made without incurring a significant drop in Conversational Accuracy.

A preliminary study comparing the production times for developing knowledge bases for different dialog systems was made, which concluded that CONCUR's development turnover rate that was exceptionally faster than rest of the field. This expedited expertise modeling method was made possible through CONCUR's domain-independent Knowledge Manager. While this evaluation theme emphasized the advantages of domain-independent knowledge management, the next theme addresses another context-driven feature of CONCUR, its dialog openness.

Evaluation Theme D: Open Dialog

Dialog openness remains the final theme demonstrated by CONCUR. This feature addresses the idea that users have full flexibility in the types of utterances they can give to a conversation agent. Such a response policy enables a higher sense of conversational naturalness, as the common alternative would be a less natural, menu-driven discourse style, as seen in the AlexDSS dialog manager. This latter form of input highly constrains the user's utterance space to a limited number of options, certainly not reflective of a natural conversation flow. CONCUR gets rid of this expectation-driven method, by adopting a more open dialog discourse approach.

Assessing the open dialog trait took into consideration both algorithmic analysis and data set comparison. The following questions were answered to confirm CONCUR's dialog openness:

1. Does CONCUR allow users more response flexibility than AlexDSS?
2. Does CONCUR allow users more response flexibility than a generic question-answer agent?
3. Are users more verbose with CONCUR than with AlexDSS?

Question D.1: Does CONCUR allow users more response flexibility than AlexDSS?

To address the issue of response flexibility, cyclomatic, or conditional, complexity (McCabe, 1976) was computed for the CONCUR and AlexDSS agents. This complexity gives an analytical treatment on the magnitude of decision points that can occur in software. (McCabe, 1976) In terms of dialog agent software, this translates to the number of valid utterances that the system can expect the user to say. Thus, it is surmised that greater cyclomatic complexity is indicative of more dialog openness.

Cyclomatic complexity is computed through a topological analysis of a software system's source code control flow diagram. This analysis relies on counting the number of nodes, edges, and connected components in the control flow graph. Figure 28 and Figure 29 depict highly simplified models of the AlexDSS and CONCUR source code control flows, respectively. These graphs assume that X number of possible topics or contexts that exist in the knowledge base, with a depth of Y statements per topic or context, where both X and Y must equal one or greater.

The idea here is to ensure that the response corpus sizes of both dialog manager models are of equal magnitudes. In this case, each knowledge base size is $X \times Y$ data items. Additionally, for simplicity's sake, all topic or context transitions are represented as a single path for AlexDSS. This is a result of its characteristically linear method of information deployment. Hence, each topic or context node serves as an agent response action rather than an actual decision node in the discourse model.

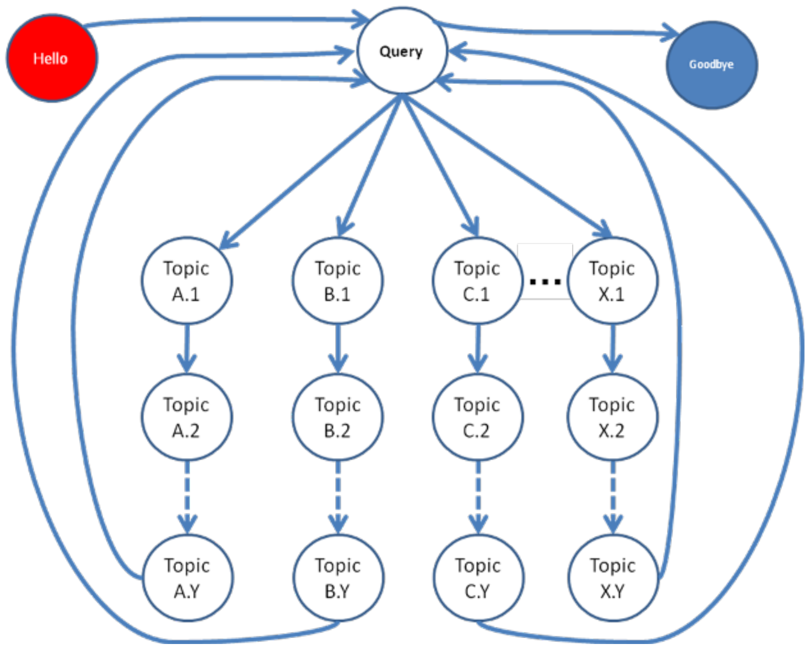


Figure 28: Control flow graph of AlexDSS

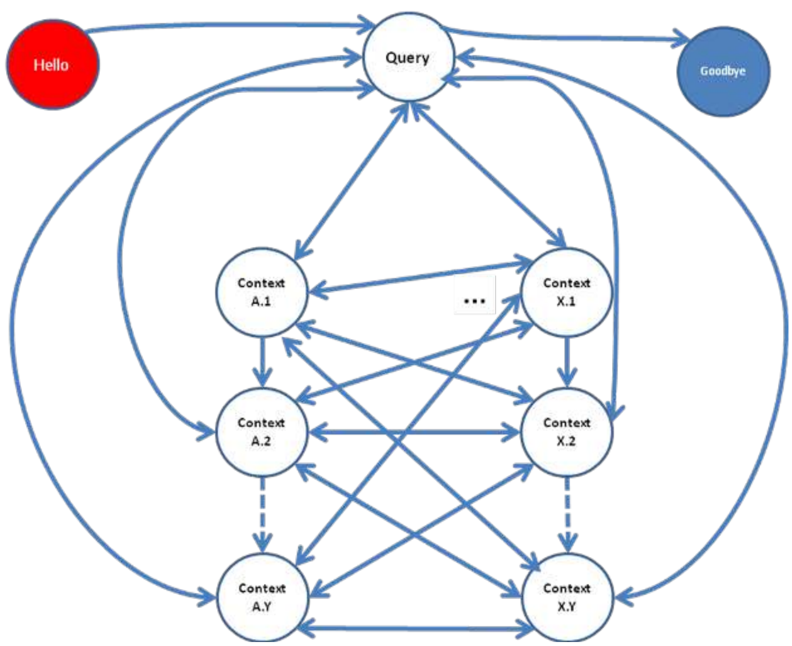


Figure 29: Control flow graph of CONCUR

Upon visual inspection of each dialog manager's control flow graphs, it is immediately apparent that the CONCUR software is rife with decision points. This reflects the aspect of CxBR that facilitates traversal flexibility between contexts. To quantitatively analyze these graphs, a mathematical treatment of *cyclomatic complexity* must be introduced.

The cyclomatic complexity, *M*, counts the number of linearly independent paths within the code. (McCabe, 1976) Effectively, *M* signifies the number of different ways to accomplish any single conversational goal from "Hello" to "Goodbye." *M* is a function of *E* edges, *N* nodes, and *P* connected components:

$$M = E - N + 2P\,. \tag{2}$$

Given the assumptions of *X* topics, with a topic depth of *Y* for the knowledge domain of AlexDSS and CONCUR, the cyclomatic complexities of the dialog system software models in Figure 28 and Figure 29 are as follows:

$$M_{AlexDSS} = X + 1\,. \tag{3}$$

$$M_{CONCUR} = X^2Y^2 - XY^2 + 3XY - X + 2\,. \tag{4}$$

From the equations in (3) and (4), it is evident that the polynomial expression found for CONCUR's *M* is more conditionally complex than the AlexDSS dialog manager. This can be further verified by setting the topic depth of *Y* to 3, meaning the agent is capable of deploying three statements pertaining to a certain topic:

$$M_{AlexDSS} = X + 1\,. \tag{6}$$

$$M_{CONCUR} = 9X^2 - X + 2\,. \tag{7}$$

Plotting these equations gives the curves found in Figure 30.

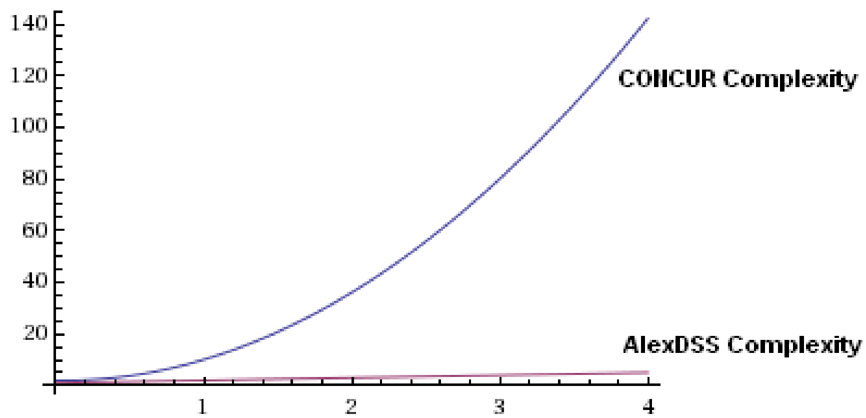


Figure 30: AlexDSS versus CONCUR cyclomatic complexity curve comparison

This figure graphically demonstrates that cyclomatic complexity (McCabe, 1976) dictates that CONCUR's conditional path count magnitude is greater than that of AlexDSS. This greater conditional complexity translates directly to CONCUR's greater response flexibility over AlexDSS. The next question compares the response flexibility between CONCUR and a generic question-answer agent.

Question D.2: Does CONCUR allow users more response flexibility than a generic question-answer agent?

For comparison's sake, CONCUR's response flexibility can also be analyzed against a question-answer agent, such as ELIZA (Weizenbaum, 1966) or Sergeant Star (Artstein et al, 2009) through a cyclomatic complexity analysis. These systems can be generalized as question-answer agents because of their extremely shallow depths of follow-up questioning per context. For simplicity, a depth of 1 was used to model this style of agent.

The source control flow graph of a generic question-answer agent is depicted in Figure 31, with the assumption that it can handle $X \times Y$ responses, the same number of responses for the

CONCUR model in Figure 29. Again, the idea here is that these question-answer agents do not maintain a sense of *depth* in their discourse, expecting a different concept or context from each user turn.

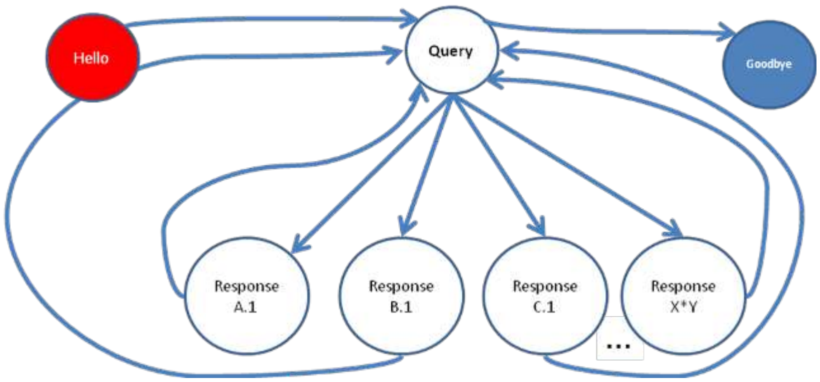


Figure 31: Control flow graph of a question-answer agent

Computing the cyclomatic complexity of the question-answer agent from (2) gives the following expression:

$$M_{Question-AnswerAgent} = XY + 1. \tag{8}$$

From the equations in (4) and (8), it is apparent that the polynomial function found for CONCUR's M is more cyclomatically complex than a question-answer chatbot, such as ELIZA (Weizenbaum, 1966) or Sergeant Star. (Artstein et al, 2009) Further verification was performed by setting the topic depth of Y to 3:

$$M_{Question-AnswerAgent} = 3X + 1. \tag{9}$$

Plotting the complexity curves from (7) and (9) gives the graph in Figure 32. This curve comparison shows the conditional complexity of CONCUR versus the question-answer agent for a corpora containing up to only five topics. The parabolic increase in decision point complexity for CONCUR dwarfs the other agent's linear progressions. This vast difference in conditional paths implies that CONCUR is more robust in terms of more open response expectations.

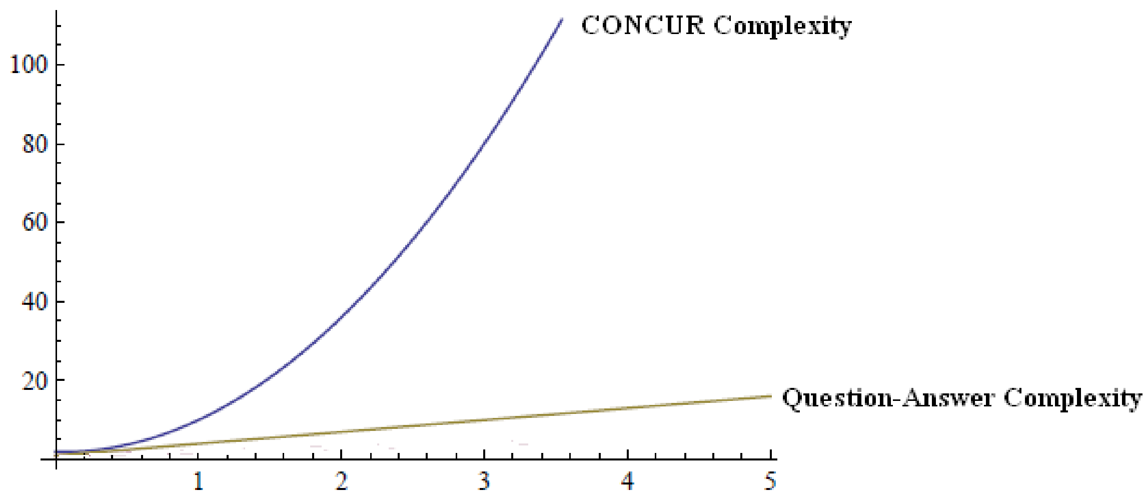


Figure 32: Cyclomatic complexity for CONCUR and Question-Answer agent

By demonstrating the cyclomatic complexity of CONCUR and a generic question-answer agent, it was established that the user is afforded more conversational flexibility with a CONCUR-based system. The next question examines the experimentally verifiable response flexibility between CONCUR and AlexDSS through an analysis of user verbosity.

<u>Question D.3: Are users more verbose with CONCUR than with AlexDSS?</u>

This question investigates how comfortable users felt in giving verbose requests to the CONCUR and AlexDSS agents. The verbosity of a conversation agent user can indicate how robustly it can handle more complex responses. This openness indicator is encapsulated in the Words Per Turn metric. The data involving the user and agents word counts per turn for CONCUR and AlexDSS are presented in Table 56. Each of the NSF I/UCRC-based agents was examined in this investigation. Specifically, the AlexDSS ECA and both the ECA-based and chatbot versions of CONCUR were involved.

Table 56: AlexDSS versus CONCUR Words Per Turn

	Data Set 1: AlexDSS ECA	Data Set 2: CONCUR ECA	Data Set 3: CONCUR Chatbot
User Words Per Turn	2.83	4.94	5.02
System Words Per Turn	28.56	29.09	28.22

All three systems utilized approximately the same number of system response word counts. This indicates that users were exposed the nearly the same amount of system prompts and information deployment actions. The most important conclusion from this table is the idea that CONCUR agents gathered more verbose responses over AlexDSS. CONCUR's users averaged nearly 5 words per turn to communicate with the agent, while AlexDSS user inputs had a mean of only 3 words per turn. This was nearly a 77% increase in user words, implying that test subjects tended to be more verbose with CONCUR than with the AlexDSS agent. Hence, the data from Table 56 indicates that users were more willing to engage in open dialog with the CONCUR agent than the AlexDSS ECA.

Analysis Summary

The point of this Open Dialog Analysis was to provide evidence of CONCUR's dialog openness, especially when compared to a menu-driven dialog manager in AlexDSS. A previous study showed that the AlexDSS ECA produced better Conversational Accuracy than the CONCUR ECA, a result of the influx of errors caused by specific QA requests in an open dialog setting. What this section strives to achieve is a justification of this decrease in Conversational Accuracy by demonstrating the dialog openness afforded by the CONCUR infrastructure. Effectively, a tradeoff between Conversational Accuracy and user input flexibility was made in designing the context-based dialog manager.

This section used the combination of experimental data and algorithmic complexity analysis to endorse CONCUR's ability to convey dialog openness. Cyclomatic complexity comparisons with the AlexDSS agent as well as with a generic question-answer chatbot demonstrated the response flexibility that CONCUR was designed to handle. An examination of the interaction trends among the two NSF I/UCRC dialog managers showed that while the system turn count remained the same, users employed more words when speaking to the CONCUR ECA than when conversing with the AlexDSS ECA. This implies that the CONCUR system yields a higher level of utterance flexibility over its AlexDSS counterpart.

Chapter Summary

This chapter discussed the different aspects and challenges of conversation agent evaluation. The main idea in this research realm is that both quality and quantitative methods must be considered, since both subjective and objective measures are involved in ECA design. In the recent past, many suggestions for generalizing chatbot assessment were made, but none were standardized. The evaluation metrics for the CONCUR prototype were derived from the overlapping measures featured in many of these attempts.

The second portion of this chapter presented a data set acquisition setup for CONCUR. Four different agents were described, each of which was tested to attain four individual data sets. Each agent represented a permutation of user input style and expertise knowledge. These were designed to express the three novel features of the prototypical dialog management system: 1) overcoming ASR limitations, 2) allowing for domain-independent knowledge management, and 3) providing open dialog. The collected metrics were a combination of quantitative measures, questionnaire results, and quantitative analyses of quality data. Upon analysis of the results, each of the three themes was experimentally and algorithmically verified.

The next chapter wraps up the work of this dissertation, beginning with a summary of what was done, followed by a presentation of final conclusions, and then concluding with a discussion of future research efforts.

CHAPTER SEVEN: SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

This chapter gives a brief summary of what was done in this research, followed by a conclusion to the work presented in this dissertation. The final section of this chapter offers recommendations for the problem investigated, and it discusses implications for future studies.

Summary

The work in this dissertation dealt with spoken HCI with emphasis on natural conversation flow. Specifically, it presented a context-centric method of dialog management to fortify the robustness of assistive speech-based ECAs. The particular areas of improvement were concentrated in three themes: overcoming ASR limitations, promoting an open dialog, and providing a domain-independent knowledge management system.

The first issue tackles a challenging technical aspect of speech-based interfaces, namely the current state of ASR. Speech recognition technology has proven to be quite unreliable, even in perfect environmental conditions. The obvious consequence of this is a deteriorated channel of spoken communication between machines and humans. The second problem deals with the constrained, expectation-based input style of many modern spoken dialog systems. What results from this method of discourse is an unnatural, menu-driven dialog that restricts users' conversational flexibility. The third theme addresses ECA knowledge bases, which have often been closely tied to the conversation agent design itself. Separating the expertise from the rest of the system allows for a quicker turnover for building agents. This idea lends itself to a plug-and-play, method of knowledge management that is domain-independent.

In response to these issues, a spoken dialog manager for an ECA was proposed, one that closely embraced the use of contextual information to navigate through a conversation. This contextualization process was central in addressing inaccurate ASR systems, creating an open dialog feel, and promoting expertise interchangeability via domain-independent knowledge management. A background study of modern technologies associated with creating this dialog manager was made, focusing on three major areas: NLP, dialog systems, and context-based methods. Each of these topics was presented in a general manner, followed by a collection of associated real-world applications. These three topics align to the permeating themes of ASR limitations, domain-independent knowledge management, and open dialog.

Following the background survey, an approach to building the context-centric dialog manager was presented. Three primary design decisions were identified for developing a dialog manager: input processing method, knowledge management, and discourse model. The foremost methods associated with each of these sub-systems choices were acknowledged and analyzed. A particular solution for each component was presented and justified to be included in the proposed dialog manager. Specifically, a keyphrase-based input processor, an encyclopedia corpus-based knowledge manager, and a CxBR discourse model were selected and tied together in a generalized framework. A prototypical treatment of this approach was embodied in the CONCUR dialog system. The central theme of CONCUR's major components is the use of contextual information, presenting itself as a dialog management application of the CxBR architecture.

Evaluation of CONCUR required both quantitative and qualitative methods, as dictated by the historically challenging realm of conversation agent evaluation. Past endeavors in assessing chatbot performance has relied on both subjective and objective metrics. Recent attempts were made to provide a generalized evaluation process, but with little success. The evaluation metrics for CONCUR were established from a conglomeration of these past efforts. Guided by these assessment tactics, the evaluation setup for CONCUR was presented, consisting of four different agents to be tested. These systems reflected a different permutation of user input style and expertise knowledge, and were designed to prove out the three themes of context-based dialog management: overcoming ASR limitations, promoting open dialog, and supporting domain-independent knowledge management. The collected data sets consisted of quantitative metrics, survey responses, and quantitative analyses of quality data. Analyzing these data led to the experimental validation of the three aforementioned themes.

Briefly, the contributions of this dissertation align with the development of HCI techniques involved with speech-based ECAs. In particular, the work here provides a robust method of open dialog that overcomes weak-performing speech recognition facilities. Additionally, the research presented a CxBR-based mixed-initiative spoken dialog management architecture with a domain-independent knowledge management system. Finally, a prototype metrics system was established and incorporated in evaluating the devised approach, resulting in publishable data touting the effectiveness of a context-centric dialog manager.

Conclusions

This dissertation brings forth insights in the realm of speech-based HCI. Specifically, four themes were touched upon by this work: conversation agent design, ASR, open dialog, and domain-independent knowledge management. Conclusions backed by this research regarding these areas are presented in the remainder of this section.

Conversation Agent Design

With the advent of ECAs in the last two decades, the progression beyond Weizenbaum's work has essentially been directed to putting a face to a voice, and eventually, a body to a face. The work in this dissertation verified that this direction of conversation agent evolution is indeed one that leads toward augmented reality, as experimental data reflected a more favorable disposition from users to advocate the naturalness of a physically present agent over a purely text-based entity.

Another aspect of Conversation Agent Design affected by this dissertation is the centralized use of contextual information. In the past, contextual methods only served as secondary support systems, assisting in auxiliary tasks such as speech recognition or NLP disambiguation. In CONCUR, however, contextualization serves as the primary driving force for the entire conversation agent dialog management process. Everything from input processing to knowledge management to discourse behavior is directed by contextual information. The idea behind this context-centric design is the agent's actions are focused on the overall conversation goals rather than the immediate syntactic mechanics of the user input.

Automatic Speech Recognition

The effect of ASR-related errors proved to be minimized through the use of discourse-based techniques designed to overcome these limitations. CONCUR's employment of such algorithms proved to be the case when reviewing its experimental data. Hence it was concluded that high WER is not necessarily an obstacle for speech-based ECA design because of the use of a heavily contextual reasoning-based method, as seen in CONCUR.

As ASR technology improves over time, it may appear that the work in this dissertation would suffer from obsolescence. The assumption is that as machines perfect the art of speech recognition, the use of any other method to interpret user speech input, such as that found in CONCUR, would be unnecessary. This, however, does not justify the call to exterminate contextual methods in dialog management. The opposite of this would come into play, as conversational cognition would take the limelight once syntax-level recognition is conquered. While this dissertation supports the idea of circumventing ASR errors through context-level processing instead of purely linguistic levels, it also presents a method of modeling conversational behavior at the conceptual level. With the advancements in ASR accuracy, this context-level discourse will also improve, as less context identification errors will be made with the increase in confidence of user utterance recognition. Hence, this research gives a glimpse into the next echelon of conversation-level computation, once syntax-level processing has been perfected.

Open Dialog

The issue of open dialog is still an elusive challenge to HCI researchers. While this dissertation makes an attempt to address it, a solution in its purest form could not be achieved. Instead, the work in this research has brought forth some auxiliary insight toward what may work while open dialog is resolved.

One conclusion regarding CONCUR's attempt at open dialog servicing is that agents should instill in the user that *specific* questions can be answered. Because of CONCUR's limitations associated with ASR errors, the ability to pinpoint exact answers to highly detailed questions caused it to deploy very general declarative responses. As users continued through conversations, their input utterances would eventually evolve from full sentences to one or two-word remarks, often taking the form of particular keyphrases recognizable by CONCUR's Input Processor. This is a direct result of CONCUR's discourse model insistence on providing general, informational responses rather than serve as a full-blown QA system. While the former is acceptable for early ECA work, more interactively advanced replies will be expected when users subject machines to longer, more detailed information requests.

A second conclusion states that open dialog discourse can achieve goal completion rates and conversational accuracy comparable to less complex dialog systems. CONCUR was able to show that users can be afforded the luxury of using a natural conversational style, instead of a set of command-style word phrases, without jeopardizing the machine's ability to negotiate a conversation with natural and useful responses.

Domain-Independent Knowledge Management

The corpus-based knowledge manager facilitated the ability of CONCUR to understand what it knows and does not know. When confronted with user responses comprised entirely of keyphrases not found in its knowledge corpora, the agent pleaded ignorance regarding the user's request and attempted to steer the conversation toward a topic it did know. This led to an observation about domain-independent knowledge management – it is equally important what the *agent* knows as what the *user* knows or expects to know.

In experimenting with different knowledge corpora, it was exhibited that more tightly knit knowledge bases were more effective than corpora that had a variety of different topics.

When the user is posed with topics that are perceived as spread out, such as a set of randomly picked current events, it is assumed that the agent knows much more than is in its corpus. Hence, it was observed that the user should have some pre-conceived notion of what the agent does know before engaging in a conversation. This was definitely the case for the NSF I/UCRC corpus-based CONCUR.

It was also concluded that the explicit separation of expertise knowledge from the discourse model can maintain consistent conversational accuracy and speed up agent development times. CONCUR's modularized style of knowledge management reflect these effects in its experimental data sets.

From these conclusions, the themes of ASR limitations, open dialog and domain-independent knowledge management were touched upon. The evaluation process featured in this dissertation responded to the idea that a context-based dialog management system could be harnessed to handle each of these themes. Its results did indeed verify that such an approach that centralized around a contextual information layer could provide an effective framework to overcome ASR limitations and to provide a conversational discourse model and a domain-independent knowledge manager for a speech-based, open dialog ECA.

Future Research

Future extensions of the work in this dissertation can be categorized into three avenues of exploration: HCI improvements, knowledge management enhancements and ethical considerations. Each of these general topics expands on the fundamental issues associated with speech-based ECA technology. The CONCUR dialog manager could be greatly enhanced if any of this future research is conducted. The following section presents the type of work that can improve the contribution of this dissertation.

HCI Improvements

The ECA work associated with CONCUR could be improved for more effective HCI. Such enrichments include accurate speech recognition, automated response synthesis, affective dialog computing, and gesture-based interactions. These measures augment the sense of realism experienced by users.

Accurate speech recognition simply refers to the continued efforts of researchers to provide an ASR system with negligible WERs. While the CONCUR dialog manager tolerates high speech recognition error rates, perfectly accurate ASR would promote the processing of the entirety of a user utterance, rather than just the context-centric keyphrases. With total word recognition, user intent can easily be identified. This improved user goal recognition allows for the use of a more complicated discourse model, which would augment ECA realism. Such a discourse model could disambiguate very specific user requests, as the words beyond the keyphrases can be analyzed.

Automated response synthesis describes the process of dynamically computing an agent output, as opposed to a simple repetition of the prescribed knowledge found in a response corpus. Such a system enables an ECA to provide variations in its syntactic behavior without disrupting its semantic intents. The resultant effect promotes a sense of realistic responses by providing unique utterances and eliminating repetition. NLP tools can assist in this response automation through the use of synonym replacement and syntax structure reconstruction.

Affective dialog computing incorporates the use of emotional cues to augment an ECA experience. The use of emotions can be employed by both the input and the output of an agent. On the input side, user voice inflections can be processed to detect for such phenomena as anger or elation. Additionally, video hardware can capture facial expressions to make similar analysis. For the agent's output, its responses can be enhanced to convey certain emotions, such as confusion or happiness. Adding this affective dimension to speech-based ECAs leads to an improved HCI experience since the user's contextual intent can be processed both syntactically and emotionally. CONCUR can use affective information to aid in its conversation discourse navigation.

Gesture-based interaction extends the agent input to include not only the user's voice, but also her/his body language. Essentially, the keyboard and mouse combination can be compared to the voice and body language duo. Here, the keyboard and voice serve as the language-based input channels, while the mouse and body language enables the machine to determine spatial context. CONCUR would benefit from gesticulation processing in that it can incorporate an added layer of contextual cues for discourse processing.

Knowledge Management Enhancements

CONCUR's knowledge manager relies heavily upon contextual indicators. Fetching data from its knowledge base simply requires a query based on a current context. While this method allows for an effective way to present the user with general contextual awareness, the simplicity of the system does not allow for specific QA requests. More research needs to be done for effective QA in speech-based HCI, especially when particular answers are expected. Such an effort would call for an extended knowledge representation technique. While CONCUR insists on using an encyclopedia-style corpus for its response repository, the corpus would have to be processed further to provide individual declarative statements.

Another enhancement to the knowledge management system would be an incorporation of Web-based corpus sources, such as Wikipedia. The information organization style found in these encyclopedic entries lends itself to the CONCUR architecture. Having such a vast resource of information from the WWW would be invaluable as conversation agent designs gravitate toward omniscient bodies of knowledge.

Ethical Considerations

One of the advantages of CONCUR's domain-independent knowledge management is its ability to provide a reusable speech-based ECA dialog manager framework. This reusability is important when it is desired that different human beings can be represented for an embodied agent. It is very much a realistic assumption that these are representations of real people. Hence, this dissertation suggests that ethical issues involved with replicated human personas should be addressed. While the main focus of this work was to provide an ECA dialog manager, it must be acknowledged that this tool could be inappropriately used to misrepresent a real human being. Responsible use of virtual personalities must be taken seriously.

APPENDIX A: SURVEY INSTRUMENT

	Disagree			Neutral			Agree
If I told someone the character in this tool was real they would believe me.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
I would be more productive if I had this system in my place of work.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
The character on the screen seemed smart.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
I felt like I was having a conversation with a real person.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
The tool provided me with the information I was looking for.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
I found this to be a useful way to get information.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
This tool made it harder to get information than talking to a person or using a website.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
This does not seem like a reliable way to retrieve information from a database.	1	2	3	4	5	6	7

	Disagree			Neutral			Agree
This did not feel like a real interaction with another person.	1	2	3	4	5	6	7

APPENDIX B: IRB APPROVAL LETTER

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Victor Hung and Co-PI: Avelino J. Gonzalez**

Date: **January 06, 2010**

Dear Researcher:

On 1/06/2010, the IRB approved the following modifications/human participant research until 1/5/2011 inclusive:

Type of Review:	Submission Response for UCF Initial Review Submission Form
Project Title:	Towards Lifelike Computer Interfaces that Learn: Examining Context-Driven Dialog Management For Speech-Based Embodied Conversation Agents
Investigator:	Victor Hung
IRB Number:	SBE-09-06613
Funding Agency:	National Science Foundation
Grant Title:	Collaborative Research: Towards Lifelike Computer Interfaces that Learn
Research ID:	1047443

The Continuing Review Progress Report must be submitted 2 – 4 weeks prior to the expiration date for studies that were previously expedited, and 8 weeks prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 1/5/2011, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Joseph Bielitzki, DVM, UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 01/06/2010 01:23:53 PM EST

Joanne Muratori

IRB Coordinator

APPENDIX C: COMPLETE NSF I/UCRC CORPUS

//Overview of Planning Grant

::About the planning grant

The purpose of the planning grant is to help a pending center or site to secure membership and fund the planning grant meeting needed to establish an IUCRC. The planning grant award is worth 10000 dollars. The support fund for the evaluator is additional. The award is to be used for travel for site directors to recruit prospective companies, fund the planning grant meeting, and to provide release time for the P I. The planning grants are not intended for the travel and lodging expenses for companies. Please see the current solicitation for more detail.

::Deadlines

A Letter of Intent is due January 1 or June 26 annually. Planning grant and full center proposals are due March 6 and September 25 annually. A letter of intent is only needed when submitting a planning grant proposal for the next immediate deadline. If the planning grant proposal deadline is missed, a new letter of intent will need to be submitted. It is always a good idea to check with the program director regarding submission of a LOI.

::Considerations for Writing a Planning Grant Proposal

Please refer to the current solicitation for more detailed information on preparing a planning grant proposal. Multiple universities involved in submitting planning grant proposals must submit a collaborative proposal. Please call FastLane helpdesk for more information regarding submission of a collaborative proposal.

::Rejected IUCRC Proposal

A proposal may be rejected for a number of reasons, missing one or more documents by any of the co P eyes failure to abide by the NSF's Grant Proposal Guidelines in area such as margins, spacing, and etc or for not filling in a project summary properly.

The planning grant proposal must state that it shall follow the IUCRC policies, procedures, and organizational structure. The planning grant proposal must contain specific research project proposals that include expected deliverables, deadlines, etc. Even though the projects are only considered envisioned, the panel must be able to confirm that the center's focus is innovative in their field and that the research is of quality.

::Joining an existing center

In order to join an existing center, the planning proposal must explicitly state that there is an existing center and that this proposal is to add a research site to it. The new site's research must be synergistic with the existing center and shall augment the research base. The new site must adopt all existing center policy and procedures which include membership rates, membership agreement, and memorandum of understanding. The proposal should include a letter of endorsement from the Center Director of the existing center in the supplementary documents section.

//Planning Grant Paper

::Planning Grant Proposal

The planning grant proposal is required to acquire the 10000 dollar planning grant award. See the current solicitation for guidelines to preparing a planning grant proposal.

::Title

The title for a planning grant must be headed as "Planning Grant IUCRC for AREA" where area is the research area for which the center is being proposed.

::Project Summary Section

The project summary is a one page description of the industry, research focus, and university capabilities. This section needs to explicitly state the Intellectual Merit and Broader Impact of the Proposal. See the current solicitation for guidelines to preparing a Project Summary.

::Objective Section

The Objective of the planning grant is to have a planning grant meeting between university and industry to agree on an initial research agenda.

::Project Description Section

The Project Description section is a writeup of all the proposed projects for the envisioned center to address companies' needs and interests.

::Supplementary Documents

The supplementary documents section contains the following required documents, Marketing Plan, Staff Plan, Membership Agreement, Draft Agenda, and Letters of Interest. See the current solicitation for guidelines to preparing the supplementary documents section.

::Marketing Plan

The marketing plan explains how the center shall make itself attractive to potential industrial members. See the current solicitation for guidelines to preparing the marketing plan.

::Staff Plan

The purpose of the staff plan is to identify the university's capability to allocate the human resources necessary for an IUCRC. See the current solicitation for guidelines to preparing a planning grant proposal.

::Membership Agreement

The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Any company or organization may join a Center provided the membership agreement form is signed and fees paid. Note that non-profit organizations, associations and non-NSF federal agencies require special attention. See typical membership agreement template on IUCRC homepage.

::Draft Agenda

A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center.

::Letter of Interest

A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. A letter of commitment which states that the company will join the center and a letter of support merely states that the company believes the idea of the center is good. Each university should include at least 6 letters of interest with their planning grant proposal.

::Budget

The planning grant proposal budget is primarily for travel, the planning meeting, and faculty time. Note any other sources of funds to be used in this study.

//Planning Grant Meeting

::About the Planning Grant Meeting

After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects.

::Project Presentations

A prospective center should present about 6 potential projects that the IUCRC would pursue. Poor project selection and presentation is an easy way for a center to lose credibility in the eye of the industry members. Time is limited during the meetings and therefore it is the center and site directors' responsibility to choose and weed out weak project. The presentations should use a 15-20 minute per project format. All projects that are presented should utilize the Executive Summary.

Each project must include scope of the work, deliverables for the first year, milestones and the proposed budget. At the end of each project, LIFE forms must be filled in by all potential members of the center.

::Executive Summary

The executive summary is a one page description of project that contains following materials, project name, P I or co P eyes, description of project, a brief experimental plan, potential member company benefits, expected deliverables, project budget, and how this project differs from other activities.

::Avoid Presentation Pitfalls

Project presentations should be on specific projects and not a group of projects. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Project presentations should provide a high level abstraction of the work. The executive summary covers the important aspects of a project. Be careful not to go into too much detail on the technical side. Give a basic overview of the technique that shall be applied to the problem that gives it novel merit. Multiple project presentations in a limited time frame can easily go over the limit. Ensure the presentation and the question and answer session are timely.

A LIFE form should be filled out immediately after its presentation. This ensures that it captures the initial reaction of the company representative before he forgets it. The LIFE forms should be filled out and returned to the center's evaluator before the start of the next presentation. Furthermore the forms can be filled in anonymously. Ensure that the submitters know this so that they won't restrain their response.

::Life Forms

The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. The varying interest levels can be compiled to show the general industrial interest in the project.

The center should prepare LIFE forms before the meeting. It can use online LIFE Forms or paper. A single LIFE form should include the title of the project, identification number, the P I and/or co P eyes involved.

::Industry Needs and Expectation Workshop

The purpose of this session is to provide an opportunity to all industry participants to discuss the research that they considered important and relevant to their companies and which were not covered in earlier presentations. This is truly a brainstorming session. Faculty members and students are more than welcome to stay in the room, however, no one is allowed to speak unless spoken to. This session is usually moderated by a designated person from the industry participants.

::NSF Closed Session

The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. The NSF representative will then ask for a show of hands to see how many attendees plan on recommending obtaining a membership into the center.

APPENDIX D: COMPLETE NSF I/UCRC CONCUR TRANSCRIPTS

Trial 1
(01/14/10 10:44:26) Avatar to User: It's good to see you. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What would you like to discuss?

(01/14/10 10:44:43) User to Avatar: anything you want

(01/14/10 10:44:44) Avatar to User: I'm not sure what you said. What else would you like to know about?

(01/14/10 10:44:51) User to Avatar: draft agenda

(01/14/10 10:44:52) Avatar to User: A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. How about we discuss something else now? How else can I help you out?

(01/14/10 10:45:09) User to Avatar: about letter of interest

(01/14/10 10:45:10) Avatar to User: A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. Do you have questions about that information?

(01/14/10 10:45:31) User to Avatar: membership agreement

(01/14/10 10:45:32) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(01/14/10 10:46:02) User to Avatar: no ethics of (NO I DON'T THINK SO)

(01/14/10 10:46:03) Avatar to User: I may not have got that right but here is some useful information. Any company or organization may join a Center provided the membership agreement form is signed and fees paid. Do you have questions about that information?

(01/14/10 10:46:21) User to Avatar: good bye

(01/14/10 10:46:21) Avatar to User: I see you need to go now. Keep the peace.

Trial 2
(01/14/10 10:47:33) Avatar to User: My name is Alex. What's your name?

(01/14/10 10:47:39) User to Avatar: Kevin Simmons

(01/14/10 10:47:41) Avatar to User: Hello, friend. I remember you. What would you like to discuss?

(01/14/10 10:47:49) User to Avatar: I'm not given shall the (I'M NOT KEVIN SULLIVAN)

(01/14/10 10:47:49) Avatar to User: I'm not sure what you said. What else would you like to know about?

(01/14/10 10:48:01) User to Avatar: deadlines

(01/14/10 10:48:01) Avatar to User: A Letter of Intent is due January 1 or June 26 annually. Do you have questions about that information?

(01/14/10 10:48:12) User to Avatar: yes

(01/14/10 10:48:12) Avatar to User: Let's continue with what we were just talking about. How else can I help you out?

(01/14/10 10:48:22) User to Avatar: what is due on January 1
(01/14/10 10:48:23) Avatar to User: I'm not sure I heard you right, but I have news for you. Planning grant and full center proposals are due March 6 and September 25 annually. A letter of intent is only needed when submitting a planning grant proposal for the next immediate deadline. Is there something further I can do for you?

(01/14/10 10:48:49) User to Avatar: no

(01/14/10 10:48:49) Avatar to User: What else would you like to know about?

(01/14/10 10:48:55) User to Avatar: membership agreement

(01/14/10 10:48:55) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(01/14/10 10:49:25) User to Avatar: tell me about the staff plan

(01/14/10 10:49:26) Avatar to User: The purpose of the staff plan is to identify the university's capability to allocate the human resources necessary for an IUCRC. See the current solicitation for guidelines to preparing a planning grant proposal. Let's change topics. What else would you like to know about?

(01/14/10 10:49:52) User to Avatar: that's all right (UH, THAT'S ALL, THANK YOU)

(01/14/10 10:49:53) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(01/14/10 10:49:59) User to Avatar: Interruption-will provide (GOOD BYE)

(01/14/10 10:49:59) Avatar to User: I see you've interrupted me. I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(01/14/10 10:50:16) User to Avatar: that nothing thank you think of a (NOTHING, THANK YOU AND GOOD BYE)

(01/14/10 10:50:16) Avatar to User: I guess we're done now. Keep the peace.

Trial 3

(01/14/10 11:31:45) Avatar to User: My name is Alex. What's your name?

(01/14/10 11:31:51) User to Avatar: Richard
(01/14/10 11:31:51) Avatar to User: Let my try that again. What's your name?

(01/14/10 11:31:57) User to Avatar: Richard Voyles

(01/14/10 11:31:57) Avatar to User: It's good to see you. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What do you want to talk about?

(01/14/10 11:32:21) User to Avatar: But let's talk about the draft agenda (UH, I WOULD LIKE TO TALK ABOUT DRAFT AGENDA)

(01/14/10 11:32:21) Avatar to User: A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. Let's change topics. What else would you like to know about?

(01/14/10 11:32:36) User to Avatar: Interruption-of the accepted it (UH, CAN YOU ...)

(01/14/10 11:32:37) Avatar to User: I see you've interrupted me. I'm having trouble figuring out what you want to know. How else can I help you out?

(01/14/10 11:32:48) User to Avatar: the member agreement

(01/14/10 11:32:49) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(01/14/10 11:33:19) User to Avatar: us what about the intellectual property (YES, WHAT ABOUT INTELLECTUAL PROPERTY)

(01/14/10 11:33:20) Avatar to User: I'm not sure I heard you right, but I have news for you. Any company or organization may join a Center provided the membership agreement form is signed and fees paid. Note that non-profit organizations, associations and non-NSF federal agencies require special attention. Do you have questions about that information?

(01/14/10 11:33:55) User to Avatar: of today so that IS lee was about the operators of what Polk and I are a good terms of (WHAT ABOUT, WELL UH, EXCUSE ME ALEX, WAIT. UM, EXCUSE ME CAN I ASK ABOUT, HOW CAN I INTERACT...WHAT ABOUT INTELLECTUAL PROPERTY)

(01/14/10 11:33:57) Avatar to User: See typical membership agreement template on IUCRC homepage. Let's change topics. What else would you like to know about?

(01/14/10 11:34:14) User to Avatar: the label the planning grant (TELL ME ABOUT THE PLANNING GRANT)

(01/14/10 11:34:14) Avatar to User: I can tell you about a number of things. You can see the available options on the screen. How else can I help you out?

(01/14/10 11:34:26) User to Avatar: a planning grant meeting (THE PLANNING GRANT MEETING)

(01/14/10 11:34:26) Avatar to User: After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects. Let's change topics. What else would you like to know about?

(01/14/10 11:34:58) User to Avatar: 014 (HOW ABOUT LIFE FORMS)
(01/14/10 11:34:58) Avatar to User: I don't know what you mean. How else can I help you out?

(01/14/10 11:35:05) User to Avatar: life forms

(01/14/10 11:35:05) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(01/14/10 11:35:43) User to Avatar: yes

(01/14/10 11:35:43) Avatar to User: What else would you like to know about?

(01/14/10 11:35:50) User to Avatar: you tell me about the NSF closed session (CAN YOU TELL ME ABOUT THE NSF CLOSED SESSION)

(01/14/10 11:35:51) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. Is there something further I can do for you?

(01/14/10 11:36:16) User to Avatar: delegates that's a real excited (NO I GUESS THAT'S IT ALEX, THANK YOU)

(01/14/10 11:36:17) Avatar to User: The NSF representative will then ask for a show of hands to see how many attendees plan on recommending obtaining a membership into the center. How about we discuss something else now? How else can I help you out?

(01/14/10 11:36:33) User to Avatar: Interruption-as those goals (WAIT A MINUTE)
(01/14/10 11:36:34) Avatar to User: I'm having trouble figuring out what you want to know. What else would you like to know about?

(01/14/10 11:36:46) User to Avatar: mouth and (NOTHING)

(01/14/10 11:36:47) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. How else can I help you out?

(01/14/10 11:36:57) User to Avatar: by Alex (GOOD BYE ALEX)

(01/14/10 11:36:58) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 4
(01/14/10 13:30:37) Avatar to User: My name is Alex. What's your name?

(01/14/10 13:30:42) User to Avatar: Hiding (HEIDI)

(01/14/10 13:30:45) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What can I do for you?

(01/14/10 13:31:10) User to Avatar: from playmate tests don't just testing the bulls system here (UM, I AM A TEST. I'M JUST TESTING THE WHOLE SYSTEM HERE. I HAVE A QUESTION ABOUT THE I/UCRC)

(01/14/10 13:31:12) Avatar to User: I don't know what you mean. What else would you like to know about?

(01/14/10 13:31:22) User to Avatar: how do I may as described by project (UM, HOW DO I DESCRIBE MY PROJECTS)

(01/14/10 13:31:22) Avatar to User: We can discuss several topics. Take a look at the screen to my right. How else can I help you out?

(01/14/10 13:31:32) User to Avatar: is not the case that may be out at the (WHAT DO I INCLUDE IN THE SUMMARY)

(01/14/10 13:31:33) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. What else would you like to know about?

(01/14/10 13:31:44) User to Avatar: clout with cyclone the marketing plan (WHAT SHOULD I INCLUDE IN THE MARKETING PLAN)

(01/14/10 13:31:45) Avatar to User: The marketing plan explains how the center shall make itself attractive to potential industrial members. Do you have questions about that information?

(01/14/10 13:31:58) User to Avatar: with the latest to make it attractive (WHAT DO I USE TO MAKE THEM MORE ATTRACTIVE, BROCHURES?)

(01/14/10 13:31:59) Avatar to User: See the current solicitation for guidelines to preparing the marketing plan. Let's change topics. What else would you like to know about?

(01/14/10 13:32:12) User to Avatar: budget

(01/14/10 13:32:12) Avatar to User: The planning grant proposal budget is primarily for travel, the planning meeting, and faculty time. Note any other sources of funds to be used in this study. How about we discuss something else now? How else can I help you out?

(01/14/10 13:32:32) User to Avatar: V letter of interest (THE LETTER OF INTEREST)

(01/14/10 13:32:32) Avatar to User: A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. Do you have questions about that information?

(01/14/10 13:32:51) User to Avatar: yes ado (YES I DO)

(01/14/10 13:32:51) Avatar to User: A letter of commitment which states that the company will join the center and a letter of support merely states that the company believes the idea of the center is good. Each university should include at least 6 letters of interest with their planning grant proposal. Let's change topics. What else would you like to know about?

(01/14/10 13:33:19) User to Avatar: project presentations

(01/14/10 13:33:19) Avatar to User: A prospective center should present about 6 potential projects that the IUCRC would pursue. Poor project selection and presentation is an easy way for a center to lose credibility in the eye of the industry members. Time is limited during the meetings and therefore it is the center and site directors' responsibility to choose and weed out weak project. Is there something further I can do for you?

(01/14/10 13:33:55) User to Avatar: can you explain life forms

(01/14/10 13:33:55) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Do you have questions about that information?

(01/14/10 13:34:42) User to Avatar: can you explain further

(01/14/10 13:34:42) Avatar to User: I may not have got that right but here is some useful information. The varying interest levels can be compiled to show the general industrial interest in the project. Is there something further I can do for you?

(01/14/10 13:35:03) User to Avatar: headway avoid presentation pitfalls (HOW DO I AVOID PRESENTATION PITFALLS)

(01/14/10 13:35:04) Avatar to User: Project presentations should be on specific projects and not a group of projects. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Do you have questions about that information?

(01/14/10 13:35:29) User to Avatar: what is the NSF closed session

(01/14/10 13:35:30) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. Is there something further I can do for you?

(01/14/10 13:35:48) User to Avatar: now I've got unfinished (NO, I'M DONE, I'M FINISHED)

(01/14/10 13:35:48) Avatar to User: In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. The NSF representative will then ask for a show of hands to see how many attendees plan on recommending obtaining a membership into the center. Let's change topics. What else would you like to know about?

(01/14/10 13:36:18) User to Avatar: from the nest (I'M FINISHED, GOOD BYE)

(01/14/10 13:36:19) Avatar to User: I'm not sure what you said. How else can I help you out?

(01/14/10 13:36:26) User to Avatar: a (UM...)

(01/14/10 13:36:26) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. What else would you like to know about?

(01/14/10 13:36:37) User to Avatar: I read that gets a (ARRIVEDERCI)

(01/14/10 13:36:37) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 5

(01/14/10 13:39:00) Avatar to User: My name is Alex. What's your name?

(01/14/10 13:39:04) User to Avatar: and Lindsay (I'M LINDSEY)

(01/14/10 13:39:06) Avatar to User: It's good to see you. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. How can I help you?

(01/14/10 13:39:26) User to Avatar: lot to ask about the devaluation of the center's (LET'S ASK ABOUT THE EVALUATION OF CENTERS)

(01/14/10 13:39:27) Avatar to User: I'm not sure what you said. What else would you like to know about?

(01/14/10 13:39:38) User to Avatar: can we talk about our rejected IUCRC proposal (CAN WE TALK ABOUT A REJECTED IUCRC PROPOSAL)

(01/14/10 13:39:38) Avatar to User: A proposal may be rejected for a number of reasons, missing one or more documents by any of the co P eyes failure to abide by the NSF's Grant Proposal Guidelines in area such as margins, spacing, and etc or for not filling in a project summary properly. The planning grant proposal must state that it shall follow the IUCRC policies, procedures, and organizational structure. Is there something further I can do for you?

(01/14/10 13:40:23) User to Avatar: yet if (YES)

(01/14/10 13:40:23) Avatar to User: The planning grant proposal must contain specific research project proposals that include expected deliverables, deadlines, etc. Even though the projects are only considered envisioned, the panel must be able to confirm that the center's focus is innovative in their field and that the research is of quality. Let's change topics. What else would you like to know about?

(01/14/10 13:40:46) User to Avatar: Interruption-triumphs and (CAN I INTERRUPT HIM)

(01/14/10 13:40:47) Avatar to User: I'm not sure what you said. How else can I help you out?

(01/14/10 13:40:52) User to Avatar: Alex going to (ALEX CAN WE TALK ABOUT...)

(01/14/10 13:40:53) Avatar to User: I don't know what you mean. Let's try that again; can you speak a little more slowly this time? What else would you like to know about?

(01/14/10 13:41:05) User to Avatar: IAB meeting

(01/14/10 13:41:06) Avatar to User: After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects. Let's change topics. How else can I help you out?

(01/14/10 13:41:38) User to Avatar: printout about life forms (CAN WE TALK ABOUT LIFE FORMS)

(01/14/10 13:41:38) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(01/14/10 13:41:47) User to Avatar: Interruption-is (...)

(01/14/10 13:41:47) Avatar to User: I see you've interrupted me. I may not have got that right but here is some useful information. The varying interest levels can be compiled to show the general industrial interest in the project. Do you have questions about that information?

(01/14/10 13:42:08) User to Avatar: yes

(01/14/10 13:42:09) Avatar to User: Let's continue with what we were just talking about. How else can I help you out?

(01/14/10 13:42:18) User to Avatar: to about project presentations (TALK ABOUT PROJECT PRESENTATIONS)

(01/14/10 13:42:19) Avatar to User: We're switching gears. A prospective center should present about 6 potential projects that the IUCRC would pursue. Do you have questions about that information?

(01/14/10 13:42:35) User to Avatar: yeah (YEAH HOW DO I AVOID)

(01/14/10 13:42:35) Avatar to User: I may not have got that right but here is some useful information. Poor project selection and presentation is an easy way for a center to lose credibility in the eye of the industry members. Time is limited during the meetings and therefore it is the center and site directors' responsibility to choose and weed out weak project. Is there something further I can do for you?

(01/14/10 13:43:05) User to Avatar: about the executive summary (HOW ABOUT THE EXECUTIVE SUMMARY)

(01/14/10 13:43:06) Avatar to User: The executive summary is a one page description of project that contains following materials, project name, P I or co P eyes, description of project, a brief experimental plan, potential member company benefits, expected deliverables, project budget, and how this project differs from other activities. Let's change topics. What else would you like to know about?

(01/14/10 13:43:47) User to Avatar: that saw good bye (THAT'S ALL GOOD BYE)

(01/14/10 13:43:47) Avatar to User: I see you need to go now. Keep the peace.

Trial 6

(01/14/10 13:51:54) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(01/14/10 13:52:00) User to Avatar: my name is George Vermont oh (MY NAME IS GEORGE VERMONT, ALEX)

(01/14/10 13:52:03) Avatar to User: Let my try that again. What's your name?

(01/14/10 13:52:10) User to Avatar: George form of what the state (GEORGE VERMONT, LIKE THE STATE)

(01/14/10 13:52:10) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. How can I help you?

(01/14/10 13:52:35) User to Avatar: told knew little about the basic IUCRC (TELL ME A LITTLE ABOUT THE BASIC IUCRC PROGRAM)

(01/14/10 13:52:35) Avatar to User: A proposal may be rejected for a number of reasons, missing one or more documents by any of the co P eyes failure to abide by the NSF's Grant Proposal Guidelines in area such as margins, spacing, and etc or for not filling in a project summary properly. The planning grant proposal must state that it shall follow the IUCRC policies, procedures, and organizational structure. Do you have questions about that information?

(01/14/10 13:53:20) User to Avatar: how much money can I get from a broker (HOW MUCH MONEY CAN I GET FROM THE PROGRAM)

(01/14/10 13:53:21) Avatar to User: The planning grant proposal must contain specific research project proposals that include expected deliverables, deadlines, etc. Even though the projects are only considered envisioned, the panel must be able to confirm that the center's focus is innovative in their field and that the research is of quality. Let's change topics. What else would you like to know about?

(01/14/10 13:54:01) User to Avatar: code that it'll go C around Alex (INTERESTING. I GOTTA GO NOW, SEE YOU AROUND ALEX)

(01/14/10 13:54:03) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(01/14/10 13:54:09) User to Avatar: Interruption-that's

(01/14/10 13:54:09) Avatar to User: I see you've interrupted me. I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(01/14/10 13:54:14) User to Avatar: Interruption-do then sold it questions or (YOU'VE ANSWERED ALL MY QUESTIONS)

(01/14/10 13:54:15) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 7

(01/14/10 15:24:02) Avatar to User: My name is Alex. What's your name?

(01/14/10 15:24:06) User to Avatar: goal of (GLENN)

(01/14/10 15:24:06) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What can I do for you?

(01/14/10 15:24:23) User to Avatar: the layout work and money from the IUCRC program (TELL ME HOW I CAN GET MONEY FROM THE IUCRC PROGRAM)

(01/14/10 15:24:24) Avatar to User: A proposal may be rejected for a number of reasons, missing one or more documents by any of the co P eyes failure to abide by the NSF's Grant Proposal Guidelines in area such as margins, spacing, and etc or for not filling in a project summary properly. The planning grant proposal must state that it shall follow the IUCRC policies, procedures, and organizational structure. Is there something further I can do for you?

(01/14/10 15:25:12) User to Avatar: tell me about the organization structure of an IUCRC

(01/14/10 15:25:12) Avatar to User: The planning grant proposal must contain specific research project proposals that include expected deliverables, deadlines, etc. Do you have questions about that information?

(01/14/10 15:25:33) User to Avatar: improve much attention the mayor of the computer you do real or (YOU'VE PAID AS MUCH ATTENTION TO ME AS IN REAL LIFE)

(01/14/10 15:25:34) Avatar to User: Even though the projects are only considered envisioned, the panel must be able to confirm that the center's focus is innovative in their field and that the research is of quality. Let's change topics. What else would you like to know about?

(01/14/10 15:25:57) User to Avatar: what goes up, old look for and a proposal (WHAT DOES A PANEL LOOK FOR IN A PROPOSAL)

(01/14/10 15:25:58) Avatar to User: I'm not sure what you said. How else can I help you out?

(01/14/10 15:26:07) User to Avatar: how are proposing review (HOW ARE PROPOSALS REVIEWED)

(01/14/10 15:26:08) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. What else would you like to know about?

(01/14/10 15:26:27) User to Avatar: what specifications to review word for four of a proposal (WHAT SPECIFICATIONS DO REVIEWERS LOOK FOR IN A PROPOSAL)

(01/14/10 15:26:28) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 8

(01/14/10 17:21:27) Avatar to User: My name is Alex. What's your name?

(01/14/10 17:21:32) User to Avatar: lawyers are (MY NAME IS ERIC)

(01/14/10 17:21:36) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What do you want to talk about?

(01/14/10 17:21:56) User to Avatar: more workshop mode where you for a vote of both were no more than a planning (OH I WANT TO TALK ABOUT WHAT YOU THINK THE BEST WAY OF RUNNING A PLANNING MEETING IS)

(01/14/10 17:21:57) Avatar to User: I don't know what you mean. How else can I help you out?

(01/14/10 17:22:09) User to Avatar: phone planning workshop for a run can either aren't sure she (PLANNING WORKSHOP FOR A UH CANDIDATE IUCRC)

(01/14/10 17:22:10) Avatar to User: There is a range of information I can provide including the following. You can see the available options on the screen. What else would you like to know about?

(01/14/10 17:22:24) User to Avatar: about the planning grant (ABOUT THE PLANNING GRANT MEETING)

(01/14/10 17:22:24) Avatar to User: The purpose of the planning grant is to help a pending center or site to secure membership and fund the planning grant meeting needed to establish an IUCRC. Do you have questions about that information?

(01/14/10 17:22:44) User to Avatar: for which the best of two before a planning grant (SURE, WHAT'S THE BEST AGENDA FOR A PLANNING GRANT MEETING)

(01/14/10 17:22:44) Avatar to User: The planning grant award is worth 10000 dollars. Is there something further I can do for you?

(01/14/10 17:22:59) User to Avatar: are no one should read to your menu the planning grant meeting of (UH HUH YEAH, I WANT YOU TO ADD TO YOUR MENU THE PLANNING GRANT MEETING AGENDA)

(01/14/10 17:22:59) Avatar to User: The support fund for the evaluator is additional. Do you have questions about that information?

(01/14/10 17:23:08) User to Avatar: colored (ALL RIGHT, THAT'S GOOD)

(01/14/10 17:23:08) Avatar to User: I may not have got that right but here is some useful information. The award is to be used for travel for site directors to recruit prospective companies, fund the planning grant meeting, and to provide release time for the P I. The planning grants are not intended for the travel and lodging expenses for companies. Is there something further I can do for you?

(01/14/10 17:23:39) User to Avatar: will find such group (NO THANKS, THAT'S GREAT)

(01/14/10 17:23:40) Avatar to User: Please see the current solicitation for more detail. Let's change topics. What else would you like to know about?

(01/14/10 17:24:14) User to Avatar: good bye

(01/14/10 17:24:14) Avatar to User: I see you need to go now. Keep the peace.

Trial 9

(01/14/10 18:32:40) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(01/14/10 18:32:46) User to Avatar: muddy of buffalo's (MARIA PALA....)

(01/14/10 18:32:48) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What would you like to discuss?

(01/14/10 18:33:12) User to Avatar: in your father of story of life forms (NEA...UH...LIFE FORMS)

(01/14/10 18:33:12) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(01/14/10 18:33:54) User to Avatar: not think it (NO THANK YOU)

(01/14/10 18:33:54) Avatar to User: I'm not sure I heard you right, but I have news for you. The varying interest levels can be compiled to show the general industrial interest in the project. The center should prepare LIFE forms before the meeting. Do you have questions about that information?

(01/14/10 18:34:17) User to Avatar: yet now thank you (YEAH, NO. THANK YOU)

(01/14/10 18:34:17) Avatar to User: I may not have got that right but here is some useful information. It can use online LIFE Forms or paper. Is there something further I can do for you?

(01/14/10 18:34:35) User to Avatar: comedian how they would get online (COULD YOU HELP ME WITH THE ONLINE)

(01/14/10 18:34:36) Avatar to User: A single LIFE form should include the title of the project, identification number, the P I and/or co P eyes involved. How about we discuss something else now? How else can I help you out?

(01/14/10 18:34:57) User to Avatar: avoid presentation pitfalls

(01/14/10 18:34:57) Avatar to User: Project presentations should be on specific projects and not a group of projects. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Is there something further I can do for you?

(01/14/10 18:35:19) User to Avatar: not (NO)

(01/14/10 18:35:19) Avatar to User: I'm not sure I heard you right, but I have news for you. Project presentations should provide a high level abstraction of the work. The executive summary covers the important aspects of a project. Do you have questions about that information?

(01/14/10 18:35:39) User to Avatar: EINO (NO)

(01/14/10 18:35:39) Avatar to User: I may not have got that right but here is some useful information. Be careful not to go into too much detail on the technical side. Give a basic overview of the technique that shall be applied to the problem that gives it novel merit. Is there something further I can do for you?

(01/14/10 18:36:01) User to Avatar: no

(01/14/10 18:36:01) Avatar to User: I guess we're done now. Keep the peace.

Trial 10
(01/14/10 18:36:56) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(01/14/10 18:37:02) User to Avatar: harm run the legal (HI I'M RODNEY KWICKO)

(01/14/10 18:37:05) Avatar to User: It's good to see you. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What do you want to talk about?

(01/14/10 18:37:28) User to Avatar: of a letter of interest

(01/14/10 18:37:28) Avatar to User: A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. Do you have questions about that information?

(01/14/10 18:37:47) User to Avatar: no

(01/14/10 18:37:47) Avatar to User: How else can I help you out?

(01/14/10 18:37:53) User to Avatar: know what about the marketing plan (UM, WHAT ABOUT THE MARKETING PLAN)

(01/14/10 18:37:53) Avatar to User: The marketing plan explains how the center shall make itself attractive to potential industrial members. See the current solicitation for guidelines to preparing the marketing plan. Do you have questions about that information?

(01/14/10 18:38:11) User to Avatar: no

(01/14/10 18:38:11) Avatar to User: I guess we're done now. Keep the peace.

Trial 11
(01/14/10 18:40:02) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(01/14/10 18:40:07) User to Avatar: Los (RIMAS)

(01/14/10 18:40:09) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What can I do for you?

(01/14/10 18:40:30) User to Avatar: De Chile mobile missile closed session (CAN YOU TELL ME ABOUT THE NSF CLOSED SESSION)

(01/14/10 18:40:31) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. Is there something further I can do for you?

(01/14/10 18:40:55) User to Avatar: while once the goal of the program (AH, WHAT'S THE GOAL OF THE PROGRAM)

(01/14/10 18:40:56) Avatar to User: The NSF representative will then ask for a show of hands to see how many attendees plan on recommending obtaining a membership into the center. Let's change topics. What else would you like to know about?

(01/14/10 18:41:12) User to Avatar: Interruption-called how does one of the ship (HOW DOES ONE ATTAIN A MEMBERSHIP)

(01/14/10 18:41:13) Avatar to User: I see you've interrupted me. I'm having trouble figuring out what you want to know. How else can I help you out?

(01/14/10 18:41:29) User to Avatar: of the government (HOW I CAN GET A MEMBERSHIP)

(01/14/10 18:41:30) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(01/14/10 18:41:40) User to Avatar: called like today and a number (I WOULD LIKE TO GAIN A MEMBERSHIP)

(01/14/10 18:41:41) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 12

(01/14/10 18:44:23) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(01/14/10 18:44:30) User to Avatar: of input how it looked cool weather (HI ALEX, BERT CALWELL HERE)

(01/14/10 18:44:35) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What can I do for you?

(01/14/10 18:44:59) User to Avatar: world that a built-in about the EINO project that the planning grant proposal would (WELL CAN YOU TELL ME A LITTLE ABOUT THE PLANNING GRANT PROPOSAL MEETING)

(01/14/10 18:44:59) Avatar to User: The planning grant proposal is required to acquire the 10000 dollar planning grant award. Do you have questions about that information?

(01/14/10 18:45:16) User to Avatar: know is that what hard to do work that you told about the action (YEAH, IS THAT LIKE HARD TO DO, OR CAN YOU TELL ME ABOUT THE MEMBERSHIP AGREEMENT)

(01/14/10 18:45:18) Avatar to User: See the current solicitation for guidelines to preparing a planning grant proposal. Let's change topics. What else would you like to know about?

(01/14/10 18:45:34) User to Avatar: while low of a fundamental workshop (WELL I'M INTERESTED IN THE WORKSHOP)

(01/14/10 18:45:35) Avatar to User: The purpose of this session is to provide an opportunity to all industry participants to discuss the research that they considered important and relevant to their companies and which were not covered in earlier presentations. This is truly a brainstorming session. Do you have questions about that information?

(01/14/10 18:46:02) User to Avatar: though, that the planning grant meeting (YEAH, TELL ME ABOUT THE PLANNING GRANT MEETING)

(01/14/10 18:46:02) Avatar to User: We're switching gears. After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. Is there something further I can do for you?

(01/14/10 18:46:20) User to Avatar: of (UM)

(01/14/10 18:46:20) Avatar to User: The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects. How about we discuss something else now? How else can I help you out?

(01/14/10 18:46:37) User to Avatar: Interruption-photo (OK)

(01/14/10 18:46:38) Avatar to User: I don't know what you mean. What else would you like to know about?

(01/14/10 18:46:46) User to Avatar: of (UMMMM)

(01/14/10 18:46:46) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. How else can I help you out?

(01/14/10 18:46:58) User to Avatar: I think with the EINO (I THINK I'VE HAD ENOUGH)

(01/14/10 18:46:58) Avatar to User: I'm not sure what you said. I guess we're done now. Keep the peace.

Trial 13
(01/15/10 11:04:10) Avatar to User: My name is Alex. What's your name?

(01/15/10 11:04:15) User to Avatar: of mines group (MY NAME IS CRAIG)

(01/15/10 11:04:19) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What do you want to talk about?

(01/15/10 11:04:39) User to Avatar: the planning meetings (UH, PLANNING MEETINGS)

(01/15/10 11:04:40) Avatar to User: There is a range of information I can provide including the following. You can see the available options on the screen. How else can I help you out?

(01/15/10 11:04:56) User to Avatar: will best fitness meeting held on for you (WELL AFTER THIS MEETING, HOW I CAN FIND YOU, ARE YOU ONLINE)

(01/15/10 11:04:57) Avatar to User: After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects. Let's change topics. What else would you like to know about?

(01/15/10 11:05:29) User to Avatar: of 14 of false (AVOIDING PITFALLS)

(01/15/10 11:05:30) Avatar to User: I'm not sure what you said. How else can I help you out?

(01/15/10 11:05:40) User to Avatar: where the pitfalls of planning meetings there should be aborted (WHAT ARE THE PITFALLS OF PLANNING MEETINGS THAT SHOULD BE AVOIDED)

(01/15/10 11:05:41) Avatar to User: I can tell you about a number of things. Take a look at the screen to my right. What else would you like to know about?

(01/15/10 11:05:55) User to Avatar: avoid presentation pitfalls

(01/15/10 11:05:55) Avatar to User: Project presentations should be on specific projects and not a group of projects. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Is there something further I can do for you?

(01/15/10 11:06:20) User to Avatar: that tell me about the planning grant meeting (YEAH TELL ME ABOUT THE PLANNING GRANT MEETING)

(01/15/10 11:06:20) Avatar to User: Let's change topics. What else would you like to know about?

(01/15/10 11:06:29) User to Avatar: the planning grant meeting

(01/15/10 11:06:29) Avatar to User: How about we discuss something else now? How else can I help you out?

(01/15/10 11:06:42) User to Avatar: all the focus of that talk about life forms (UH, LET'S TALK ABOUT LIFE FORMS)

(01/15/10 11:06:42) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(01/15/10 11:07:25) User to Avatar: who should review the life form data

(01/15/10 11:07:25) Avatar to User: The center should prepare LIFE forms before the meeting. A single LIFE form should include the title of the project, identification number, the P I and/or co P eyes involved. Do you have questions about that information?

(01/15/10 11:07:44) User to Avatar: Interruption-in the (DO YOU HAVE)

(01/15/10 11:07:45) Avatar to User: The varying interest levels can be compiled to show the general industrial interest in the project. It can use online LIFE Forms or paper. Let's change topics. What else would you like to know about?

(01/15/10 11:08:08) User to Avatar: about workshops receive workshops (WHAT ABOUT WORKSHOPS, I SEE, WORKSHOPS)

(01/15/10 11:08:09) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(01/15/10 11:08:24) User to Avatar: own (UM, LET'S CHANGE TOPICS)

(01/15/10 11:08:24) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(01/15/10 11:08:34) User to Avatar: Interruption-of the staff plan (STAFF PLAN)

(01/15/10 11:08:34) Avatar to User: I see you've interrupted me. The purpose of the staff plan is to identify the university's capability to allocate the human resources necessary for an IUCRC. Do you have questions about that information?

(01/15/10 11:08:51) User to Avatar: the S (YES)

(01/15/10 11:08:52) Avatar to User: See the current solicitation for guidelines to preparing a planning grant proposal. Let's change topics. What else would you like to know about?

(01/15/10 11:09:07) User to Avatar: marketing plan (THE MARKETING PLAN)

(01/15/10 11:09:07) Avatar to User: The marketing plan explains how the center shall make itself attractive to potential industrial members. See the current solicitation for guidelines to preparing the marketing plan. How about we discuss something else now? How else can I help you out?

(01/15/10 11:09:28) User to Avatar: about the membership agreement (HOW ABOUT THE MEMBERSHIP AGREEMENT)

(01/15/10 11:09:29) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. Do you have questions about that information?

(01/15/10 11:09:47) User to Avatar: publication delays

(01/15/10 11:09:48) Avatar to User: I may not have got that right but here is some useful information. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(01/15/10 11:10:13) User to Avatar: objective selection

(01/15/10 11:10:13) Avatar to User: The Objective of the planning grant is to have a planning grant meeting between university and industry to agree on an initial research agenda. Let's change topics. What else would you like to know about?

(01/15/10 11:10:33) User to Avatar: evaluator responsibility (EVALUATOR RESPONSIBILITIES)

(01/15/10 11:10:34) Avatar to User: I'm not sure what you said. How else can I help you out?

(01/15/10 11:10:47) User to Avatar: evaluator Irving your porch (EVALUATOR, ERR, ANNUAL REPORTS)

(01/15/10 11:10:48) Avatar to User: I don't know what you mean. Let's try that again; can you speak a little more slowly this time? What else would you like to know about?

(01/15/10 11:11:02) User to Avatar: industry needs and expectation (INDUSTRY NEEDS AND EXPECTATIONS)

(01/15/10 11:11:03) Avatar to User: The purpose of this session is to provide an opportunity to all industry participants to discuss the research that they considered important and relevant to their companies and which were not covered in earlier presentations. This is truly a brainstorming session. Do you have questions about that information?

(01/15/10 11:11:32) User to Avatar: this is a workshop up at the planning meeting (IS THIS A WORKSHOP AT THE PLANNING MEETING)

(01/15/10 11:11:33) Avatar to User: Faculty members and students are more than welcome to stay in the room, however, no one is allowed to speak unless spoken to. This session is usually moderated by a designated person from the industry participants. How about we discuss something else now? How else can I help you out?

(01/15/10 11:11:56) User to Avatar: the NSF closed session

(01/15/10 11:11:56) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. Do you have questions about that information?

(01/15/10 11:12:23) User to Avatar: mouth like you are thicker and Don M. Regan (WELL, THANK YOU, I THINK I'M DONE, WELL THANK YOU)

(01/15/10 11:12:24) Avatar to User: The NSF representative will then ask for a show of hands to see how many attendees plan on recommending obtaining a membership into the center. How about we discuss something else now? How else can I help you out?

(01/15/10 11:12:33) User to Avatar: Interruption-O (OH)

Trial 14

(01/14/10 14:22:16) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(01/14/10 14:22:23) User to Avatar: Ballots (ALEX)

(01/14/10 14:22:25) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What would you like to discuss?

(01/14/10 14:22:48) User to Avatar: out of buildup of what the law planning meeting (HOW DO YOU BUILD A, HOW DO I DO A PLANNING MEETING)

(01/14/10 14:22:49) Avatar to User: The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects. Is there something further I can do for you?

(01/14/10 14:23:16) User to Avatar: that what see what should be the agenda of such a big (WHAT'S THE, WHAT SHOULD BE THE AGENDA OF SUCH A MEETING)

(01/14/10 14:23:16) Avatar to User: We're switching gears. A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. Do you have questions about that information?

(01/14/10 14:23:38) User to Avatar: SI wanna know what you suggest the agenda should look like (YES, I WANT TO KNOW WHAT DO YOU SUGGEST THE AGENDA SHOULD LOOK LIKE)

(01/14/10 14:23:38) Avatar to User: A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. Is there something further I can do for you?

(01/14/10 14:23:55) User to Avatar: a set of good enough to know what the way (IS THAT A GOOD ANSWER IN A WAY)

(01/14/10 14:23:56) Avatar to User: After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. Let's change topics. What else would you like to know about?

(01/14/10 14:24:18) User to Avatar: looking though good bye (OK BYE)

(01/14/10 14:24:18) Avatar to User: I see you need to go now. Keep the peace.

Trial 15

(01/15/10 11:40:37) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(01/15/10 11:40:46) User to Avatar: all my name is Arlene O work in NSF (UH, MY NAME IS ARLENE, I WORK AT NSF)

(01/15/10 11:40:51) Avatar to User: It's good to see you. I remember you. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What would you like to discuss?

(01/15/10 11:41:15) User to Avatar: I'd like to discuss what does a comprises so all work your house or a presentation pitfalls (I'D LIKE TO DISCUSS WHAT COMPRISES, OH, WHAT UH, OH HOW TO AVOID PRESENTATION PITFALLS)

(01/15/10 11:41:16) Avatar to User: Project presentations should provide a high level abstraction of the work. The LIFE forms should be filled out and returned to the center's evaluator before the start of the next presentation. Is there something further I can do for you?

(01/15/10 11:41:37) User to Avatar: I I'm not clear (UH, I'M NOT CLEAR) <3/4> 1

(01/15/10 11:41:38) Avatar to User: I may not have got that right but here is some useful information. Project presentations should be on specific projects and not a group of projects. Do you have questions about that information?

(01/15/10 11:41:58) User to Avatar: the U. S. could you describe life forms (UH YES, COULD YOU DESCRIBE LIFE FORMS)

(01/15/10 11:41:58) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(01/15/10 11:42:39) User to Avatar: of that can you explain (UH, CAN YOU EXPLAIN THE AVATAR)

(01/15/10 11:42:40) Avatar to User: I'm not sure I heard you right, but I have news for you. The varying interest levels can be compiled to show the general industrial interest in the project. The center should prepare LIFE forms before the meeting. Do you have questions about that information?

(01/15/10 11:43:01) User to Avatar: of knowledge to (UH NO I DON'T)

(01/15/10 11:43:02) Avatar to User: It can use online LIFE Forms or paper. A single LIFE form should include the title of the project, identification number, the P I and/or co P eyes involved. Let's change topics. What else would you like to know about?

(01/15/10 11:43:24) User to Avatar: Interruption-offering (AVATARS)

(01/15/10 11:43:25) Avatar to User: I see you've interrupted me. I'm having trouble figuring out what you want to know. How else can I help you out?

(01/15/10 11:43:30) User to Avatar: Interruption-avatar avatars met her at (AVATAR)

(01/15/10 11:43:31) Avatar to User: I don't know what you mean. Let's try that again; can you speak a little more slowly this time? What else would you like to know about?

(01/15/10 11:43:40) User to Avatar: Interruption-avatar (ABOUT AVATARS)

(01/15/10 11:43:40) Avatar to User: I'm not sure what you said. I guess we're done now. Keep the peace.

Trial 16

(02/19/10 11:46:27) Avatar to User: My name is Alex. What's your name?

(02/19/10 11:46:31) User to Avatar: are no (MY NAME IS CARLOS LEON-BARTH)

(02/19/10 11:46:32) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. How can I help you?

(02/19/10 11:46:56) User to Avatar: of mice and men showed let's talk about the project description section (OH WELL NICE TO MEET YOU. I WOULD LIKE TO TALK ABOUT, UM, PROJECT DESCRIPTION SECTION)

(02/19/10 11:46:56) Avatar to User: The Project Description section is a writeup of all the proposed projects for the envisioned center to address companies' needs and interests. Let's change topics. What else would you like to know about?

(02/19/10 11:47:20) User to Avatar: womb (UMM)

(02/19/10 11:47:20) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 11:47:26) User to Avatar: to know the research and the visit would (GIVE ME A SECOND WAIT)

(02/19/10 11:47:27) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(02/19/10 11:47:38) User to Avatar: fitness center (GIVE ME A SECOND)

(02/19/10 11:47:38) Avatar to User: In order to join an existing center, the planning proposal must explicitly state that there is an existing center and that this proposal is to add a research site to it. The new site's research must be synergistic with the existing center and shall augment the research base. Is there something further I can do for you?

(02/19/10 11:48:06) User to Avatar: of assurance of about something else (UM SURE LET'S TALK ABOUT SOMETHING ELSE)

(02/19/10 11:48:07) Avatar to User: I may not have got that right but here is some useful information. The new site must adopt all existing center policy and procedures which include membership rates, membership agreement, and memorandum of understanding. Do you have questions about that information?

(02/19/10 11:48:27) User to Avatar: know (NO)

(02/19/10 11:48:28) Avatar to User: The proposal should include a letter of endorsement from the Center Director of the existing center in the supplementary documents section. How about we discuss something else now? How else can I help you out?

(02/19/10 11:48:43) User to Avatar: notice of about life forms (YEAH LET'S TALK ABOUT LIFE FORMS)

(02/19/10 11:48:44) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(02/19/10 11:49:23) User to Avatar: our own (UMM GIVE...)

(02/19/10 11:49:23) Avatar to User: I may not have got that right but here is some useful information. The varying interest levels can be compiled to show the general industrial interest in the project. Do you have questions about that information?

(02/19/10 11:49:44) User to Avatar: and though the start of a private presentation (NO LET'S TALK ABOUT PROJECT PRESENTATION)

(02/19/10 11:49:45) Avatar to User: Project presentations should be on specific projects and not a group of projects. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Is there something further I can do for you?

(02/19/10 11:50:09) User to Avatar: no (NO)

(02/19/10 11:50:09) Avatar to User: I guess we're done now. Keep the peace.

Trial 17

(02/19/10 16:26:19) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(02/19/10 16:26:25) User to Avatar: my name is for us (MY NAME IS FERAS)

(02/19/10 16:26:29) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. How can I help you?

(02/19/10 16:26:44) User to Avatar: phone (UM CAN I)

(02/19/10 16:26:44) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(02/19/10 16:26:53) User to Avatar: bilodeau of the budget (OK I WANT TO KNOW ABOUT THE BUDGET)

(02/19/10 16:26:53) Avatar to User: The planning grant proposal budget is primarily for travel, the planning meeting, and faculty time. Note any other sources of funds to be used in this study. Let's change topics. What else would you like to know about?

(02/19/10 16:27:14) User to Avatar: with you want to focus on (WHAT DO YOU MEAN BY FACULTY TIME)

(02/19/10 16:27:15) Avatar to User: I'm not sure what you said. How else can I help you out?

(02/19/10 16:27:23) User to Avatar: office of those offices for faculties of (UM YOU JUST SAID THE BUDGET WAS USED FOR FACULTY TIME)

(02/19/10 16:27:24) Avatar to User: I don't know what you mean. Let's try that again; can you speak a little more slowly this time? What else would you like to know about?

(02/19/10 16:27:38) User to Avatar: one of the remaining Y and faculty far (WHAT DO YOU MEAN BY FACULTY TIME)

(02/19/10 16:27:39) Avatar to User: I'm having trouble figuring out what you want to know. I guess we're done now. Keep the peace.

Trial 18

(02/19/10 12:01:23) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(02/19/10 12:01:28) User to Avatar: place of (LISA)

(02/19/10 12:01:30) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What do you want to talk about?

(02/19/10 12:01:47) User to Avatar: less talk about the workshop (LET'S TALK ABOUT THE WORKSHOP)

(02/19/10 12:01:48) Avatar to User: The purpose of this session is to provide an opportunity to all industry participants to discuss the research that they considered important and relevant to their companies and which were not covered in earlier presentations. This is truly a brainstorming session. Do you have questions about that information?

(02/19/10 12:02:18) User to Avatar: but he tell me about the executive summary (UH CAN YOU TELL ME ABOUT THE EXECUTIVE SUMMARY)

(02/19/10 12:02:18) Avatar to User: The executive summary is a one page description of project that contains following materials, project name, P I or co P eyes, description of project, a brief experimental plan, potential member company benefits, expected deliverables, project budget, and how this project differs from other activities. How about we discuss something else now? How else can I help you out?

(02/19/10 12:02:47) User to Avatar: Interruption-Barry above grade level (ALRIGHT THAT'S GREAT)

(02/19/10 12:02:48) Avatar to User: I don't know what you mean. What else would you like to know about?

(02/19/10 12:02:57) User to Avatar: will study of the NSF closed session (LET'S TALK ABOUT THE NSF CLOSED SESSION)

(02/19/10 12:02:57) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. Do you have questions about that information?

(02/19/10 12:03:23) User to Avatar: know what our life forms (NO BUT WHAT ARE LIFE FORMS)

(02/19/10 12:03:24) Avatar to User: The center should prepare LIFE forms before the meeting. It can use online LIFE Forms or paper. Is there something further I can do for you?

(02/19/10 12:03:40) User to Avatar: doing more information about life forms (DO YOU HAVE ANY MORE INFORMATION ABOUT LIFE FORMS)

(02/19/10 12:03:40) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. Do you have questions about that information?

(02/19/10 12:04:00) User to Avatar: no

(02/19/10 12:04:00) Avatar to User: I guess we're done now. Keep the peace.

Trial 19

(02/19/10 12:33:10) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(02/19/10 12:33:17) User to Avatar: delete the (TALITHA)

(02/19/10 12:33:18) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. How can I help you?

(02/19/10 12:33:41) User to Avatar: does some ailing Deng (UH SHOW ME WHAT YOU DO)

(02/19/10 12:33:42) Avatar to User: I don't know what you mean. What else would you like to know about?

(02/19/10 12:33:50) User to Avatar: let's talk about life forms

(02/19/10 12:33:50) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Is there something further I can do for you?

(02/19/10 12:34:31) User to Avatar: - (-)

(02/19/10 12:34:31) Avatar to User: I may not have got that right but here is some useful information. The varying interest levels can be compiled to show the general industrial interest in the project. Do you have questions about that information?

(02/19/10 12:34:51) User to Avatar: unknown is talking (I DON'T EVEN KNOW WHAT HE'S TALKING)

(02/19/10 12:34:52) Avatar to User: I'm not sure I heard you right, but I have news for you. The center should prepare LIFE forms before the meeting. Is there something further I can do for you?

(02/19/10 12:34:57) User to Avatar: Interruption-of year related topics (OH HERE RELATED TOPICS)

(02/19/10 12:34:59) Avatar to User: It can use online LIFE Forms or paper. A single LIFE form should include the title of the project, identification number, the P I and/or co P eyes involved. Let's change topics. What else would you like to know about?

(02/19/10 12:35:26) User to Avatar: home at the yield, he and (UM THE FACT THAT YOU DON'T HAVE A HAND)

(02/19/10 12:35:27) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(02/19/10 12:35:33) User to Avatar: Interruption-the

(02/19/10 12:35:33) Avatar to User: I see you've interrupted me. I don't know what you mean. Let's try that again; can you speak a little more slowly this time? What else would you like to know about?

(02/19/10 12:35:38) User to Avatar: Interruption-fiery (SORRY)

(02/19/10 12:35:38) Avatar to User: I'm not sure what you said. I guess we're done now. Keep the peace.

Trial 20

(02/19/10 13:20:00) Avatar to User: My name is Alex. What's your name?

(02/19/10 13:20:05) User to Avatar: room and inspire (HEY MY NAME IS PETER)

(02/19/10 13:20:08) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What do you want to talk about?

(02/19/10 13:20:25) User to Avatar: all that the planning grant (I'D LIKE TO TALK ABOUT THE PLANNING GRANT)

(02/19/10 13:20:25) Avatar to User: We can discuss several topics. Take a look at the screen to my right. What else would you like to know about?

(02/19/10 13:20:37) User to Avatar: of women (UH THE MEETING)

(02/19/10 13:20:37) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 13:20:46) User to Avatar: closed up one (UH THE STAFF PLAN)

(02/19/10 13:20:46) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. Is there something further I can do for you?

(02/19/10 13:21:14) User to Avatar: avoid presentation pitfalls

(02/19/10 13:21:14) Avatar to User: Project presentations should be on specific projects and not a group of projects. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Do you have questions about that information?

(02/19/10 13:21:36) User to Avatar: when is of the (UH YES I DO)

(02/19/10 13:21:37) Avatar to User: I may not have got that right but here is some useful information. Project presentations should provide a high level abstraction of the work. Is there something further I can do for you?

(02/19/10 13:21:56) User to Avatar: of the top vote of (I'D LIKE TO TALK ABOUT LIFE FORMS)

(02/19/10 13:21:56) Avatar to User: I'm not sure I heard you right, but I have news for you. The executive summary covers the important aspects of a project. Be careful not to go into too much detail on the technical side. Give a basic overview of the technique that shall be applied to the problem that gives it novel merit. Do you have questions about that information?

(02/19/10 13:22:23) User to Avatar: no end in a (NO I DO NOT)

(02/19/10 13:22:24) Avatar to User: I may not have got that right but here is some useful information. Multiple project presentations in a limited time frame can easily go over the limit. Is there something further I can do for you?

(02/19/10 13:22:41) User to Avatar: no

(02/19/10 13:22:41) Avatar to User: I guess we're done now. Keep the peace.

Trial 21

(02/19/10 13:58:30) Avatar to User: My name is Alex. What's your name?

(02/19/10 13:58:35) User to Avatar: my name is shown (MY NAME IS SEAN)

(02/19/10 13:58:40) Avatar to User: It's good to see you. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What would you like to discuss?

(02/19/10 13:58:56) User to Avatar: membership agreement

(02/19/10 13:58:56) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(02/19/10 13:59:25) User to Avatar: letter of interest

(02/19/10 13:59:26) Avatar to User: We're switching gears. A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. Do you have questions about that information?

(02/19/10 13:59:51) User to Avatar: zone for a (NO I'M GOOD BUDGET)

(02/19/10 13:59:51) Avatar to User: A letter of commitment which states that the company will join the center and a letter of support merely states that the company believes the idea of the center is good. Each university should include at least 6 letters of interest with their planning grant proposal. Let's change topics. What else would you like to know about?

(02/19/10 14:00:20) User to Avatar: of about the planning grant green (UM ABOUT THE PLANNING GRANT MEETING)

(02/19/10 14:00:20) Avatar to User: The purpose of the planning grant is to help a pending center or site to secure membership and fund the planning grant meeting needed to establish an IUCRC. The planning grant award is worth 10000 dollars. Is there something further I can do for you?

(02/19/10 14:00:41) User to Avatar: deadlines

(02/19/10 14:00:41) Avatar to User: A Letter of Intent is due January 1 or June 26 annually. Planning grant and full center proposals are due March 6 and September 25 annually. Do you have questions about that information?

(02/19/10 14:01:01) User to Avatar: rejected IUCRC proposal

(02/19/10 14:01:01) Avatar to User: We're switching gears. A proposal may be rejected for a number of reasons, missing one or more documents by any of the co P eyes failure to abide by the NSF's Grant Proposal Guidelines in area such as margins, spacing, and etc or for not filling in a project summary properly. The planning grant proposal must state that it shall follow the IUCRC policies, procedures, and organizational structure. Is there something further I can do for you?

(02/19/10 14:01:48) User to Avatar: no

(02/19/10 14:01:48) Avatar to User: I guess we're done now. Keep the peace.

Trial 22

(02/19/10 13:39:18) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(02/19/10 13:39:24) User to Avatar: all rule and (OH RUBEN)

(02/19/10 13:39:26) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. How can I help you?

(02/19/10 13:39:29) User to Avatar: Interruption-then (-)

(02/19/10 13:39:29) Avatar to User: I'm not sure what you said. What else would you like to know about?

(02/19/10 13:39:38) User to Avatar: joining an existing center

(02/19/10 13:39:38) Avatar to User: In order to join an existing center, the planning proposal must explicitly state that there is an existing center and that this proposal is to add a research site to it. Do you have questions about that information?

(02/19/10 13:39:57) User to Avatar: mall 05 (NO THAT'S FINE)

(02/19/10 13:39:57) Avatar to User: I may not have got that right but here is some useful information. The new site's research must be synergistic with the existing center and shall augment the research base. Is there something further I can do for you?

(02/19/10 13:40:18) User to Avatar: of all the planning grant (ABOUT THE PLANNING GRANT)

(02/19/10 13:40:18) Avatar to User: We can discuss several topics. Take a look at the screen to my right. What else would you like to know about?

(02/19/10 13:40:31) User to Avatar: no consideration for writing a planning grant proposal (EL CONSIDERATION FOR WRITING A PLANNING GRANT PROPOSAL)

(02/19/10 13:40:31) Avatar to User: We're switching gears. The planning grant proposal is required to acquire the 10000 dollar planning grant award. Do you have questions about that information?

(02/19/10 13:40:53) User to Avatar: but wasn't (NO QUESTIONS)

(02/19/10 13:40:54) Avatar to User: See the current solicitation for guidelines to preparing a planning grant proposal. How about we discuss something else now? How else can I help you out?

(02/19/10 13:41:10) User to Avatar: all with the draft agenda (UH WHAT'S THE DRAFT AGENDA)

(02/19/10 13:41:10) Avatar to User: A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. Let's change topics. What else would you like to know about?

(02/19/10 13:41:29) User to Avatar: of the sort of thing (UH THAT'S ALL I THINK)

(02/19/10 13:41:30) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(02/19/10 13:41:40) User to Avatar: the bill sold by banks (THAT'S ALL GOOD BYE THANKS)

(02/19/10 13:41:41) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(02/19/10 13:41:49) User to Avatar: Interruption-advice (GOOD BYE)

(02/19/10 13:41:49) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 23

(02/19/10 12:50:57) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(02/19/10 12:51:02) User to Avatar: one is job (MY NAME IS JOHN)

(02/19/10 12:51:06) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. How can I help you?

(02/19/10 12:51:07) User to Avatar: Interruption-are you (HOW ARE YOU)

(02/19/10 12:51:08) Avatar to User: I don't know what you mean. What else would you like to know about?

(02/19/10 12:51:15) User to Avatar: from EINO membership agreement (TELL ME ABOUT MEMBERSHIP AGREEMENTS)

(02/19/10 12:51:16) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(02/19/10 12:51:45) User to Avatar: are consuming more (UH CAN YOU TELL ME MORE)

(02/19/10 12:51:45) Avatar to User: Any company or organization may join a Center provided the membership agreement form is signed and fees paid. Do you have questions about that information?

(02/19/10 12:52:01) User to Avatar: of the pillar of interest (UH TELL ME ABOUT THE LETTER OF INTEREST)

(02/19/10 12:52:01) Avatar to User: A letter of interest is one in which company states it is interested in joining the center if it is created. This is not to be confused with a letter of commitment or a letter of support. Is there something further I can do for you?

(02/19/10 12:52:22) User to Avatar: of time of a letter of support (UH TELL ME ABOUT THE LETTER OF SUPPORT)

(02/19/10 12:52:22) Avatar to User: A letter of commitment which states that the company will join the center and a letter of support merely states that the company believes the idea of the center is good. Each university should include at least 6 letters of interest with their planning grant proposal. Let's change topics. What else would you like to know about?

(02/19/10 12:52:51) User to Avatar: of the staff plan (UH TELL ME ABOUT THE STAFF PLAN)

(02/19/10 12:52:51) Avatar to User: The purpose of the staff plan is to identify the university's capability to allocate the human resources necessary for an IUCRC. Do you have questions about that information?

(02/19/10 12:53:10) User to Avatar: about the title (TELL ME ABOUT THE TITLE)

(02/19/10 12:53:10) Avatar to User: The title for a planning grant must be headed as "Planning Grant IUCRC for AREA" where area is the research area for which the center is being proposed. Let's change topics. What else would you like to know about?

(02/19/10 12:53:33) User to Avatar: of the (UM NOTHING)

(02/19/10 12:53:34) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 12:53:43) User to Avatar: moment (NOTHING)

(02/19/10 12:53:44) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. What else would you like to know about?

(02/19/10 12:53:52) User to Avatar: Interruption-and nothing (NOTHING)

(02/19/10 12:53:52) Avatar to User: I see you've interrupted me. I guess we're done now. Keep the peace.

Trial 24
(02/19/10 14:03:12) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(02/19/10 14:03:17) User to Avatar: rhyme (BRIAN)

(02/19/10 14:03:19) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What do you want to talk about?

(02/19/10 14:03:37) User to Avatar: about the planning grant meeting

(02/19/10 14:03:37) Avatar to User: After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. The purpose of the planning grant meeting is to showcase the prospective center to potential members and address their interests in the projects. Let's change topics. What else would you like to know about?

(02/19/10 14:04:10) User to Avatar: live forms (LIFE FORMS)

(02/19/10 14:04:10) Avatar to User: The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. Do you have questions about that information?

(02/19/10 14:04:26) User to Avatar: no

(02/19/10 14:04:26) Avatar to User: How else can I help you out?

(02/19/10 14:04:33) User to Avatar: executive summary

(02/19/10 14:04:33) Avatar to User: The executive summary is a one page description of project that contains following materials, project name, P I or co P eyes, description of project, a brief experimental plan, potential member company benefits, expected deliverables, project budget, and how this project differs from other activities. Let's change topics. What else would you like to know about?

(02/19/10 14:05:17) User to Avatar: something random

(02/19/10 14:05:18) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 14:05:24) User to Avatar: draft agenda

(02/19/10 14:05:24) Avatar to User: A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. How about we discuss something else now? What else would you like to know about?

(02/19/10 14:05:43) User to Avatar: membership agreement (HOW ABOUT THE MEMBERSHIP AGREEMENT)

(02/19/10 14:05:43) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. Do you have questions about that information?

(02/19/10 14:05:58) User to Avatar: known (NO)

(02/19/10 14:05:58) Avatar to User: I may not have got that right but here is some useful information. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(02/19/10 14:06:22) User to Avatar: from the (I'M GOOD)

(02/19/10 14:06:22) Avatar to User: I'm not sure I heard you right, but I have news for you. Any company or organization may join a Center provided the membership agreement form is signed and fees paid. Do you have questions about that information?

(02/19/10 14:06:39) User to Avatar: no

(02/19/10 14:06:39) Avatar to User: I guess we're done now. Keep the peace.

Trial 25
(02/19/10 12:39:02) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(02/19/10 12:39:09) User to Avatar: My name is Steven Bosco (MY NAME IS STEVEN KOBOSCO)

(02/19/10 12:39:13) Avatar to User: Let me repeat myself; I'll say that again. What's your name?

(02/19/10 12:39:21) User to Avatar: My name is Steven Liang

(02/19/10 12:39:21) Avatar to User: Hello, friend. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What would you like to discuss?

(02/19/10 12:39:44) User to Avatar: are five to know about supplemental funding (UM I'D LIKE TO KNOW ABOUT SUPPLEMENTAL FUNDING)

(02/19/10 12:39:45) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(02/19/10 12:39:55) User to Avatar: from about the planning grant meeting (UM HOW ABOUT THE PLANNING GRANT MEETING)

(02/19/10 12:39:56) Avatar to User: After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. Do you have questions about that information?

(02/19/10 12:40:11) User to Avatar: of the start of the life forms (UM LET'S TALK ABOUT LIFE FORMS)

(02/19/10 12:40:11) Avatar to User: We're switching gears. The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. Is there something further I can do for you?

(02/19/10 12:40:30) User to Avatar: from once the NSF closed session (UM WHAT'S THE NSF CLOSED SESSION)

(02/19/10 12:40:31) Avatar to User: The NSF Closed Session excludes all university personnel except for a person who is familiar with most of the industry attendees. In this session, the NSF representative will discuss with the industry about how they felt about the meeting and the center in general. Do you have questions about that information?

(02/19/10 12:40:54) User to Avatar: yes

(02/19/10 12:40:54) Avatar to User: What else would you like to know about?

(02/19/10 12:41:00) User to Avatar: but no more information (I WOULD LIKE TO KNOW MORE INFORMATION)

(02/19/10 12:41:00) Avatar to User: The NSF representative will then ask for a show of hands to see how many attendees plan on recommending obtaining a membership into the center. Let's change topics. How else can I help you out?

(02/19/10 12:41:21) User to Avatar: model I avoid presentation pitfalls (UH HOW DO I AVOID PRESENTATION PITFALLS)

(02/19/10 12:41:21) Avatar to User: Project presentations should be on specific projects and not a group of projects. Do you have questions about that information?

(02/19/10 12:41:35) User to Avatar: ES (YES)

(02/19/10 12:41:35) Avatar to User: I'm not sure I heard you right, but I have news for you. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Project presentations should provide a high level abstraction of the work. Is there something further I can do for you?

(02/19/10 12:42:03) User to Avatar: R1 is the industry needs and expectation workshop (UH WHAT IS THE INDUSTRY NEEDS AND EXPECTATION WORKSHOP)

(02/19/10 12:42:04) Avatar to User: The purpose of this session is to provide an opportunity to all industry participants to discuss the research that they considered important and relevant to their companies and which were not covered in earlier presentations. This is truly a brainstorming session. Do you have questions about that information?

(02/19/10 12:42:30) User to Avatar: no

(02/19/10 12:42:30) Avatar to User: I guess we're done now. Keep the peace.

Trial 26
(02/19/10 13:01:15) Avatar to User: My name is Alex. What's your name?

(02/19/10 13:01:20) User to Avatar: the crash (UH CHRIS)

(02/19/10 13:01:22) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What would you like to discuss?

(02/19/10 13:01:43) User to Avatar: a novel about the finishing this proposal of working out what the web content are you looking for (HOW DO I GO ABOUT FINISHING THIS PROPOSAL I'M WORKING ON, WHAT WHAT CONTENT ARE YOU LOOKING FOR)

(02/19/10 13:01:45) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(02/19/10 13:01:56) User to Avatar: how do I submit the proposal

(02/19/10 13:01:56) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(02/19/10 13:02:09) User to Avatar: what is the marketing plan

(02/19/10 13:02:09) Avatar to User: The marketing plan explains how the center shall make itself attractive to potential industrial members. See the current solicitation for guidelines to preparing the marketing plan. Is there something further I can do for you?

(02/19/10 13:02:30) User to Avatar: the liquid in the project description section (UH WHAT DO I PUT IN THE PROJECT DESCRIPTION SECTION)

(02/19/10 13:02:30) Avatar to User: The Project Description section is a writeup of all the proposed projects for the envisioned center to address companies' needs and interests. Let's change topics. What else would you like to know about?

(02/19/10 13:02:55) User to Avatar: wouldn't want it would limit appointed to the membership agreement (WHAT DO WHAT DO I NEED TO PUT IN THE MEMBERSHIP AGREEMENT)

(02/19/10 13:02:56) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. The membership agreement is the heart of the IUCRC program. The agreement is not perfect, but it is one that is effective in fostering industry and university collaboration. Is there something further I can do for you?

(02/19/10 13:03:26) User to Avatar: when I needed within a project description section (WHAT DO I NEED TO PUT IN THE PROJECT DESCRITION SECTION)

(02/19/10 13:03:27) Avatar to User: We're switching gears. The Project Description section is a writeup of all the proposed projects for the envisioned center to address companies' needs and interests. Do you have questions about that information?

(02/19/10 13:03:50) User to Avatar: five no (UH NO)

(02/19/10 13:03:50) Avatar to User: I'm not sure I heard you right, but I have news for you. Any company or organization may join a Center provided the membership agreement form is signed and fees paid. Note that non-profit organizations, associations and non-NSF federal agencies require special attention. Is there something further I can do for you?

(02/19/10 13:04:21) User to Avatar: matter any part of the list of items I can leave out (IS THERE ANY PART OF THE LIST OF ITEMS THAT I CAN LEAVE OUT)

(02/19/10 13:04:22) Avatar to User: See typical membership agreement template on IUCRC homepage. How about we discuss something else now? How else can I help you out?

(02/19/10 13:04:38) User to Avatar: what happens if I miss the deadline (WHAT HAPPENS IF I MISS A DEADLINE)

(02/19/10 13:04:39) Avatar to User: I don't know what you mean. What else would you like to know about?

(02/19/10 13:04:53) User to Avatar: of the mouth of dog died away your money if the (UH I'M I'M DONE, I'M NOT REALLY INTERESTED)

(02/19/10 13:04:54) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. How else can I help you out?

(02/19/10 13:05:05) User to Avatar: mouth of (I'M DONE)

(02/19/10 13:05:06) Avatar to User: I'm not sure what you said. I guess we're done now. Keep the peace.

Trial 27

(02/19/10 13:53:32) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(02/19/10 13:53:38) User to Avatar: the minds and (UH MY NAME IS AMANDA)

(02/19/10 13:53:41) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What would you like to discuss?

(02/19/10 13:53:56) User to Avatar: the (UH)

(02/19/10 13:53:56) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 13:54:06) User to Avatar: flick of the grant proposal (LET'S LOOK AT THE GRANT PROPOSAL)

(02/19/10 13:54:07) Avatar to User: I can tell you about a number of things. Take a look at the screen to my right. What else would you like to know about?

(02/19/10 13:54:18) User to Avatar: planning

(02/19/10 13:54:18) Avatar to User: We can discuss several topics. You can see the available options on the screen. How else can I help you out?

(02/19/10 13:54:29) User to Avatar: of the planning grant (UH ABOUT THE PLANNING GRANT)

(02/19/10 13:54:30) Avatar to User: There is a range of information I can provide including the following. Take a look at the screen to my right. What else would you like to know about?

(02/19/10 13:54:43) User to Avatar: considerations

(02/19/10 13:54:43) Avatar to User: Please refer to the current solicitation for more detailed information on preparing a planning grant proposal. Multiple universities involved in submitting planning grant proposals must submit a collaborative proposal. Is there something further I can do for you?

(02/19/10 13:55:06) User to Avatar: the balance (UH DEADLINES)

(02/19/10 13:55:06) Avatar to User: Please call FastLane helpdesk for more information regarding submission of a collaborative proposal. How about we discuss something else now? How else can I help you out?

(02/19/10 13:55:22) User to Avatar: of the executives and (UH THE EXECUTIVE SUMMARY)

(02/19/10 13:55:23) Avatar to User: I'm having trouble figuring out what you want to know. What else would you like to know about?

(02/19/10 13:55:37) User to Avatar: joining an existing center

(02/19/10 13:55:37) Avatar to User: In order to join an existing center, the planning proposal must explicitly state that there is an existing center and that this proposal is to add a research site to it. Do you have questions about that information?

(02/19/10 13:56:03) User to Avatar: in the room of (NOT REALLY)

(02/19/10 13:56:04) Avatar to User: I'm not sure I heard you right, but I have news for you. The new site's research must be synergistic with the existing center and shall augment the research base. The new site must adopt all existing center policy and procedures which include membership rates, membership agreement, and memorandum of understanding. Is there something further I can do for you?

(02/19/10 13:56:34) User to Avatar: the second and (NOPE I THINK I'M DONE)

(02/19/10 13:56:34) Avatar to User: The proposal should include a letter of endorsement from the Center Director of the existing center in the supplementary documents section. Let's change topics. What else would you like to know about?

(02/19/10 13:56:54) User to Avatar: from the complete (UM NOTHING GOOD BYE)

(02/19/10 13:56:55) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 13:57:01) User to Avatar: the man (GOOD BYE)

(02/19/10 13:57:01) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. What else would you like to know about?

(02/19/10 13:57:07) User to Avatar: Interruption-Peter close (COMPUTER CLOSE)

(02/19/10 13:57:08) Avatar to User: I'm not sure what you said. I guess we're done now. Keep the peace.

Trial 28

(02/19/10 15:44:33) Avatar to User: My name is Alex. What's your name?

(02/19/10 15:44:40) User to Avatar: to hide as a mechanic said lehman's marks of (HI NICE TO MEET YOU ALEX MY NAME IS MAHSA)

(02/19/10 15:44:49) Avatar to User: Welcome. I don't think we've met. I am the virtual avatar of Alex Schwarzkopf of the National Science Foundation's IUCRC program. What do you want to talk about?

(02/19/10 15:45:11) User to Avatar: I've seen each year and fell 5/8 on information about the (NICE TO MEET YOU AND UH I NEED SOME INFORMATION ABOUT LETTER OF INTEREST)

(02/19/10 15:45:12) Avatar to User: I'm having trouble figuring out what you want to know. How else can I help you out?

(02/19/10 15:45:21) User to Avatar: of (UMM)

(02/19/10 15:45:21) Avatar to User: I don't know what you mean. Let's try that again; can you speak a little more slowly this time? What else would you like to know about?

(02/19/10 15:45:37) User to Avatar: a can you tell me about projects presentation (AH CAN YOU TELL ME ABOUT PROJECT PRESENTATION)

(02/19/10 15:45:37) Avatar to User: Project presentations should be on specific projects and not a group of projects. Do you have questions about that information?

(02/19/10 15:45:49) User to Avatar: yet sure (YEAH SURE)

(02/19/10 15:45:49) Avatar to User: I may not have got that right but here is some useful information. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. Is there something further I can do for you?

(02/19/10 15:46:10) User to Avatar: of (UM)

(02/19/10 15:46:10) Avatar to User: I'm not sure I heard you right, but I have news for you. Project presentations should provide a high level abstraction of the work. The executive summary covers the important aspects of a project. Do you have questions about that information?

(02/19/10 15:46:31) User to Avatar: yet they talk about the planning grant meeting (YEAH CAN YOU TALK ABOUT THE PLANNING GRANT MEETING)

(02/19/10 15:46:32) Avatar to User: We're switching gears. After the planning grant proposal is awarded, the prospective center shall hold a planning meeting. Is there something further I can do for you?

(02/19/10 15:46:50) User to Avatar: IAB what is the life forms (YEAH WHAT IS LIFE FORMS)

(02/19/10 15:46:51) Avatar to User: We're switching gears. The center should prepare LIFE forms before the meeting. Do you have questions about that information?

(02/19/10 15:47:03) User to Avatar: the ensure a Ku down the (YEAH SURE CAN YOU TELL ME WHAT IS THE INDUSTRY NEEDS)

(02/19/10 15:47:04) Avatar to User: I may not have got that right but here is some useful information. The Level of Interest Feedback Evaluation (LIFE) has been credited as a key element of the IUCRC program. Is there something further I can do for you?

(02/19/10 15:47:27) User to Avatar: the Pleas of explain more about it (YEAH PLEASE EXPLAIN MORE ABOUT IT)

(02/19/10 15:47:27) Avatar to User: They are used after each project presentation to capture the immediate reaction of an IAB member. This is important since the comments and concerns of the industry member are necessary to the development of experimental plan of the project. Do you have questions about that information?

(02/19/10 15:47:58) User to Avatar: when is executive summary (WHAT IS EXECUTIVE SUMMARY EXACTLY)

(02/19/10 15:47:59) Avatar to User: The executive summary is a one page description of project that contains following materials, project name, P I or co P eyes, description of project, a brief experimental plan, potential member company benefits, expected deliverables, project budget, and how this project differs from other activities. Is there something further I can do for you?

(02/19/10 15:48:39) User to Avatar: sounded as though (NO THAT WAS GOOD THANK YOU)

(02/19/10 15:48:39) Avatar to User: I'm not sure I heard you right, but I have news for you. The varying interest levels can be compiled to show the general industrial interest in the project. It can use online LIFE Forms or paper. Do you have questions about that information?

(02/19/10 15:49:03) User to Avatar: all no thanks you bet good bye (OH NO THANK YOU GOOD BYE)

(02/19/10 15:49:04) Avatar to User: I see you need to go now. Keep the peace.

Trial 29

(02/19/10 13:27:37) Avatar to User: My name is Alex Schwarzkopf. What's your name?

(02/19/10 13:27:43) User to Avatar: of the social (UH CHRIS SOSA)

(02/19/10 13:27:43) Avatar to User: It's good to see you. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What do you want to talk about?

(02/19/10 13:28:09) User to Avatar: of joining an existing center (UH JOINING AN EXISTING CENTER)

(02/19/10 13:28:09) Avatar to User: In order to join an existing center, the planning proposal must explicitly state that there is an existing center and that this proposal is to add a research site to it. The new site's research must be synergistic with the existing center and shall augment the research base. Is there something further I can do for you?

(02/19/10 13:28:39) User to Avatar: of consideration for a planning grant will (UH CONSIDERATIONS FOR WRITING A PLANNING GRANT PROPOSAL)

(02/19/10 13:28:39) Avatar to User: We can discuss several topics. Take a look at the screen to my right. What else would you like to know about?

(02/19/10 13:28:50) User to Avatar: move planning grant who (UH PLANNING A GRANT PROPOSAL)

(02/19/10 13:28:50) Avatar to User: There is a range of information I can provide including the following. You can see the available options on the screen. How else can I help you out?

(02/19/10 13:29:04) User to Avatar: planning grant proposal

(02/19/10 13:29:04) Avatar to User: We're switching gears. The planning grant proposal is required to acquire the 10000 dollar planning grant award. Do you have questions about that information?

(02/19/10 13:29:20) User to Avatar: of the marketing plan (UH THE MARKETING PLAN)

(02/19/10 13:29:21) Avatar to User: The marketing plan explains how the center shall make itself attractive to potential industrial members. See the current solicitation for guidelines to preparing the marketing plan. Let's change topics. What else would you like to know about?

(02/19/10 13:29:43) User to Avatar: more than one of its (WHAT WERE THE ORIGINAL TOPICS)

(02/19/10 13:29:44) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 13:29:51) User to Avatar: most of them (UH THE STAFF PLAN)

(02/19/10 13:29:51) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. What else would you like to know about?

(02/19/10 13:30:06) User to Avatar: staff plan

(02/19/10 13:30:06) Avatar to User: The purpose of the staff plan is to identify the university's capability to allocate the human resources necessary for an IUCRC. See the current solicitation for guidelines to preparing a planning grant proposal. Let's change topics. How else can I help you out?

(02/19/10 13:30:30) User to Avatar: to know what it (THE UH BUDGET)

(02/19/10 13:30:30) Avatar to User: I'm having trouble figuring out what you want to know. What else would you like to know about?

(02/19/10 13:30:37) User to Avatar: Interruption-can be both a (THE BUDGET)

(02/19/10 13:30:38) Avatar to User: I see you've interrupted me. I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. How else can I help you out?

(02/19/10 13:30:52) User to Avatar: membership agreement

(02/19/10 13:30:52) Avatar to User: The membership agreement is a contract between an IUCRC center, the university, and an industry member. Do you have questions about that information?

(02/19/10 13:31:09) User to Avatar: and draft agenda (HMM DRAFT AGENDA)

(02/19/10 13:31:09) Avatar to User: A draft agenda is the planning grant meeting that is used to determine the feasibility and viability of building a center. How about we discuss something else now? How else can I help you out?

(02/19/10 13:31:29) User to Avatar: to move this is a more aggressive and (I SEE YOU GUYS DON'T HAVE A WAY TO)

(02/19/10 13:31:30) Avatar to User: I'm having trouble figuring out what you want to know. What else would you like to know about?

(02/19/10 13:31:40) User to Avatar: of the mobile (UH NOTHING ELSE)

(02/19/10 13:31:40) Avatar to User: I'm not sure what you said. Can you please repeat what you just said; I didn't hear you. How else can I help you out?

(02/19/10 13:31:51) User to Avatar: of the bar (UH GOOD BYE)

(02/19/10 13:31:52) Avatar to User: I don't know what you mean. I guess we're done now. Keep the peace.

Trial 30
(02/19/10 15:59:40) Avatar to User: I'm Alex Schwarzkopf. What's your name?

(02/19/10 15:59:46) User to Avatar: they'll is busy all of (HEY ALEX IT'S JOHANN)

(02/19/10 15:59:48) Avatar to User: It's good to see you. I don't think we've met. I am the Project LifeLike Avatar, a 3-year collaborative effort sponsored by the NSF between the University of Central Florida and the University of Illinois at Chicago. What would you like to discuss?

(02/19/10 16:00:13) User to Avatar: A Aamodt always wondered both the idea right and executive summary (AH UM I'VE ALWAYS WONDERED HOW DO YOU WRITE AN EXECUTIVE SUMMARY)

(02/19/10 16:00:13) Avatar to User: The executive summary is a one page description of project that contains following materials, project name, P I or co P eyes, description of project, a brief experimental plan, potential member company benefits, expected deliverables, project budget, and how this project differs from other activities. Let's change topics. What else would you like to know about?

(02/19/10 16:00:54) User to Avatar: on (UM)

(02/19/10 16:00:54) Avatar to User: I don't know what you mean. How else can I help you out?

(02/19/10 16:01:01) User to Avatar: one BUZ second of summary (WHEN DO YOU USE THE EXEUCTIVE SUMMARY)

(02/19/10 16:01:02) Avatar to User: We can discuss several topics. Take a look at the screen to my right. What else would you like to know about?

(02/19/10 16:01:13) User to Avatar: the project presentations (UH PROJECT PRESENTATIONS)

(02/19/10 16:01:13) Avatar to User: A prospective center should present about 6 potential projects that the IUCRC would pursue. Do you have questions about that information?

(02/19/10 16:01:27) User to Avatar: home (UM)

(02/19/10 16:01:28) Avatar to User: I may not have got that right but here is some useful information. Poor project selection and presentation is an easy way for a center to lose credibility in the eye of the industry members. Time is limited during the meetings and therefore it is the center and site directors' responsibility to choose and weed out weak project. Is there something further I can do for you?

(02/19/10 16:01:52) User to Avatar: Interruption-are at Rei none of the more than one of the pit bulls (I ALREADY KNOW THAT WHAT ARE THE PITFALLS)

(02/19/10 16:01:52) Avatar to User: I see you've interrupted me. The presentations should use a 15-20 minute per project format. All projects that are presented should utilize the Executive Summary. Do you have questions about that information?

(02/19/10 16:02:12) User to Avatar: the mobile will cut a pit bulls will cut the bulls from board (WHAT KIND OF PITFALLS SHOULD I AVOID)

(02/19/10 16:02:14) Avatar to User: Each project must include scope of the work, deliverables for the first year, milestones and the proposed budget. At the end of each project, LIFE forms must be filled in by all potential members of the center. Let's change topics. What else would you like to know about?

(02/19/10 16:02:27) User to Avatar: Interruption-what's that falls (WHAT PITFALLS)

(02/19/10 16:02:28) Avatar to User: I see you've interrupted me. I don't know what you mean. How else can I help you out?

(02/19/10 16:02:35) User to Avatar: by what the goals (WHAT PITFALLS)

(02/19/10 16:02:35) Avatar to User: I'm having trouble figuring out what you want to know. I didn't catch that. What else would you like to know about?

(02/19/10 16:02:44) User to Avatar: Interruption-what project presentation pitfalls (WHAT PROJECT PRESENTATION PITFALLS)

(02/19/10 16:02:44) Avatar to User: I see you've interrupted me. Project presentations should be on specific projects and not a group of projects. Project presentations should provide a high level abstraction of the work. Multiple project presentations in a limited time frame can easily go over the limit. Do you have questions about that information?

(02/19/10 16:03:13) User to Avatar: U. S. offer no (UM I THINK THAT'S ALL FOR NOW)

(02/19/10 16:03:13) Avatar to User: I'm not sure I heard you right, but I have news for you. It is much easier for potential members to comment on and make suggestion for changes if presented with specific projects. The executive summary covers the important aspects of a project. Is there something further I can do for you?

(02/19/10 16:03:38) User to Avatar: no thank you (NO THANK YOU)

(02/19/10 16:03:38) Avatar to User: I may not have got that right but here is some useful information. Be careful not to go into too much detail on the technical side. Give a basic overview of the technique that shall be applied to the problem that gives it novel merit. Do you have questions about that information?

(02/19/10 16:03:51) User to Avatar: Interruption-no

(02/19/10 16:03:52) Avatar to User: I'm not sure I heard you right, but I have news for you. Ensure the presentation and the question and answer session are timely. A LIFE form should be filled out immediately after its presentation. This ensures that it captures the initial reaction of the company representative before he forgets it. Is there something further I can do for you?

(02/19/10 16:04:24) User to Avatar: no

(02/19/10 16:04:24) Avatar to User: I guess we're done now. Keep the peace.

APPENDIX E: COMPLETE CURRENT EVENTS CORPUS

//Sports

::NBA Playoffs

The Boston Celtics held off the Orlando Magic's fourth quarter surge to defeat the Magic 92-88 to take a 1-0 lead in their best of seven series in the NBA Eastern Conference Championship in front 17,461 Magic fans at the Amway Arena in Orlando. The Celtics defense held the Magic offense to a 41.6% shooting percentage. Ray Allen led the Celtics with 25 points while grabbing 7 rebounds. Paul Pierce had a fine all around game with 22 points, 9 rebounds, and 5 assists. Rasheed Wallace came off the bench to score 13 points and played strong defense against Magic center Dwight Howard. Celtics point guard Rajon Rondo contributed 8 points and 8 assists. *(Stephens, 2010)*

The Lakers haven't played since completing a sweep of Utah a week ago, but the circumstances are more considerable, the results more weighty when they begin the Western Conference finals Monday against the Phoenix Suns at Staples Center. They're four victories away from a 31st appearance in the NBA Finals, but they'll get there only if Kobe Bryant and Andrew Bynum shake off knee injuries that have basically kept them off the practice court the last week. *(Bresnahan, 2010)*

::Tiger Woods

Tiger Woods' ailing neck isn't bad enough to make him hesitate about scheduling future tournaments. Woods officially entered the July 15-18 British Open today. That of course was widely anticipated to happen at some point, given that this year's Brit is on the Old Course at St. Andrews, where Woods had dominating victories in 2000 and 2005. *(Cherner, 2010)*

::Tony Romo

Cowboys QB Tony Romo missed an opportunity to qualify for a PGA Tour event on Monday in order to be at practice with his Dallas teammates. Romo was scheduled for a 10:57 a.m. tee time in the qualifying for the Byron Nelson Championship, set for this Thursday-Sunday in Irving, Texas. But the time conflicted with a Cowboys OTA session, which is voluntary. Romo chose the Cowboys practice over the golf event. Romo had said last week he hoped to get an afternoon tee time so he could meet both commitments. But organizers of the golf competition were unable to accommodate him. Romo made clear last week he would not be tempted to skip the Cowboys practice. *(Leahy, 2010)*

::World Cup

Iraqi security forces have detained an al-Qaida militant suspected of planning an attack targeting the World Cup in South Africa next month, an official said Monday. Major General Qassim al-Moussawi, a spokesman for Baghdad security services, said Abdullah Azam Saleh al-Qahtani was an officer in the Saudi army. He is suspected of planning a "terrorist act" in South Africa during the World Cup beginning June 11, al-Moussawi told a news conference in Baghdad. He said al-Qahtani entered Iraq in 2004 and is suspected in several attacks in the capital and elsewhere in the country. In South Africa, a police spokesman said Iraq has not notified them of the arrest. *(Yaccoub, 2010)*

//Current News

::BP oil leak

After more than three weeks of trying to stop a gushing oil leak in the Gulf of Mexico, BP engineers have achieved some success using a mile-long pipe to capture some of the oil and divert it to a drill ship on the surface some 5,000 feet above the wellhead, company officials said Monday. After two false starts, engineers successfully inserted a narrow tube into the damaged pipe from which most of the oil is leaking. Doug Suttles, BP's chief operating officer, who appeared on several network morning shows Monday, said that the mile-long, 4-inch-wide tube was capturing a little more than 1,000 barrels of oil a day from the blown well and its 21-inch-wide riser pipe, and funneling the oil into the tanker ship. *(Dewan, 2010)*

::NYC bomber

The suspected driver in a failed car bombing of Times Square fits the profile of a recent wave of "homegrown" terrorists threatening America, New York police officials warned Tuesday. The officials said Faisal Shahzad and other suspects like Najibullah Zazi - the admitted leader of a plot to bomb the New York subway system - had roots in working- or middle-class society, some college education and no previous criminal records, but became radicalized in part by traveling to overseas terrorist hotbeds. The Times Square threat was "a classic case of homegrown terrorism," Police Commissioner Raymond Kelly said at a briefing for private security executives. *(Hays, 2010)*

::Icelandic volcano

Iceland's volcano has something out for European airline passengers. London's Heathrow and Amsterdam's Schiphol airports have been shut down again due to drifting ash from volcano. Heathrow was close at 1 a.m. Monday and is expected to reopen later Monday morning. Amsterdam's airport closed at 2 in the morning and is expected to remain closed until early afternoon on Monday. This is in addition the airports that were closed on Sunday due to the ash cloud. The immediate message to business travelers from these latest rounds of closures is that the volcano/ash impact isn't over yet. If you're planning on traveling in Europe today or possibly this week, be ready for some disruption. *(Kelly, 2010)*

::Bangkok protests

Troops and anti-government protesters are clashing in the Thai capital, Bangkok. More than 30 people have been killed and 200 injured in five days of violence that began on Thursday. The clashes are part of a two-month stand-off between "red-shirt" protesters and the government. *("Bangkok protests", 2010)*

//Health

::Health Care reform

The National Federation of Independent Business, never a fan of Democratic-inspired health care reform, announced Friday that it would join 20 states that are suing to have the law invalidated. The NFIB especially objects to the law's employer mandate, though in explaining its decision to back the legal challenge, the NFIB also takes aim at the fee on health insurance providers, new 1099 reporting requirements, and even the tax credits for small business, which it views as too meager "to make purchasing insurance more affordable for small firms." The lawsuit, however, does not focus on these issues - instead, it challenges the constitutionality of the individual mandate, which will require most Americans to purchase health insurance. *(Mandelbaum, 2010)*

::Cell phone health risks

Using a mobile phone does not appear to increase the risk of developing certain types of brain cancer, the largest study of its kind has concluded. Analysis of more than 10,000 people by the International Agency for Research on Cancer (IARC) found no relationship between years of use and risk. There is no known biological mechanism by which mobiles could cause cancer, but there has been public concern. It is hoped this study will allay some anxieties, as research continues. The overall rate of brain cancer has not risen in countries where use has long been prevalent - like Sweden, and studies have mostly found no evidence of an increased risk. This latest research is consistent with this. *(Murphy, 2010)*

::Fast food-related asthma

A burger and fries are not only bad for the waistline, they might also exacerbate asthma, a new study suggests. Patients with asthma who ate a high-fat meal had increased inflammation in their airways soon afterward, and did not respond as well to treatment as those who ate a low-fat meal, the researchers found. The results provide more evidence that environmental factors, such as diet, can influence the development of asthma, which has increased dramatically in recent years in westernized countries where high-fat diets are common. In 2007, about 34.1 million Americans had asthma, according to the American Academy of Allergy, Asthma and Immunology. From 1980 through 1994, the prevalence of asthma increased 75 percent. While the results are preliminary, they suggest cutting down on fat might be one way to help control asthma. *(Rettner, 2010)*

LIST OF REFERENCES

Agarwal, R. (1997). Towards a PURE spoken dialogue system for information access. Proceedings of the ACL/EACL Workshop on Interactive Spoken Dialog Systems: bringing speech and NLP together in real applications, pp. 90-97.

Aha, D. W., McSherry, D., and Yang, Q. (2005). Advances in conversational case-based reasoning. The Knowledge Engineering Review, 20(3), pp. 247-254.

Ahad, A., Jung, B., and Prendinger, H. (2007). Neva: A Conversational Agent based Interface for Library Information Systems. From Proceeding Human-Computer Interaction.

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L., and Stent, A. (2001). Towards Conversational Human-Computer Interaction. AI magazine.

Allen, J. F., Miller, B. W., Ringger, E. K., and Sikorski, T. (1996). Robust understanding in a dialogue system. Proceedings of the 34th Association for Computational Linguistics.

Alm, C., Roth, D., and Sproat, R. (2005). Emotions from Text: Machine Learning for Text-based Emotion Prediction. HLT/EMNLP 2005.

Al-Mubaid, H. (2006). Context-Based Technique for Biomedical Term Classification. IEEE Congress on Evolutionary Computation, pp. 5726-5733.

Andernach, J., ter Doest, H., op den Akker, H., van der Hoeven, G., Schaake, J., van der Burgt, S., and Nijholt, A. (1995). Language analysis for dialogue management in a theatre information and booking system. Artificial Intelligence '95: Fifteenth International Conference (Language Engineering 95), pp. 351-362.

Artstein, R., Gandhe, S., Gerten, J., Leuski, A., and Traum, D. (2009). Semi-formal Evaluation of Conversational Characters. Languages: From Formal to Natural: Essays Dedicated to Nissim Francez on the Occaision of his 65th Birthday, pp. 22-35.

Atwell, E., and Shawar, A. (2007). Chatbots: Are they really Useful? LDV-Forum Journal for Computational Linguistics and Language Techonology, 22(1), pp. 29-49.

Augello, A., Pilato, G., and Gaglio, S. (2009). A conversational agent to support decisions in SimCity like games. IEEE International Conference on Semantic Computing, pp. 367-372.

Augello, A., Santangelo, A., Sorce, S., Pilato, G., Gentile, A., Genco, A., and Gaglio, S. (2006). MAGA: A Mobile Archaeological Guide at Agrigento. Giornata nazionale su Guide Mobili Virtuali 2006, ACM-SIGCHI.

Babina, O. (2006). NLP-based Patent Information Retrieval. Web Journal of Formal, Computational & Cognitive Linguistics, 8.

Babu, S., Schmugge, S., Barnes, T., and Hodges, L. (2006). What Would You Like to Talk About? An Evaluation of Social Conversations with a Virtual Receptionist. 6th International Conference on Intelligent Virtual Agents.

Baca, J., Zheng, F., Gao, H., and Picone, J. (2003). Dialog Systems for Automotive Environments. 8th European Conference on Speech Communication and Technology, pp. 1929-1932.

Bain, W. (1986). Case-Based Reasoning: A computer model of subjective assessment. Ph. D. Thesis, Yale University.

Baker, K. L., Franz, A. M., and Jordan, P W. (1994). Coping With Ambiguity in Knowledge-based Natural Language Analysis. Proceedings of FLAIRS-94.

Ball, G. (1999). Mixing Scripted Interaction with Task-Oriented Language Processing in a Conversational Interface. Proceedings of the 1999 International Conference on Intelligent User Interfaces, pp. 101-103.

Bangkok protests day-by-day. (2010, May 17). BBC News. 17 May 2010. Retrieved July 12, 2010 from http://www.bbc.co.uk.

Barrows, R. C., Busuioc, M., and Friedman, C. (2000). Limited parsing of notational text visit notes: ad-hoc vs. NLP approaches. AMIA 2000 Symposium. pp. 51–55.

Batacharia, B., Levy, D., Catizone, R., Krotov, A., and Wilks, Y. (1997). CONVERSE: a Conversational Companion. 1st International Workshop on Human-Computer Conversation.

Béchet, F., Gorin, A., Wright, J., and Hakkani-Tür, D. (2004). Detecting and extracting named entities from spontaneous speech in a mixed-initiative spoken dialogue context: How May I Help You? Speech Communication, 42(2), pp. 207-225.

Belvin, R., Burns, R., and Hein, C. (2005). Development of the HRL Route Navigation Dialogue System. HLT-2001.

Bentivogli, L., and Pianta, E. (2004). Extending WordNet with Syntagmatic Information. 2nd Global WordNet Conference, pp. 47-53.

Beringer, N., Kartal, U., Louka, K., Schiel, F., and Turk, U. (2002). PROMISE - a procedure for multimodal interactive system evaluation. Proceedings of the LREC Workshop on Multimodal Resources and Multimodal Systems Evaluation.

Bernardi, R., Jijkoun, V., Mishne, G., and de Rijke, M. (2003). Selectively Using Linguistic Resources Throughout the Question Answering Pipeline. Proceedings 2nd CoLogNET-ElsNET Symposium.

Bernsen, N., and Dybkjær, L. (2005). User Evaluation of Converational Agent H. C. Andersen. 9th European Conference on Speech.

Bersen, N., Charfuelàn, M., Corradini, A., Dybkjær, L., Hansen, T., Kiilerich, S., Kolodnytsky, M., Kupkin, D., and Mehta, M. (2004). First Prototype of Conversational H.C. Andersen. Information Systems Animations and Virtual Realities.

Bickmore, T. and Cassell, J. (2000). 'How about this weather?' - Social Dialogue with Embodied Conversational Agents. Proceedings of the AAAI Fall Symposium on Socially Intelligent Agents.

Bickmore, T. W., and Picard, R. W. (2004). Towards Caring Machines. Computer Human Interaction.

Bickmore, T., Pfeifer, L., and Jack, B. (2009). Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. CHI 2009, pp. 1265-1274.

Bohus, D., and Rudnicky, A. (2002). LARRI: A language-based maintenance and repair assistant, IDS-2002, p. 13.

Bohus, D., and Rudnicky, A. (2003). RavenClaw: Dialog Management Using Hierarchical Task Decomposition and an Expectation Agenda. Proceedings of the Eighth European Conference on Speech Communication and Technology (Eurospeech 2003), pp. 597-600.

Branting, K., Lester, J., and Mott, B. (2004). Dialogue Management for Conversational Case-Based Reasoning. Proceedings of the Seventh European Conference on Case-Based Reasoning, pp. 77-90.

Breazeal, C. (2005). Socially intelligent robots. Interactions, 12(2), pp. 19-22.

Bresnahan, M. (2010, May 16) Lakers limping into Western Conference finals. Los Angeles Times. 16 May 2010. Retrieved July 12, 2010 from http://www.latimes.com.

Brill, E., Dumais, S., and Banko, M. (2002). An analysis of the AskMSR question-answering system. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing.

Brown, A., and Knight, K. (1990). Letters cues as retrieval aids in semantic memory.American Journal of Psychology, 103(1), pp. 101-113.

Brüninghaus, S., and Ashley, K. D. (1998) How Machine Learning Can Be Beneficial for Textual CBR. Papers from the AAAI-98 Workshop on Machine Learning for Text

Classification. Canisius, S., and van den Bosch, A. (2003). A Memory-Based Shallow Parser for Spoken Dutch. CLIN 2003.

Bui, T. H. (2006). Multimodal Dialogue Management - State of the art. Technical Report TR-CTIT-06-01 Centre for Telematics and Information Technology, University of Twente, Enschede.

Canisius, S., and van den Bosch, A. (2004). A memory-based shallow parser for spoken Dutch. Fourteenth Computational Linguistics in the Netherlands, pp. 31-45.

Cao, J., Roussinov, D., Robles, J., and Nunamaker, J. (2005). Automated Question Answering From Videos: NLP vs. Pattern Matching. Hawaii International Conference on System Sciences (HICSS-38).

Cardie, C. (1994). Domain-Specific Knowledge Acquisition for Conceptual Sentence Analysis, Ph.D. Thesis, University of Massachusetts, Amherst, MA.

Cassell, J., Ananny, M., Basu, A., Bickmore, T., Chong, P., Mellis, D., Ryokai, K., Smith, J., Vilhjálmsson, H., and Yan, H. (2000). Shared reality: Physical collaboration with a virtual peer. Proceedings of CHI 2000.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, I., Chang, K., Vilhjálmsson, H., and Yan, H. (1999). Requirements for An Architecture for Embodied Conversational Characters. Computer Animation and Simulation '99 (Eurographics Series), pp. 109-120.

Catrambone, R. (2002). Anthropomorphic Agents as a User Interface Paradigm: Experimental Findings and a Framework for Research. 24th Annual Conference of the Cognitive Science Society, pp. 166-171.

Cenek, P. (2001). Dialogue Interfaces for Library Systems. FI MU Report Series.

Chai, J., Horvath, V., Nicolov, N., Stys-budzikowska, M., Kambhatla, A., and Zadrozny, W. (2001). Natural Language Sales Assistant - a Web-based Dialog System for Online Sales. 13th Innovative Applications of Artificial Intelligence Conference.

Charfuelán, M., Gómez, L. H., López, C. E., and Hemsen, H. (2002). A XML-based tool for evaluation of SLDS. Proceedings of the Third International Conference on Language Resources and Evaluation.

Chen, Y., Zhou, M., and Wang, S. (2006). Reranking Answers for Definitional QA Using Language Modeling. ACL 2006.

Cheng, C. K., Pan, X., and Kurfess, F. J. (2003). Ontology-Based Semantic Classification of Unstructured Documents. Adaptive Multimedia Retrieval, pp. 120-131.

Cherner, R. (2010, May 17). Tiger Woods commits to play British Open. USA Today. 17 May 2010. Retrieved July 12, 2010 from http://www.usatoday.com.

Churcher, G., Atwell, E. S., and Souter, C. (1997). Dialogue management systems: a survey and overview. Report 97.6, School of Computer Studies, University of Leeds.

Cimiano, P., Haase, P., and Heizmann, J. (2007). Porting natural language interfaces between domains: an experimental user study with the ORAKEL system. Intelligent User Interfaces, pp. 180-189.

Clark, H. and Fox Tree, J. (2002). Using uh and um in spontaneous speech. Cognition, 84, pp. 73-111.

Clark, H., and Marshall, C. (1978). Reference diaries. Association for Computing Machinery Theoretical issues in natural language processing, 2, pp. 57-63.

Colby, K. (1973). Simulation of belief systems. Computer Models of Thought and Language.

de Almeida, R., and Libben, G. (2005). Changing morphological structures: The effect of sentence context on the interpretation of structurally ambiguous English trimorphemicwords. Language and Cognitive Processes, 20(1), 373-394.

DeMara, R., and Moldovan, D. (1993). The SNAP-1 Parallel AI Prototype. IEEE Transactions on Parallel and Distributed Systems, 4(8), pp. 841–854.

DeMara, R., Gonzalez, A., Jones, S., Johnson, A., Hung, V., Leon-Barth, C., Dookhoo, R., Leigh, J., Renambot, L., Lee, S., and Carlson, G. (2008). Toward Interactive Training with an Avatar-based Human-Computer Interface. Interservice/Industry Training, Simulation & Education Conference.

Denecke, M., and Yasuda, N. (2005). Does this answer your Question? Towards Dialogue Management for Restricted Domain Question Answering Systems. 6th SIGdial Workshop on Discourse and Dialogue.

Dewan, S. (2010, May 16). BP Reports Some Success in Capturing Leaking Oil. The New York Times. Retrieved July 12, 2010 from http://www.nytimes.com.

Dybkjær, L., and Bernsen, N. (2001). Usability evaluation in spoken language dialogue systems. Annual Meeting of the ACL Worshop on Evaluation for Language and Dialogue Systems, 9.

Edlund, J., Heldner, M., and Gustafson, J. (2006). Two faces of spoken dialogue systems. INTERSPEECH-2006 - ICSLP Satellite Workshop Dialogue on Dialogues: Multidisciplinary Evaluation of Advanced Speech-Based Interactive Systems.

Edmondson, W. H. (1999). Context and Contextualization. International Conference on Cognitive Sciences.

Eliassi-Rad, T., and Shavlik, J. (2003). Intelligent Web Agents that Learn to Retrieve and Extract Information. Intelligent Exploration of the Web.

Ferguson, G., Allen, J., Galescu, L., Quinn, J., and Swift, M. (2009). CARDIAC: An Intelligent Conversational Assistant for Chronic Heart Failure Patient Heath Monitoring. AAAI Fall Symposium on Virtual Healthcare Interaction.

Field, D., and Grasso, F. (2004). ELIZA, but Cleverer: Designing Persuasive Artificial Agents. Advanced Topics in Modeling Natural Language Dialog Workshop.

Field, D., and Ramsay, A. (2007). Deep-reasoning-centred dialogue. 11th European Workshop on Natural Language Generation, pp. 131–138.

Fleischman, M., and Roy, D. (2005). Intentional Context in Situated Language Learning. Proc. of 9th Conference on Comp. Natural Language Learning.

Flycht-Eriksson, A., and Jönsson, A. (2000). Dialogue and domain knowledge management in dialogue systems. Proceedings of the 1st SIGdial workshop on Discourse and dialogue, 10, pp. 121-130.

Forrest, S. (1991). Parallelism and Programming in Classifier Systems. Morgan Kaufmann Publishers, Inc.

Frankish, C., and Turner, E. (2007). Sihgt and sunod: The role of orthography and phonology inthe perception of transposed letter anagrams. Journal of Memory and Language, 56(2), pp. 189-211.

Friedman, C. (2000). A Broad-Coverage Natural Language Processing System. Proceeding of the AMIA Symposium, 19(19), pp. 270-274.

Fügen, C., Holzapfel, H, and Waibel, A. (2004). Tight coupling of speech recognition and dialog management - dialog-context dependent grammar weighting for speech recognition. INTERSPEECH-2004, 169-172.

Fung, P., and Yee, L. Y. (1998). An IR Approach for Translating New Words from Nonparallel, Comparable Texts. Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics, pp. 414-420.

Galescu, L., Allen, J., Ferguson, G., Quinn, J., and Swift, M. (2009). Speech Recognition in a Dialog System for Patient Health Monitoring. IEEE International Conference on Bioinformatics and Biomedicine Workshop on NLP Approaches for Unmet Information Needs in Health Care.

Galibert, O., Illouz, G., Rosset, S. (2005). Ritel: An Open-Domain, Human-Computer Dialog Systems. INTERSPEECH-2005, pp. 909-912.

Galvão, A., Barros, F., Neves, A., and Ramalho, G. (2004). Persona-AIML: An Architecture Developing Chatterbots with Personality. AAMAS, pp. 1266-1267.

Gandhe, S., Whitman, N., Traum, D., and Artstein, R. (2009). An Integrated Authoring Toolkit for Tactical Questioning Dialogue Systems. IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems.

Garrod, S., and Pickering, M. J. (2004). Why is conversation so easy? TRENDS in Cognitive Sciences, 8(1), pp. 8-11.

Gatterbauer, W., Bohunsky, P., Herzog, M., Krüpl, B., and Pollak, B. (2007). Towards domain-independent information extraction from web tables. Proceedings of the 16th international conference on World Wide Web, pp. 71-80.

Geib, C., and Steedman, M. (2007). On Natural Language Processing and Plan Recognition. IJCAI, pp. 1612-1617.

Gockley, R., Bruce, A.,Forlizzi, J., Michalowski, M. P., Mundell, A., Rosenthal, S., Sellner, B. P., Simmons, R., Snipes, K., Schultz, A., and Wang, J. (2005). Designing Robots for Long-Term Social Interaction. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 2199 - 2204.

Goh, O., Fung, C., Wong, K., and Depickere, A. (2006). An Embodied Conversational Agent for Intelligent Web Interaction on Pandemic Crisis Communication. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 397-400.

Gonzalez, A. J., and Dankel, D. D. (1993). The Engineering of Knowledge-Based Systems Theory and Practice. Prentice-Hall, Incorporated, Englewood Cliffs, New Jersey.

Gonzalez, A., and Ahlers, R. (1998). Context-based representation of intelligent behavior in training simulations. Transactions of the Society for Computer Simulation International, 15(4), pp. 153-166.

Gonzalez, A., Stensrud, B., and Barrett, G. (2008). Formalizing context-based reasoning: A modeling paradigm for representing tactical human behavior. International Journal of Intelligent Systems, 23(7), pp. 822-847.

González-Ferreras, C., Cardeñoso-Payo, V., and Sanchis-Arnal, E. (2008). Experiments In Speech Driven Question Answering. IEEE Workshop on Spoken Language Technology.

Gorin, A. L., Riccardi, G., Wright, J. H. (1997). How may I help you. Speech Communication, 23, pp. 113-127.

Goulian, J., Antione, J., and Poirier, F. (2003). How NLP Techniques can improve speech understanding using ROMUS – a Robust Chunk-based Message Understanding System Using Link. EUROSPEECH 2003.

Gray, D. and Walters, G. (1998). Managing the industry/university cooperative research center: a guide for directors and other stakeholders. Columbus, Ohio: Battelle Press.

Grice, H. P. (1975). Logic and conversation. In p. Cole and J. Morgan, editors, Syntax and Semantics: Vol. 3, Speech Act, pp. 43-58, Academic Press, New York.

Gruber, T. R. (1993). A translation approach to portable ontologies. Knowledge Acquisition, 5(2), pp. 199-220.

Gustafson, J., Heldner, M., and Edlund, J. (2008). Potential Benefits of Human-Like Dialogue Behaviour in the Call Routing Domain. PIT 2008, pp. 240-251.

Haav, H. (2003). An Application of Inductive Concept Analysis to Construction of Domain-specific Ontologies. 29th International Conference on Very Large DataBases.

Hajdinjak, M. and Mihelic, F. (2007). A Dialogue-management Evaluation Study. Journal of Computing and Information Technology, 15(2), pp. 111-121.

Hammer, J., Garcia-Molina, H., Cho, J., Aranha, R., and Crespo, A. (1997). Extracting Semistructured Information from the Web. Workshop on Management of Semistructured Data (PODS/SIGMOD'97).

Han, C., Han, N., Ko, E., and Palmer, M. (2002). Development and Evaluation of a Korean Treebank and its Application to NLP. Korean Society for Language and Information, 6(1), pp. 123-138.

Harabagiu, S., and Moldovan, D. (1996). Textnet - a text-based intelligent system. Working Notes of the AAAI Fall Symposium on Knowledge Representation Systems Based on Natural Language, pp. 32-43.

Harabagiu, S., Pasca, M., and Maiorano, S. (2000). Experiments with open-domain textual question answering. Proceedings of the COLING-2000.

Hardy, H., Strzalkowski, T., and Wu, M. (2003). Dialogue management for an automated multilingual call center. Proceedings of the HLT-NAACL 2003 workshop on Research directions in dialogue processing, 7.

Hardy, H., Strzalkowski, T., Wu, M., Ursu, C., Webb, N., Biermann, A., Inouye, R. B., and McKenzie, A. (2004). Data-driven strategies for an automated dialogue system. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, p.71.

Harris, T. K., and Rosenfeld, R. (2004). A Universal Speech Interface for Appliances. Proc. ICSLP 2004.

Hassel, L., and Hagen, E. (2005). Evaluation of a Dialogue System in an Automotive Environment. 6th SIGdial Workshop on Discourse and Dialogue.

Hays, T. (2010, May 11). Ray Kelly: Faisal Shahzad 'Homegrown' Terrorist. The Huffington Post. Retrieved July 12, 2010 from http://www.huffingtonpost.com.

Hermjakob, U., Hovy, E. H., and Lin, C. (2000). Knowledge-based question answering. TREC-2000.

Hickl, A., Wang, P., Lehmann, J. and Harabagiu, S. (2006). FERRET: Interactive Question-Answering for Real-World Environments. ACL 2006

Hori, C., Hori, T., Isozaki, H., Maeda, E., Katagiri, S., and Furui, S. (2003). Study on spoken interactive open domain question answering. SSPR2003, pp. 111-113 .

Hoshino, A., Kato, K., Takeuchi, J., and Tsujino, H. (2005). A Chat Information Service System using A Humanoid Robot. IEEE International Workshop on Robots and Human Interactive Communication, pp. 468-473.

Huang, C., Xu, P., Zhang, X., Zhao, S., Huange, T., and Xu, B. (1999). LODESTAR: A Mandarin Spoken Dialogue System for Travel Information Retrieval. Proceedings of Eurospeech'1999, 3, pp. 1155-1158.

Huang, J., Zhou, M., and Yang, D. (2007). Extracting Chatbot Knowledge from Online Discussion Forums. IJCAI 2007, pp. 423-428.

Huck, G., Fankhauser, P., Aberer, K., Neuhold, E. (1998). Jedi: Extracting and Synthesizing Information from the Web. Conference on Cooperative Information Systems.

Huhns, M., and Singh, M. (1997). Conversational Agents. IEEE Internet Computing, 1(2), pp. 73-75.

Inui, N., Koiso, T., Nakamura, J., Kotani, Y. (2003). Fully corpus-based natural language dialogue system. Natural Language Generation in Spoken and Written Dialogue, AAAI Spring Symposium.

Isaacs, E., and Clark, H. (1987). References in conversations between experts and novices. Journal of Experimental Psychology: General, 116, pp. 26-37.

Isbister, K., and Doyle, P. (2002). Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy. AAMAS Workshop: Embodied Conversatinal Agents.

Jia, J. (2003). The Study of the Application of a Keywords-based Chatbot System on the Teaching of Foreign Languages. CoRR: Computers and Society.

Jia, J. (2004). CSIEC (Computer Simulator in Educational Communication): A Virtual Context-Adaptive Chatting Partner for Foreign Language Learners. ICALT 2004.

Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. Knowledge-Based Systems, 22(4), pp. 249-255.

Jiang, J., and Zhai, C. (2007). Instance weighting for domain adaptation in NLP. Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, pp. 264-271.

Johnsen, M. H., Svendsen, T., Amble, T., Holter, T., and Harborg, E. (2000). TABOR – A Norwegian Spoken Dialogue System for Bus Travel Information. International Conference on Spoken Language Processing.

Jokinen, K. (2003). Natural Interaction in Spoken Dialogue Systems. Proceedings of the Workshop on Ontologies and Multilinguality in User Interfaces. HCI International 2003, 4, pp. 730-734.

Jokinen, K., and McTear, M. (2009). Spoken Dialogue Systems. Morgan & Claypool Publishers.

Jokinen, K., Hurtig, T., Hynna, K., Kanto, K., Kaipanen, M., and Kerminen, A. (2001). Self-Organizing Dialogue Management. Proceedings of the Natural Language Pacific Rim Symposium (NLPRS) Workshop: Neural Networks and Natural Language Processing.

Kameyama, M., Kawai, G., Arima, I. (1996). A Real-Time System for Summarizing Human-Human Spontaneous Spoken Dialogues. 4th International Conference on Spoken Language Processing, pp. 681-684.

Kang, S., Lee, S., and Seo, J. (2009). Dialogue Strategies to Overcome Speech Recognition Errors in Form-Filling Dialogue. 22nd International Conference on Computer Processing of Oriental Languages, pp. 282-289.

Karanastasi, A., Zotos, A., Christodoulakis, S. (2007). The OntoNL Framework for Natural Language Interface Generation and a Domain-Specific Application. Proceedings of the DELOS Conference on Digital Libraries.

Karis, D., and Dobroth, K. M. (1991). Automating services with speech recognition over the public switched telephone network: human factors considerations. IEEE Journal on Selected Areas in Communications, 9(4), pp. 574-585.

Kelly, D. (2010, May 17). Here We Go Again. About.com Guide to Business Travel. Retrieved July 12, 2010 from http://www.about.com.

Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsela, S., and Piepol, D. (2007). Building Interactive Virtual Humans for Training Environments. I/ITSEC’07.

Kenny, P., Parsons, T., Rizzo, A. (2009). Human Computer Interaction in Virtual Standardized Patient Systems. HCI (4), pp. 514-523.

King, S., Stephenson, T., Isard, S., Taylor, P., and Strachan, A. (1998). Speech recognition via phonetically featured syllables. Proceedings of ICSLP'98.

Kitano, H., and Higuchi, T. (1991). Massively Parallel Memory-Based Parsing. Proceedings of IJCAI-91.

Kladke, R. R. (1989). A Mega-Heuristic Approach to the Problem of Component Identification in Automated Knowledge Generation. M. S. Thesis, University of Central Florida.

Kopp, S., Gesellensetter, L., Krämer, N.C., and Wachsmuth, I. (2005). A conversational agent as museum guide - design and evaluation of a real-world application. Intelligent Virtual Agents, 3661, pp. 329-343.

Krenn B., Neumayr, B., Gstrein, E., and Grice, M. (2004). Lifelike Agents for the Internet: A Cross-Cultural Case Study. Agent Culture: Human-Agent Interaction in a Multicultural World, pp. 197-229.

Kübler, S. (2006). Towards Case-Based Parsing: Are Chunks Reliable Indicators for Syntax Trees? Proceedings of the COLING/ACL Workshop on Linguistic Distances.

Kupiec, J. (1993). MURAX: a robust linguistic approach for question answering using an on-line encyclopedia. 16th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 181-190.

Kurohashi, S., and Higasa, W. (2000). Dialogue Helpsystem based on Flexible Matching of User Query with Natural Language Knowledge Base. 1st ACL SIGdial Workshop on Discourse and Dialogue, pp. 141-149.

Kwok, C., Etzioni, O. and Weld, D. S. (2001). Scaling question answering to the Web. World Wide Web, pp. 150-161.

L'Abbate, M., and Thiel, U. (2002). Helping conversational agents to find informative responses: query expansion methods for chatterbots. Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 796-797.

L'Abbate, M., Thiel, U., and Kamps, T. (2005). Can Proactive Behavior turn Chatterbots into Conversational Agents? IAT 2005, pp. 173-179.

Lamel, L., Rosset, S., Gauvain, J., Bennacef, S., Garnier-Rizet, M., and Prouts, B. (1998). The LMSI ARISE System. IEEE 4th Workshop on Interactive Voice Technology Telecommunications applications, pp. 209-214.

Larsson, S. (2002). Issue-based Dialogue Management. Ph. D. Thesis, Goteborg University.

Larsson, S. and Traum, D. (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. Natural Language Engineering, 6(3-4), pp. 323-340.

Leahy, S. (2010, May 17). Tony Romo skips tee time to qualify for PGA Tour event because it conflicted with Cowboys practice. USA Today. Retrieved July 12, 2010 from http://www.usatoday.com.

Lee, C., Sidner, C., and Kidd, C. (2005). Engagement During Dialogues with Robots. AAAI Spring Symposia.

Lee, G., Jung, H., Lee, J. (1995). Bi-directional memory-based dialog translation: The KEMDT approach. CoRR cmp-lg/9502030.

Lee, L. (1997). Simliarity-Based Approaches to Natural Language Processing. PhD thesis, Harvard University.

Lemon, O. (2004). Context-sensitive speech recognition in ISU dialogue systems: results for the grammar switching approach. CATALOG, 8th Workshop on the Semantics and Pragmatics of Dialogue.

Lemon, O., and Liu, X. (2006). DUDE: A Dialogue And Understanding Development Environment Mapping Business Process Models To Information State Update Dialogue Systems. EACL - Demos.

Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM, 38(11).

Levin, E., Narayanan, S., Pieraccini, R., Biatov, K., Bocchieri, E., DiFabbrizio, G., Eckert, W., Lee, S., Pokrovsky, A., Rahim, M., Ruscitti, P., and Walker, M. (2000). The AT&T-DARPA Communicator Mixed-Initiative Spoken Dialog System. ICSLP, pp. 122-125.

Levin, L., Glickman, O., Qu, Y., Gates, D., Lavie, A., Rose, C. P., Van Ess-Dykema, C., and Waibel, A. (1995). Using Context in Machine Translation of Spoken Language. Proceedings of Theoretical and Methodological Issues in Machine Translation (TMI-95).

Levin, L., Langley, C, Lavie, A., Gates, D., Wallace, D., and Peterson, K. (2003). Domain Specific Speech Acts for Spoken Language Translation. In Proceedings of 4th SIGdial Workshop on Discourse and Dialogue.

Lieberman, H., Faaborg, A., Daher, W., and Espinosa, J. (2005). How to Wreck a Nice Beach You Sing Calm Incense. International Conference on Intelligent User Interfaces.

Lita, L., and Carbonell, J. (2002). Unsupervised Question Answering Data Acquisition From Local Corpora. 13th Conference on Information and Knowledge Management.

Litkowski, K. (1999). Question-Answering Using Semantic Relation Triples. 8th Text Retrieval Conference.

Litman, D., Pan, S., and Walker, M. (1998). Evaluating Response Strategies in a Web-Based Spoken Dialogue Agent. COLING-ACL, pp. 780-786.

Liu, H., and Singh, P. (2004). ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal, 22.

Loos, B. (2006). On2L - A Framework for Incremental Ontology Learning in Spoken Dialog Systems. Proceedings of the COLING/ACL 2006 Student Research Workshop, pp. 61–66.

Lu, C., Chiou, G., Day, M., Ong, C., and Hsu, W. (2006). Using Instant Messaging to Provide an Intelligent Learning Environment. Intelligent Tutoring Systems 2006, pp. 575-583.

LuperFoy, S., Loehr, D., Duff, D., Miller, K., Reeder, F., and Harper, L. (1998). An architecture for dialogue management, context tracking, and pragmatic adaptation in spoken dialogue systems. 17th International Conference On Computational Linguistics, 2, pp. 794-801.

MacWhinney, B., Keenan, J. M., and Reinke, P. (1982). The role of arousal in memory for conversation. Mem Cognit, 10(4), pp. 308-317.

Magnini, B., Serafini, L., and Speranza, M. (2002). Using NLP Techniques for Meaning Negotiation. IRST Technical Report 0208-18.

Mandelbaum, R. (2010, May 17). N.F.I.B Joins Suit Against Health Reform. The New York Times. Retrieved July 12, 2010 from http://www.nytimes.com.

Massaro, D., Cohen, M., Beskow, J., and Cole, R. (2001). Developing and evaluating conversational agents. Embodied conversational agents, pp. 287-318.

Masterman, M. (1957). The Thesaurus in Syntax and Semantics. Mechanical Translation, 4(1 – 2), pp. 35-43.

Mayer, N. (2009). The Lifenaut Project: a multifaceted experiment in data storage for future machine consciousness learning. SENS4 Conference.

McBreen, H., and Jack, M. (2000). Empirical Evaluation of Animated Agents In a Multi-Modal Retail Application. AAAI Fall Symposium: Socially Intelligent Agents - The Human in the Loop, pp. 122-126.

McCabe, T. (1976). A Complexity Measure. IEEE Transactions on Software Engineering, pp. 308-320.

McIlraith, S. A., Plexousakis, D., and van Harmelen, F. (2004). The Semantic Web. Third International Semantic Web Conference.

McTear, M. F. (2002) Spoken dialogue technology: enabling the conversational interface. ACM Computing Surveys, 34(1), pp. 90-169.

Means, R. W., Nemat-Nasser, S. C., Fan, A. T., and Hecht-Nielsen, R. (2004). A Powerful and General Approach to Context Exploitation in Natural Language Processing. Proceedings of the Human Language Technology Conference, North American chapter of the Association for Computational Linguistics Annual Meeting.

Medelyan, O., and Witten, I. H. (2006). Thesaurus Based Automatic Keyphrase Indexing. Proceedings of the Joint Conference on Digital Libraries, pp. 296-297.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). WordNet: an on-line lexical database. International Journal of Lexicography, 3(4).

Milne, D., Medelyan, O., and Witten, I. H. (2006). Mining Domain-Specific Thesauri from Wikipedia: A case study. Proceedings of the International Conference on Web Intelligence.

Minker, W. (2002). Overview on Recent Activities in Speech Understanding and Dialogue Systems Evaluation. 7th International Conference on Spoken Language Processing, pp. 337-340.

Minker, W., Haiber, U., Heisterkamp, P., and Scheible, S. (2004). The SENECA spoken language dialogue system. Speech Communication, 43(1-2), pp. 89-102.

Minsky, M. (1961). Steps towards artificial intelligence. Proceedings of the IEEE, 49, pp. 8-30.

Minsky, M. (2006). The Emotion Machine. Simon and Schuster.

Misu, T., and Kawahara, T. (2007). An Interactive Framework for Document Retrieval and Presentation with Question-Answering Function in Restricted Domain. LNAI, pp. 126-134.

Mollá, D., Schwitter, R., Rinaldi, F., Dowdall, J., and Hess, M. (2003). NLP for Answer Extraction in Technical Domains. Proceedings of EACL 03 Workshop: Natural Language Processing for Question Answering.

Montero, C., and Araki, K. (2006). Is It Correct? - Towards Web-Based Evaluation of Automatic Natural Language Phrase Generation. ACL 2006.

Montero, C., and Araki, K. (2007). Semi-supervised Algorithm for Human-Computer Dialogue Mining. Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing, pp. 61-64.

Mooney, R. J. (2006). Learning Language from Perceptual Context: A Challenge Problem for AI. Proceedings of the 2006 AAAI Fellows Symposium.

Murphy, C. (2010, May 17). No proof of mobile cancer risk, major study concludes. BBC News. Retrieved July 12, 2010 from http://www.bbc.co.uk.

Nakano, M., Kanda, N., Hasegawa, Y., Torii, T., Takeuchi, Y., Nakadai, K., Tsujino, H., and Okuno, H. (2005). A Two-Layer Model For Behavior And Dialogue Planning In Conversational Service Robots. IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1542–1547.

Nederhof, M., Bouma,G., Koeling, R., and van Noord, G. (1997). Grammatical analysis in the OVIS spoken-dialogue system. Interactive Spoken Dialog Systems: Bringing Speech and NLP Together in Real Applications, pp. 66-73.

Neumann, G. and Xu, F. (2003). Mining answers in German Web pages. Proceedings IEEE/WIC International Conference on Web Intelligence, pp. 125-131.

Nguyen, A. (2005). An Agent-Based Approach to Dialogue Management in Personal Assistants. IUI-2005, pp. 137-144.

Niesler, T., Roux, J. (2001). Natural language understanding in the DACST-AST dialogue system. 12th annual symposium of the Pattern Recognition Association of South Africa (PRASA), pp. 134-136.

Nirenburg, S., and Wilks, Y. (2001). What's in a symbol: ontology, representation and language. Journal of Experimental and Theoretical Artificial Intelligence, 13(1), pp. 9-23.

Nyberg, E., Mitamura, T., Frederking, R., Pedro, V., Bilotti, M. W., Schlaikjer, A. and Hannan, K. (2005). Extending the JAVELIN QA System with Domain Semantics. AAAI 2005.

O'Shea, K., Bandar, Z., Crockett, K. (2008). A Novel Approach for Constructing Conversational Agents using Sentence Similarity Measures. World Congress on Engineering 2008, 1, pp. 321-326.

Owda, M., Bandar, Z., and Crockett, K. (2007). Conversation-Based Natural Language Interface to Relational Databases. Web Intelligence/IAT Workshops, pp. 363-367.

Ozcan, R., and Aslandogan, Y. A. (2004). Concept-based Information Retrieval Using Ontologies and Latent Semantic Analysis. Technical Report CSE-2004-8.

Pallotta, V. (2004). Cognitive Dialogue Management. Proceedings of International Workshop on Natural Language Understanding and Cognitive Science.

Pallotta, V., and Ballim, A. (2001). Robust Dialogue Understanding in HERALD. Proceedings of RANLP 2001 - EuroConference on Recent Advances in Natural Language Processing.

Pan, X. (2002). A Context-Based Free Text Interpreter. M. S. Thesis in Computer Science, Cal-Poly, San Luis Obispo, California.

Parekh, V., Gwo, J., and Finin, T. (2004) Mining domain specific texts and glossaries to evaluate and enrich domain ontologies. Proceedings the International MultiConference in Computer Science and Computer Engineering, pp. 533-540.

Pastra, K., Saggion, H., and Wilks, Y. (2003). NLP for indexing and retrieval of captioned photographs. EACL03: 10th Conference of the European Chapter of the Association for Computational Linguistics.

Pazienza, M. T., Pennacchiotti, M., Vindigni, M., and Zanzotto, F. M. (2005). AI/NLP Technologies Applied to Spacecraft Mission Design. IEA/AIE, 3533, pp. 239-248.

Peñas, A., Verdejo, F., and Gonzalo, J. (2001). Corpus-based terminology extraction applied to information access, Proceedings of the Corpus Linguistics, 13.

Perez-Marin, D., Pascual-Nieto, I., Alfonseca, E., and Rodríguez, P. (2006). Automatic Identification of Terms for the Generation of Students Concept Maps. Proceedings of the International Conference in Multimedia and Information Technologies for the Education.

Pon-Barry, H., Clark, B., Schultz, K., Bratt, E., and Peters, S. (2004). Advantages of Spoken Language in Dialogue-based Tutoring Systems. 7th International Conference on Intelligent Tutoring Systems, pp. 390-400.

Popowich, F. (2005). Using text mining and natural language processing for health care claims processing. Natural language processing and text mining, 7(1), pp. 59-66.

Porzel, R., and Strube, M. (2002). Towards Context-dependent Natural Language Processing in Computational Linguistics for the New Millennium: Divergence or Synergy. Proceedings of the International Symposium, pp. 21-22.

Porzel, R., Zorn, H., Loos, B., and Malaka, R. (2006). Towards a separation of pragmatic knowledge and contextual information. ECAI-06 Workshop on Contexts and Ontologies.

Punyakanok, V., Roth, D., and Yih, W. (2004). Mapping Dependencies Trees: An Application to Question Answering. AI&Math Special Session: Intelligent Text Processing.

Quarteroni, S., and Manandhar, S. (2007). A Chatbot-based Interactive Question Answering System. Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue, pp. 83-90.

Ragnarsdottir, M. D., and Hvannberg, E. T. (2006). Using Language Technology to Increase Efficiency and Safety in ATC Communication. Journal of Aerospace Computing, Information, and Communication, 3.

Ravichandran, D., and Hovy, E. (2002). Learning Surface Text Patterns for a Question Answering System. 40th Annual Meeting on Association for Computational Linguistics, pp. 41-47.

Ravichandran, S., and Karthik, M. N. (2004). Autogenic Training With Natural Language Processing Modules: A Recent Tool For Certain Neuro Cognitive Studies. CoRR cs. AI/0407008.

Rayson, P., and Garside, R. (1999). Comparing corpora using frequency profiling. In Proceedings of the Workshop on Comparing Corpora, pp. 1-6.

Rettner, R. (2010, May 16). Burger & Fries Worsen Asthma, Study Suggests. LiveScience. Retrieved July 12, 2010 from http://www.livescience.com.

Rich, C., Sidner, C., Lesh, N., Garland, A., Booth, S., and Chimani, M. (2005). DiamondHelp: A Graphical User Interface Framework for Human-Computer Collaboration. 5th International Workshop on Smart Appliances and Wearable Computing.

Rindflesch, T. C. (1995). Integrating natural language processing and biomedical domain knowledge for increased information retrieval effectiveness. Proceedings of the 5th Annual Dual-use Technologies and Applications Conference, pp. 260-265.

Robinson, S., Traum, D., Ittycheriah, M., and Henderer, J. (2008). What would you ask a conversational agent? Observations of Human-Agent dialogues in a museum setting. Language Resources and Evaluation Conference.

Rosset, S., Galibert, O., Illouz, G., and Max, A. (2006). Integrating Spoken Dialog and Question Answering: the Ritel Project. InterSpeech'06, pp. 1914-1917.

Roth, D. (1998). Learning to Resolve Natural Language Ambiguities: A Unified Approach. Proceedings of AAAI-98, 15th Conference of the American Association for Artificial Intelligence, pp. 806-813.

Ruiz-Casado, M., Alfonseca, E., and Castells, P. (2006). From wikipedia to semantic relationships: a semi-automated annotation approach. Proceedings of the First Workshop on Semantic Wikis - From Wiki To Semantics.

Rzepka, R., Ge, Y., and Araki, K. (2005). Naturalness of an Utterance Based on the Automatically Retrieved Commonsense. Proceedings of Nineteenth International Joint Conference on Artificial Intelligence.

Saeki, S., and Gonzalez, A. J. (2000). Soft-coding the Transitions between Contexts in CGF's: The Competing Context Approach, Proceedings of the Computer Generate Forces and Behavior Representation Conference.

Sagae, K., Christian, G., DeVault, D. and Traum, D. (2009). Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems. Short paper proceedings of the North American chapter of the Association for Computational Linguistics - Human Language Technologies conference.

Sammut, C. (2001). Managing Context in a Conversational Agent. Electronic Transactions on Artificial Intelligence, 5(B), pp. 189-202.

Sanders, G., and Scholtz, J. (2000). Measurement and Evaluation of Embodied Conversation Agents. Embodied Conversational Agents, pp. 346-373.

Sansonnet, J., Leray, D., and Martin, J. (2006). Architecture of a Framework for Generic Assisting Conversational Agents. Lecture Notes in Computer Science, 4133, pp. 145-146.

Santangelo, A., Augello, A., Gentile, A., Pilato, G., and Gaglio, S. (2006). A Chat-bot based Multimodal Virtual Guide for Cultural Heritage Tours. PSC 2006, pp. 114-120.

Sarma, A., and Palmer, D. (2004). Context-based Speech Recognition Error Detection and Correction. HLT-NAACL 2004: Short Papers, pp. 85-88.

Sawyer, P., Rayson, P., and Garside, R. (2002). REVERE: support for requirements synthesis from documents. Information Systems Frontiers Journal. 4(3), pp. 343-353.

Schaefer, D. (2001). Context-Sensitive Speech Recognition in the Air Traffic Control Simulation. Presentation given at the Fourth International Air Traffic Management R&D Seminar ATM-2001.

Schank, R. C., and Abelson, R. P. (1977). Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures (Chap. 1-3). L. Erlbaum, Hillsdale, NJ.

Schofield, E., and Zheng, Z. (2003). A speech interface for open-domain question-answering. 41st Annual Meeting on Association for Computational Lingustics, 2, pp. 177-180.

Schulman, D., and Bickmore, T. (2009). Persuading users through counseling dialogue with a conversational agent. 4th International Conference on Persuasive Technology.

Schumaker, R., Liu, Y., Ginsburg, M., & Chen, H. (2007) Evaluating the Efficacy of a Terrorism Question Answer System: The TARA Project. Communications of the ACM, 50(7), pp. 74-80.

Semeraro, G., Andersen, H. H. K., Andersen, V., Lops, P., and Abbattista, F. (2003). Evaluation and Validation of a Conversational Agent Embodied in a Bookstore. Universal Access: Theoretical Perspectives, Practice and Experience, Lecture Notes in Computer Science, 2615, pp. 360-371.

Serridge, B. (1997). Context-Dependent Modeling in a Segment-based Speech Recognition System. M. Eng. thesis, MIT Department of Electrical Engineering and Computer Science, August 1997.

Shah, H. and Henry, O. (2005) Confederate Effect in Human Machine Textual Interaction. Proceedings of 5th WSEAS Int. Conf. on Information Science, Communications and Applications (ISCA 2005), pp. 109-114.

Shawar, A., and Atwell, E. (2005). A chatbot system as a tool to animate a corpus. ICAME Journal, 29, pp. 5-24.

Shawar, B. A., and Atwell, E. (2007). Different measurements metrics to evaluate a chatbot system. Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing.

Sherwell, B. W., Gonzalez, A. J. and Nguyen, J. (2005). Contextual implementation of human problem-solving knowledge in a real-world decision support system. Proceedings of the Conference on Behavior Representation in Modeling and Simulation.

Shi, Z., Gu, B., Popowich, F., and Sarkar, A. (2005). Synonym-Based Expansion and Boosting-Based Re-Ranking: A Two-phase Approach for Genomic Information Retrieval. The Fourteenth Text REtrieval Conference (TREC 2005) Proceedings.

Sidner, C. (2002). Building Spoken Language Collaborative Interface Agents. Practical Dialogue Systems.

Simperl, E. P. B., and Schlangen, D. (2006). Creating ontologies for content representation - the OntoSeed suite. International Journal of Data Semantics.

Sing, G. O., Wong, K. W, Fung, C. C., and Depickere, A. (2006). Towards A More Natural and Intelligent Interface with Embodied Conversation Agent. Joint International Conference on CyberGames and Interactive Entertainment.

Skantze, G., Edlund, J., and Carlson, R. (2006). Talking with Higgins: Research challenges in a spoken dialogue system. Proceedings of Perception and Interactive Technologies, pp. 193-196.

Soderland, S., and Mandhani, B. (2007). Moving from Textual Relations to Ontologized Relations. Proceedings of the 2007 AAAI Spring Symposium on Machine Reading.

Sonntag, D. (2006). Towards Combining Finite State, Ontologies, and Data Driven Approaches to Dialogue Management for Multimodal Question Answering. Information Society Language Technologies Conference.

Spärck Jones, K. (1964). Synonymy and Semantic Classification. Ph.D. thesis, University of Cambridge, Cambridge, England.

Stede, M., and Schlangen, D. (2004). Information-Seeking Chat: Dialogues Driven by Topic-Structure. Proceedings of Catalog '04 (The 8th Workshop on the Semantics and Pragmatics of Dialogue, SemDial04).

Stensrud, B. S., Barrett, G. C., Trinh, V. C., and Gonzalez, A. J. (2004). Context-Based Reasoning: A Revised Specification. FLAIRS Conference 2004.

Stephens, R. (2010, May 17). Celtics Hold Off Magic In Game One of the Eastern Conference Finals. The Washington Informer. Retrieved July 12, 2010 from http://www.washingtoninformer.com.

Stibler, K., and Denny, J. (2001). A three-tiered evaluation approach for interactive spoken dialogue systems. Proceedings of the first international conference on Human language technology research, pp. 1-5.

Stoness, S. C., Allen, J., Aist, G., and Swift, M. (2005). Using real-world reference to improve spoken language understanding. Proceedings of Workshop on Spoken Language Understanding, AAAI05.

Streiter, O. (2001). Recursive top-down fuzzy match: New perspectives for memory-based parsing. PACLIC 2001.

Suh, S., Halpin, H., and Klein, E. (2006). Extracting Common Sense Knowledge from Wikipedia. ISWC2006 Workshop on Web Content Mining with Human Language technology.

Surdeanu, M., Dominguez-Sal, D., and Comas, P. R. (2006). Design and Performance Analysis of a Factoid Question Answering System for Spontaneous Speech Transcriptions. Proceedings of the Ninth International Conference on Spoken Language Processing (INTERSPEECH 2006).

Takeda, K., Fujimura, H., Itou, K., Kawaguchi, N., Matsubara, S., and Itakura, F. (2005). Construction and Evaluation of a Large In-Car Speech Corpus. IEICE Transactions 88-D(3), pp. 553-561.

Tarau, P., and Figa, E. (2004). Knowledge-Based Conversational Agents and Virtual Storytelling. ACM Symposium on Applied Computing, pp. 39-44.

Thiel, U., and Stein, A. (2000). Intelligent E-Commerce with Guiding Agents based on Personalized Interaction Tools. E-business: Key Issues, Applications and Technologies, pp. 760-766.

Thomas, C., and Sheth, A. P. (2006). Predicting Domain Specific Entities with Limited Background Knowledge. LSDIS Technical Report.

Thórisson, K. (1999). A Mind Model for Multimodal Communicative Creatures and Humanoids. International Journal of Applied Artificial Intelligence, 13(4-5), pp. 449-486.

Tomko, S., and Rosenfeld, R. (2004). Shaping Spoken Input in User-Initiative Systems. Proc. ICSLP 2004.

Towhidnejad, M. (1990). Functional Conflict Resolution in Automated Knowledge Generation. Ph. D. Thesis, University of Central Florida.

Townsend, J., Copestake, A., Murray-Rust, P., Teufel, S., and Waudby, C. (2005). Language Technology for Processing Chemistry Publications. Proceedings of UK e-Science All Hands Meeting.

Turunen, M., Halkulinen, J., Smith, C., Charlton, D., Zhang, L., and Cavazza, M. (2008). Physically Embodied Conversational Agents as Health and Fitness Companions. INTERSPEECH-2008, pp. 2466-2469.

Urbanová, L. (2001). On the Language of Authentic English Conversation. Brno Studies in English, 27, pp. 49-55.

Varges, S., Weng, F., and Pon-Barry, H. (2009). Interactive question answering and constraint relaxation in spoken dialogue systems. Natural Language Engineering, 15(1), pp. 9-30.

Vogel, D. (2003). Using Generic Corpora to Learn Domain-Specific. Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003).

Vrajitoru, D., and Ratkiewicz, J. (2004). Evolutionary Sentence Combination for Chatterbots. The IASTED International Conference on Artificial Intelligence and Applications (AIA 2004), pp. 287-292.

Walker, M. A., Litman, D. J., Kamm, C. A., and Abella, A. (1997). PARADISE: a framework for evaluating spoken dialogue agents. Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, pp.271-280.

Walker, T. (2004). Automating the Extraction of Domain Specific Information from the Web - a Case Study for the Genealogical Domain. M. S. Thesis, Brigham Young University.

Wallace, R. (2002). AIML Pattern Matching Simplified. Technical report, A.L. I.C. E. AI foundation.

Wallis, P. (2005). Believable Conversational Agents: Introducing the Intention Map. Proceedings of Creating Bonds with Humanoids Workshop at AAMAS'05.

Wallis, P., and Pelachaud, C. (2005). The Trouble with Chatbots: social skills in social world. AISB05.

Weizenbaum, J. (1966). ELIZA-a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), pp. 36-45.

Wilks, Y. (2005) The History of Natural Language Processing and Machine Translation, In Encyclopedia of Language and Linguistics, Kluwer, Amsterdam.

Wilks, Y., and Fass, D. (1992). Preference semantics: a family history. Computing and Mathematics with Applications, 23(2).

Williams, S. (1996). Dialogue management in a mixed-initiative, cooperative, spoken language system. 11th Twente Workshop on Language Technology Dialogue Management in Natural Language System.

Winograd, T. (1980). What does it mean to understand language? Cognitive Science, 4, pp. 209-241.

Wobcke, W., Ho, V., Nguyen, A., and Krzywicki, A. (2005). A BDI Agent Architecture for Dialogue Modelling and Coordination in a Smart Personal Assistant. NICTAHCSNet Multimodal User Interaction Workshop.

Wong, B., and Starner, T. (2001). Conversational speech recognition for creating intelligent agents on wearables. Human Computer Interface International.

Xiao, C., and Rösner, D. (2003). Finding high-frequent synonyms of domain-specific verbs in the English sub-language of MEDLINE abstracts using WordNet. Proceedings of 2nd Global WordNet Conference, pp. 242-247.

Yaccoub, S. (2010, May 18). Iraq security forces detains al-Qaida militant suspected of plotting attack on World Cup. Associated Press. Retrieved July 12, 2010 from http://www.thespec.com.

Yan, P., and Zheng, F. (2004). Context Directed Speech Recognition in Dialogue Systems. International Symposium on Tonal Aspects of Languages With Emphasis on Tone Languages, pp. 225-228.

Young, S. (1989). The MINDS system: using context and dialog to enhance speech. Human Language Technology Conference Workshop on Speech and Natural Language, pp. 131-136.

Zavrel, J., and Daelemans, W. (2003). Feature-Rich Memory-Based Classification for Shallow NLP and Information Extraction. Text Mining, Theoretical Aspects and Applications, Springer Physica-Verlag, pp. 33-54.

Zdravkova, K. (2000). Conceptual Framework for an Intelligent Chatterbot. 22nd International Conference on Information Technology Interfaces, pp. 189-194.

Zesch, T., Gurevych, I. (2007). Analysis of the Wikipedia Category Graph for NLP Applications. Proceedings of of the TextGraphs-2 Workshop (NAACL-HLT 2007).