Different Degrees of Explicitness in Intentional Artifacts: Studying User Goals in a Large Search Query Log

Markus Strohmaier
Graz University of Technology and Know-Center
Inffeldgasse 21a, 8010 Graz, AUSTRIA
markus.strohmaier@tugraz.at

Peter Prettenhofer
Know-Center
Inffeldgasse 21a, 8010 Graz, AUSTRIA
pprett@know-center.at

Mathias Lux
Klagenfurt University
Universitätsstraße 65-67, 9020 Klagenfurt, AUSTRIA
mlux@itec.uni-klu.ac.at

ABSTRACT
On the web, search engines represent a primary instrument through which users exercise their intent. Understanding the specific goals users express in search queries could improve our theoretical knowledge about strategies for search goal formulation and search behavior, and could equip search engine providers with better descriptions of users' information needs. However, the degree to which goals are explicitly expressed in search queries can be suspected to exhibit considerable variety, which poses a series of challenges for researchers and search engine providers. This paper introduces a novel perspective on analyzing user goals in search query logs by proposing to study *different degrees of intentional explicitness*. To explore the implications of this perspective, we studied two different degrees of explicitness of user goals in the AOL search query log containing more than 20 million queries. Our results suggest that different degrees of intentional explicitness represent an *orthogonal dimension* to existing search query categories and that understanding these different degrees is essential for effective search. The overall contribution of this paper is the elaboration of a set of theoretical arguments and empirical evidence that makes a strong case for further studies of different degrees of intentional explicitness in search query logs.

Author Keywords
Web search, user goals, query log analysis, AOL search database

ACM Classification Keywords
H3.3: Information storage and retrieval: Information search and retrieval, H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Studying users' goals on the web in general and in web search in particular has received increasing attention by scientists as well as industry recently [13,16,22]. While industry has a strong interest in learning more about user goals in order to provide better search results, enable more targeted ad campaigns or increase click-through rates, the research community aims to develop a profound theoretical understanding about the different types of goals users have on the web [4], how users express their goals [25], how goals can be identified automatically and how goal-orientation can be used to facilitate human-computer interaction [8].

The enormous power that search engines, such as Google, Yahoo and Microsoft Live, have today has been described by John Batelle in 2003 with the notion of so-called "databases of intentions"[1]. This notion refers to the fact that user goals, something sensitive and private for users for a very long time, have become explicit and – to a certain extent - public with the advent of powerful search engines on the web. John Batelle describes databases of intentions as *"the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result. […]. This information represents […] a place holder for the intentions of humankind - a massive database of desires, needs, wants, and likes that can be discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends. Such a beast has never before existed in the history of culture […]."*

---

[1] http://battellemedia.com/archives/000063.php, last accessed Nov 21, 2007

What has received only little attention so far is that the intentions represented in such "databases of intentions" can be suspected to *exhibit considerable variety* with respect to their degree of explicitness. While some goals contained in search queries might be very explicit, other queries might contain more implicit goals, which would mean that they are more difficult to recognize by, for example, an external observer. To give an example: in terms of intentional explicitness, the query "car miami" differs significantly from the query "buy a used car in Miami".

While this observation appears rather intuitive, to the best of our knowledge there is no research effort comprehensively studying different degrees of intentional explicitness in search query logs, although the implications seem profound: different degrees of intentional explicitness could put significant constraints on the general *analyzability* and ultimately the overall *utility* of so-called databases of intentions, and they could put an *upper bound on the level of service that search engines can provide*. As a result, studying different degrees of intentional explicitness in search queries appears relevant on at least two different levels:

- On a theoretical level, better understanding different degrees of intentional explicitness in search queries could increase our knowledge about the levels of abstractions users employ when searching, and could equip us with better distinctions and tools for studying, for example, the way users refine or generalize goals during search.
- On a practical level, understanding different degrees of intentional explicitness in search queries could improve the ability of search engine vendors to better tailor their search results to specific users and to link search queries at different levels of explicitness.

However, understanding the degree of explicitness of user goals in search queries poses significant research and technical challenges: First and foremost, all goals contained in search query logs are of hypothetical nature in the sense that verification is extremely hard – if not impossible. Most query logs that are available to researchers have been anonymized, and even if information about the users would be available, contacting and verifying hypothetical goals would be costly or hardly feasible due to geographical, time and other constraints. We refer to this problem as the *goal verification problem*, which is extremely hard to overcome in research on search query log analysis. Second, query logs represent huge text corpora in terms of size, which renders manual elicitation of goals by experts practically impossible. We refer to this problem as the *goal elicitation problem*. Furthermore, query logs represent a fundamentally different text corpus to mine goals from, compared to other corpora that have been studied from an intentional perspective, such as interview transcripts or organizational guidelines: The length of search queries is significantly shorter, the words used in search queries do not necessarily appear in lexica, and the text is not necessarily represented as natural language text but in some artificial language, such as an arbitrary concatenation of terms that users suspect to yield to fruitful and relevant search results (such as "car miami"). We refer to this problem as the *linguistic artificiality problem*.

While solving all of these problems in their entirety is well beyond the scope of this work, in this paper we aim to 1) increase our understanding about the notion of different degrees of explicitness in intentional artifacts *theoretically*, and 2) explore related challenges, potentials, and implications *empirically*. For that purpose, we have adopted selected concepts from the body of literature related to the notion of goals in different research areas and conducted an exploratory study of a large search query log: the AOL search database released in 2006.

WHAT ARE GOALS? DEFINITION AND RELATED WORK

To establish a theoretical understanding about the fundamental constructs we work with, we introduce the following definitions based on related work in a series of different, but related research areas. The most central concept in our paper is the concept of a *goal*, which we define in our paper as *"a condition or state of affairs in the world that some agent would like to achieve or avoid. How the goal is to be achieved or avoided is typically not specified, allowing alternatives to be considered*" (based on [21]). An *intentional artifact* is an electronic artifact produced by users or user behaviour that contain *recognizable* "*traces of intent*", i.e. traces of users' goals and intentions expressed in different degrees of explicitness. The degree to which these traces can be recognized as goals by some independent observer depends on the artifact's degree of intentional explicitness. In this paper, we assume that search query logs at large represent intentional artifacts, meaning that they contain such traces of intent at different levels of explicitness. Examples for search queries exhibiting different degrees of intentional explicitness are shown in Figure 1.

```
car, car Miami, car Miami dealer,
buy a car in Miami, buy a used car in
Miami, get loan to buy a used car in Miami
```

Figure 1. Queries with different degrees of explicitness

The notion of goals has been used by researchers in different areas to represent and frame the desires and needs users have when interacting with software. In the following, we will discuss selected research relevant to our work.

The Notion of Goals in Human Computer Interaction

Researchers have focused on studying user intentions long before the current popularity of search engines, query log analysis and the web in general. In the broader human-computer interaction (HCI) context, Norman's theory of action [19], for example, describes the inherent gap between a person's goals and intentions and a system's capabilities, features and structures. Norman's research has implicitly acknowledged the existence of different degrees of explicitness in users' goals by highlighting that user goals are often *not well specified*, *opportunistic*, *ill-formed* and *vague* and therefore hard to capture, identify and represent. Any attempt studying goals in a web search context must be suspected to face similar, if not the same, challenges. Other work in HCI identifies basic types of so-called Goal-Effect Problems, i.e. problems that characterize system performance from an intentional perspective. In their paper [23] the authors distinguish between (I) Missing cues for goal construction, where a system does not suggest appropriate goals (II) Misleading cues for goal construction, where a system suggests irrelevant goals (III) Missing cues for goal elimination, where a system does not eliminate completed goals, and (IV) misleading cues for goal elimination, where a system does eliminate incomplete goals. Translated to a web search context, these distinctions highlight some of the implications of search queries expressed on different levels of intentional explicitness. Further work in HCI, such as the work of [12] on the Lumiere project, focuses particularly on studying intentional artifacts with a low degree of explicitness.

The Notion of Goals in Requirements Engineering

Goal Oriented Requirements Engineering (GORE) conceptualizes software development as a process that aims to satisfy a series of stakeholder goals. The corresponding research community distinguishes between different types of goals such as: *achieve* and *cease goals*, which are said to generate behavior, *maintain* and *avoid goals*, which are said to restrict behaviors as well as *optimize goals*, which are said to compare behaviors [21]. The distinction between goals and softgoals in GORE can be seen as an indicator for the plausibility of studying different degrees of explicitness in goals. While, for example, in the i* framework [29] a *goal* has a clear cut criteria, a *softgoal* describes a goal for which there is no such clear-cut criterion to be used for deciding whether it is satisfied or not.

The Notion of Goals in Web Search

On the web, search represents a primary instrument through which users exercise their intent. This allows search engines to have a tremendous corpus of *intentional artifacts* at their disposal. This observation has led scientists to focus on studying user intentions in search query logs. In 2002, Broder [4] has introduced a high level categorization of web search intent, distinguishing between navigational, informational and transactional queries. Based on this early work, Rose and Levinson [22] have refined this categorization into a hierarchical taxonomy including more fine grained categories, such as entertainment or advice seeking. In 2004, [16] have presented an automatic approach that aims to tell navigational and informational goals apart based on analyzing two parameters: user-click behavior and anchor-link distribution. Baeza-Yates et al apply supervised and unsupervised learning techniques to study users' goals in search query logs [2]. Faaborg [8] has presented a prototype for goal-oriented browsing and Liu et al [17] have presented a prototype for goal-oriented search based on intentional concepts retrieved from the ConceptNet commonsense knowledge base.

While state-of-the-art research offers a set of useful categories, techniques and prototypes, *we consider the degree of intentional explicitness to be orthogonal to existing intentional categories of search queries*. In other words, we assume that within *each intentional category* (such as informational or transactional queries), goals can be expressed in different degrees of intentional explicitness. Broder, for example, makes a similar point in his 2002 paper, by mentioning that "*many informational queries are extremely wide, for instance* cars *or* San Francisco*, while some are narrow, for instance* normocytic anemia, Scoville heat units". Our work in this paper is motivated by a desire to characterize different degrees of *intentional explicitness* in search query logs, and identifying implications for the process of search. Our own previous work explored how users express their goals during search [25].

Further related work has acknowledged this problem to some extent: in the paper of [22], for example, a tool that aims to support experts in categorizing search queries into goal categories is presented. While different degrees of intentional explicitness were not in the explicit focus of this work, the development of the tool can be interpreted as an early recognition of the problems that researchers face with different degrees of intentional explicitness in search queries.

DEGREES OF EXPLICITNESS IN INTENTIONAL ARTIFACTS

In a web search context, we conceptualize the degrees of explicitness in intentional artifacts to represent a broad, continuous spectrum. On one end of this spectrum, we would have queries that describe the users' intent completely and precisely, with nothing to add from an intentional perspective. On the other end of the spectrum we would have queries that do not describe user intent at all, such as blank queries.

For reasons of simplicity, in this paper we propose to distinguish – at a high, dichotomous level – between two degrees of intentional queries only: explicit and implicit intentional queries. This allows us to study whether a distinction between implicit and explicit intentional queries is reasonable in a web search context in the first place, and whether it yields interesting insights or implications. Given that we can identify interesting differences between different degrees of intentional explicitness, it could be interesting to conduct research on more refined definitions and more fine grained degree distinctions in the future. With these arguments in mind, we introduce the following *idealized definitions* of explicit and implicit intentional query. An **explicit intentional query** is a query that can be related to a specific goal in a *recognizable, unambiguous* way. Recognizable refers to what [15] defines as "trivial to identify" by a subject within a given attention span. On a more practical level, this idealized definition is related to what other researchers have characterized as "better queries", or queries that have "more precise goals" (R. Baeza-Yates at the "Future of Web Search" workshop 2006, Barcelona). Examples of explicit intentional queries, i.e. queries that have more precise goals, would be "buy a car", "maximize adsense revenue" or "how to get revenge on neighbor within limits of law". While these queries can still be refined and elaborated, they are more unambiguous in a sense that a user searching for "how to get revenge on neighbor within limits of law" is unlikely to have the true goal of "buy a nice gift for neighbor". We define an **implicit intentional query** as a query where it is difficult or extremely hard to elicit some specific goal from the intentional artifact. Examples include blank queries, or queries such as "car" or "travel", which embody user goals on a very general level. Queries on this kind of level are likely to require further refinement in order to yield useful search results. Interestingly, a significant proportion of queries today are of length 1 or 2 (as it is evident in, for example, the AOL search database set [20]).

Distinguishing between these two broad types of queries is important for several reasons: First, explicit ("better") intentional queries could be used to disambiguate or refine *implicit* intentional queries. For example: a search engine might be able to refine the implicit intentional query "car shop" with the explicit intentional queries "shop for a car", "repair a car", "find a car shop" or "buy a car for shopping" with the help of user interaction. Second, we have found anecdotal evidence that some users organize their search in a way that can be understood as a traversal of goal graphs [25], including iterative goal refinement and generalization. This suggests that switching between more explicit and more implicit intentional queries during search is a natural cognitive activity for at least some users. Third, our own recent research has indicated that only 1.69% to 3.01% of queries have a high degree of intentional explicitness [25]. While this percentage is rather small, we do not know whether users prefer to search via implicit intentional queries, or whether users have simply adapted to the non-intentional mode in which Google, Yahoo and other search engines operate today (cf. "bag-of-word principle"). Our research is driven by a desire to understand whether explicit intentional queries have the potential to *narrow the cognitive gap between a user's goals and the queries she uses*. We are interested in the implications of distinguishing between explicit and implicit intentional queries and in learning more about the explicit goals users have on the web, with the long term vision of enabling users to more accurately express their goals in search in the long run (towards "better queries" in Baeza Yates' diction).

This is in contrast to some past work in information retrieval, for example in the area of query expansion, where the purpose of query expansion is to make the user query *resemble* more closely the *documents* it is expected to retrieve [26]. Our interest is rather the opposite: Because the precision with which users describe their goals in search queries puts an upper bound on the level of service search engines can provide, our long term interest is to make search queries *resemble* more closely the *intentions* users have (moving towards more explicit intentional queries). This could help to narrow the "gulf of execution" for users, and could help computer scientists and search engine vendors to work with more accurate descriptions of users' intent – something search engine vendors are desperate to achieve today [10]. While some researchers have already attempted to address similar issues, [1], our particular focus lies in exploring different degrees of *intentional* explicitness in large search query logs rather than ambiguity of queries in general.

AN EXPLORATORY STUDY

Equipped with a theoretical understanding about explicit and implicit intentional queries, we are now interested in empirically studying these different types of queries "in the wild". In an exploratory study, we aim to identify and better understand explicit intentional queries in the AOL search database, a large search query log database released in 2006. We want to explore whether there are differences between explicit and implicit intentional queries with respect to, for example, the number of users issuing these types of queries or the type of URLs clicked as a result. Furthermore, we were interested in learning whether there are certain words that indicate the presence of explicit intentional queries, which could represent a relevant finding for future research efforts.

Although our preliminary distinction between explicit and implicit intentional queries equips us with an intuitive criterion for classification, a sharper measure is needed to separate explicit from implicit intentional queries on an operational level. To simplify classification, we distinguish between explicit and implicit intentional queries based on the following *arbitrary criteria* A) whether a query contains at least one verb and B) whether the goal elicited from the intentional artifact conforms to our definition of a goal. Note that for other or more refined degrees of intentional explicitness, different criteria might be used. We are now using our previous example of queries to illustrate the implications of our particular distinction in Figure 2, where queries in bold represent explicit intentional queries according to our classification criteria.

```
Car, car Miami, car Miami dealer, buy a car in Miami, buy a used car in Miami, get loan to buy a used car in Miami
```

Figure 2. Distinguishing different degrees of explicitness

While our example might imply that the degree of explicitness correlates with query length only, it does not necessarily. Although the query "buying a car in the 1920's" contains a verb, it does not conform to our definition of a goal and would therefore not be considered to represent an explicit intentional query. Our criteria thus allow to distinguish between "buy a car" or "sell a car" (explicit) and "car dealer ads" (implicit). We are aware of the implications of this simplification, and we discuss them in the "Threats to validity" section at the end of this paper.

We investigated explicit and implicit intentional queries in the AOL search database. In addition to the AOL data, several other web search logs are available [13]. We used the AOL search database because it provides a very large dataset including comprehensive information about anonymous user IDs, time stamps, search queries, and click-through events. It contains ~ 20 million search queries collected from 657,426 unique user ID's between March 1, 2006 and May 31 2006 by AOL. To our knowledge, the AOL search database is also the most recent very large corpus of search queries publicly available (2006)[2]. Because applying our definition of explicit and implicit intentional queries manually to the AOL dataset with more than 20 million queries is infeasible (cf. *the goal elicitation problem*), we have developed an experimental classification approach based on a training set of queries that was used for machine learning syntactical features of explicit intentional queries. However, coming up with an automatic classifier that excels on precision and recall measures would be well beyond the scope of this paper. Instead, our approach focuses on providing us with a reasonable subset of the AOL query dataset that contains a significant higher proportion of explicit intentional queries than the entire dataset. Therefore, the goals of our experimental classification approach are more modest: it should enable us to gain a better understanding about explicit and implicit intentional queries and aid us in coupling our intuitions with empirical data. Focusing on better classification approaches could represent a promising line of future research. In the next section, we will describe some technical details of our approach.

An Experimental Classification Approach

Before using the dataset for our analysis, we sanitized it with respect to undesirable properties such as empty queries. The data representation of an entry resulting from our sanitation process has the following form: {UserID, query, timestamp, (ItemRank, URL)*}. Taking this data representation as an input, our experimental classification approach consists of two parts: part-of-speech (POS) tagging and supervised learning of syntactical goal features.

Part of Speech Tagging

Our classification approach is based on the simplified assumption that explicit intentional queries can be distinguished from implicit intentional queries by the occurrence of certain part-of-speech patterns. For this purpose the experimental setup incorporated a fast and reasonably accurate bigram part-of-speech tagger trained on a sample of the Penn Treebank corpus. We have focused on tagging queries with query length > 2 only, because of the inherent ambiguity of shorter queries, and the resulting difficulty of recognizing goals. We favored a bigram tagger over more powerful approaches such as transformation-based taggers and Hidden Markov Model taggers due to efficiency issues, the lack of contextual information and the rather naive (artificial) linguistic nature of search queries (cf. the *linguistic artificiality problem*). The tag set of the Penn Treebank corpus consists of 45 word classes [14]. The reason for choosing this particular tag set is the fact that we are mainly interested in identifying verbs and verb noun combinations. For our purpose, we don't need the finer grained word classes provided by e.g. the tag set of the brown corpus or C7. Table 1 shows a sample of word classes of the Penn Treebank tag set.

Tag	Description	Example
NN	Noun, sing. or mass	car
VB	Verb, base form	eat
VBG	Verb, gerund	eating
VBZ	Verb, 3sg pres	eats
JJ	Adjective	yellow
WRB	Wh-adverb	how, where
TO	"to"	to

Table 1. A sample of Penn Treebank tags (from [14])

The vocabulary size of the corpus is an estimated number of 13,500 words, which is rather small compared to the expected vocabulary size of the dataset (cf. the *linguistic artificiality problem*). To address this problem, we have chosen a suffix tagger as a back off strategy for the bigram tagger. The part-of-speech tagging functionality we used was provided by the natural language toolkit NLTK [18].

Supervised Learning of Goal Features

Our classification approach is similar to those reported in [5,9,11]. However, we use part-of-speech n-grams instead of word n-grams as features. In our experimentation we used binary features based on fixed size trigrams. Furthermore, we introduced markers ($ $) at the beginning and the end of a query to take the query boundary part-of-speeches into account. Thus, the query "*buying/VBG a/DT car/NN*" would be composed of the following trigrams:

$ $ VBG, $ VBG DT, VBG DT NN, DT NN $, NN $ $

[2] Because the AOL search database was retracted from AOL shortly after releasing it, we obtained a copy from a secondary source: http://www.gregsadetsky.com/aol-data/ last accessed on July 15th, 2007.

To obtain a training set, we drew a uniform random sample from the set of queries which contain at least one verb[3]. Two of the authors labeled instances in the sample consensually based on whether the queries conform to our definition of goals introduced earlier. This resulted in a training set consisting of 98 instances, 59 positives and 39 negatives. While this training set is not necessarily representative for the set of all queries under investigation, it yielded sufficient results given the exploratory nature of our research.

We trained a naive bayesian classifier [7] on the feature vectors described above using 10-fold cross-validation. In order to increase the performance of our classifier we applied a chi-squared feature selection algorithm to our training set [24]. The best results, based on 10-fold cross-validation, were achieved by reducing the feature space to the 20 most predictive features. Table 2 shows the most predictive features according to the feature selection.

$ $ NN	$ $ VBG
$ WRB TO	WRB TO VB
$ NN NN	$ VBG DT
VBG DT NN	$ VBG NN
$ $ VBZ	JJ NN $
$ VBG IN	VBG IN NN
$ VB NN	TO VB VBN

Table 2. Most predictive features based on chi-squared feature selection

The purpose of our classification technique is to provide us with a more condensed set of queries - ideally containing a higher proportion of explicit intentional queries than the entire dataset – that would allow us to study explicit intentional queries in greater detail. More sophisticated linguistic techniques such as *selectional preference* [3] might be more adequate if the goal would be doing classification with a stronger focus on precision and recall measures. For all feature selection and classification tasks, we used the WEKA toolkit [27] in our work.

In the next section, we present the results of applying our experimental classification approach to the AOL search database.

[3] 1,598,612 out of 20,494,002queries contained at least one verb according to the outcome of our part-of-speech tagging process.

STUDY RESULTS

Results of Experimental Classification

Applying our technique resulted in a condensed set of queries containing 279,260 queries. We will refer to this set of queries from here on as the "condensed dataset". The condensed dataset contains a higher proportion of explicit intentional queries than the entire dataset. The difference is significant: While the set of explicit intentional queries in the entire dataset has been estimated to lie between 1.69% and 3.01%, in the condensed dataset we estimate this ratio (based on a sample containing 500 random queries from this set) to be in a 95% confidence interval of 49.6% and 58.4%. This allows us to compare whether there are interesting differences in query sets that contain a large as opposed to a very small proportion of explicit intentional queries.

	Entire Dataset	**Condensed Dataset**
Queries	20,494,002	279,260
Explicit Intentional Queries	346,349-616,869	138,513-163,089
Implicit Intentional Queries	19,877,133-20,147,653	116,172-140,747
Explicit Intentional Queries, 95% confidence interval	1.69% - 3.01%	49.6% - 58.4%
Users	657,426	94,487

Table 3. Statistical overview of the condensed dataset

Table 3 gives an overview of some statistics of our condensed dataset. It also shows that the condensed dataset captures only part of the explicit intentional queries estimated in the entire dataset. However, the dataset provides a subset of queries with a significantly higher proportion of explicit intentional queries, which is sufficient for the kind of exploratory research questions we are interested in.

Correctly Classified Intentional Queries
"buying groceries online"
"how to get revenge on neighbor within limits of law"
"helping children handle death of a loved one"
"cleaning the ak-47"
"coughing up blood"
"dealing with the guilt of cheating"

Table 4. Examples of correctly classified queries

In addition to the statistical analysis, we want to give a qualitative account of the type of queries our technique classified correctly and incorrectly in the condensed dataset.

Examples of correctly classified queries in the condensed dataset, are depicted in Table 4. These queries all represent goals that contain at least one verb and conform to our definition of goals. In addition, the set of correctly classified explicit intentional queries does not belong to a single query category (such as the ones identified in previous research [10]), but spans several of them. "buying groceries online" for example can be categorized as a transactional query, while "helping children handle death of a loved one" can be categorized as an informational query. This observation, together with the observation that implicit intentional queries do not belong to a single category either, illustrates that the degree of intentional explicitness represents an orthogonal view to existing categories in query log analysis. Another particularly interesting query is the instance, "coughing up blood". Although conforming to our definition of a goal, it represents a rather different kind of goal compared to the other goals identified in the condensed dataset: it represents an avoid goal of a user, describing a state which the user presumably tries to change (presumably a medical symptom). Automatically distinguishing between achieve and avoid goals appears to be an interesting research question and a non-trivial research challenge. The other goals in our table represent achieve goals in a sense that a user can be reasonably suspected to pursue the goal which is represented in the query (within the limitations of the *goal verification problem*).

Examples of incorrectly classified queries are especially interesting, as they show some of the limitations of our experimental classification approach:

Incorrectly Classified Intentional Queries
"saving privat ryan"
"driving school Illinois"
"stem cell transplant"
"founding fathers temple"
"recovering the satellites lyrics"

Table 5. Examples of incorrectly classified queries

The small sample of queries listed in Table 5 gives a good overview of the challenges of identifying explicit intentional queries: "Saving private ryan", for example, is a popular Hollywood movie starring Tom Hanks, which makes it unlikely that the user issuing the query has the goal of actually saving a Private named Ryan. "Driving school Illinois" probably refers to some school where people can learn to drive, rather than the goal of driving to school in Illinois. "stem cell transplant" is very likely not a goal either. The incorrect classification is likely the result of imperfections on the part-of-speech tagging part.

Finally, we observed a significant proportion of queries that appear goal-oriented, but have the term "lyrics" as a pre- or postfix, such as "recovering the satellites lyrics" (a song performed by the Counting Crows). Utilizing domain knowledge (such as an Amazon API to detect movie or book titles) can represent one way for dealing with such kind of queries.

Results of Comparing the two Datasets

We also investigated whether the most popular websites (i.e. websites that have been selected by users as a result of their search) in our condensed dataset differ from the most popular websites in the entire search query log. If this would be the case, it would make a strong argument for the development of more advanced algorithms and techniques that have higher precision in distinguishing between different degrees of intentional explicitness in search queries.

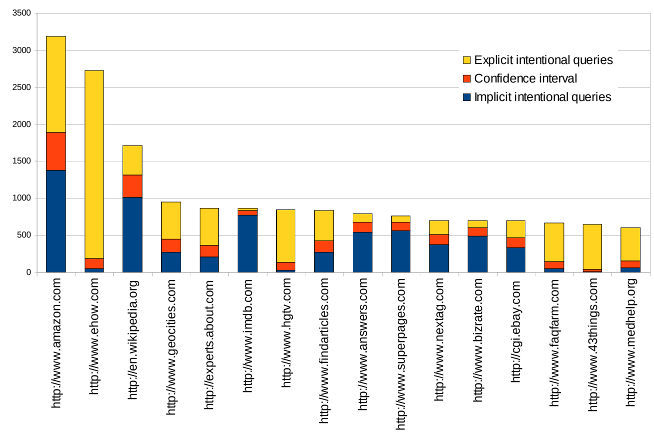


Figure 3. Top 16 websites in the condensed dataset

The histogram in figure 3 lists the top 16 websites that have been clicked by users in the condensed dataset, including websites such as amazon.com, ehow.com, en.wikipedia.org, geocities.com, medhelp.org and others.

We have taken a random sample from each set of queries associated with a URL listed in Figure 3 and evaluated it with respect to correctly and incorrectly classified queries. We calculated the 95% confidence interval of the error rate to give an estimate (middle part of each bar in figure 3). This kind of analysis revealed interesting differences: The websites that have highest proportion of correctly classified explicit intentional queries among the top 16 websites are websites that can be considered to be very goal-centric: 43things.com (a website encouraging users to share their goals in life), ehow.com (a website on how to accomplish a broad variety of tasks and goals), hgtv.com (a home improvement website), faqfarm.com (a question answering website), and medhelp.org (a medical information website). Medhelp.org is a particularly interesting result, as a large proportion of the correctly classified explicit intentional queries are queries describing medical symptoms ("coughing up blood"), which we defined as avoid goals.

The websites with a higher proportion of incorrectly classified explicit intentional queries are interestingly websites that are less goal centric such as imdb.com (a movie database, many queries were movie or series titles like “saving private ryan”, “bowling for columbine” or “meet joe black”), superpages.com (a directory website), followed by bizrate.com (a comparison shopping site, many queries for goods such as “marble fitted table cloth” or “fencing for pools”), answers.com (an online dictionary and encyclopedia, many queries focusing on definitions such as “meaning of centimeter” or “define alamo war”) and en.wikipedia.org (an online encyclopedia).

Especially amazon.com – the website associated with the highest number of queries in the condensed set – was difficult to interpret. Book titles often contain goals in their titles and it is hard to judge whether a user is searching for the specific book or using a goal as search query (e.g. “organizing your life” might be a search for the book “The Complete Idiot's Guide to Organizing Your Life”, which can be found at amazon.com). Geocities, which is a hosting company for a variety of web sites has a similar fraction of intentional queries, and is very broad regarding the range of topics identified in the queries.

In the following, we compare the entire and the condensed dataset with respect to whether they differ in the set of websites users select as a result of issuing queries.

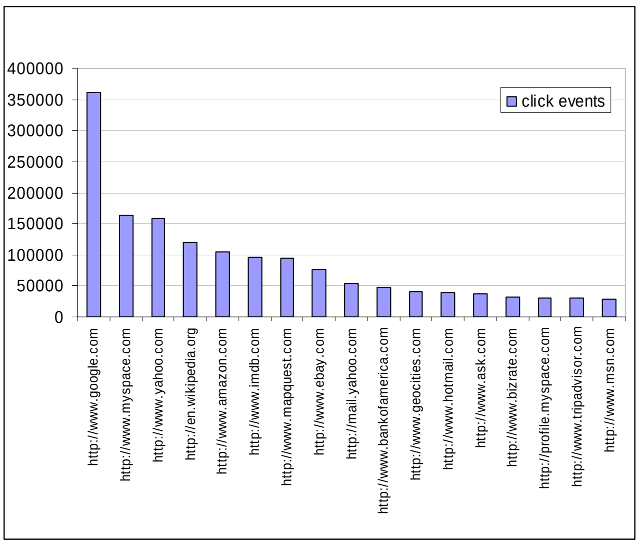
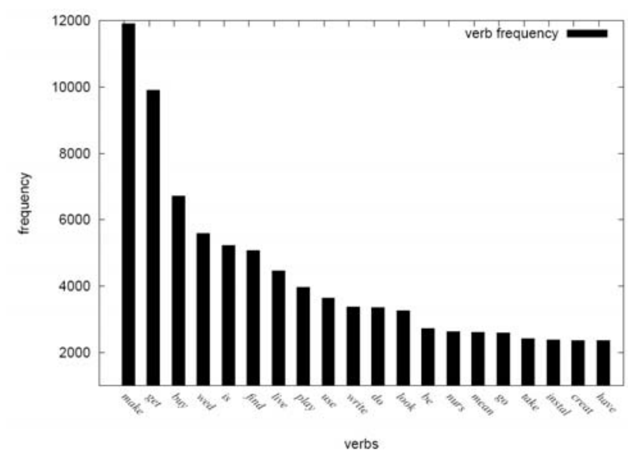


Figure 4. Top 16 websites in the entire dataset

In figure 4, we can see the list of top 16 websites that have been clicked by users in the entire search result set. The results differ significantly from the top 16 in the condensed dataset. Especially goal centric websites are affected by our experimental classification approach, such as 43things.com (moving from rank #388 in the entire dataset up to rank #15 in the condensed set), ehow.com (from #64 up to #2), hgtv.com (from #97 up to #7), and medhelp.org (from #104 up to #16). The difference between popularity of websites found in the condensed vs. the entire dataset and the observation of goal-centric websites surfacing in the condensed dataset leads us to hypothesize that there is a correlation between explicit intentional queries on one hand, and goal-oriented websites and resources on the other.

Results of Analyzing the Condensed Dataset

Beyond comparative analysis, we were interested in the distribution of verbs in our condensed dataset.



Fi

Figure 5. Verb frequency histogram

The histogram in Figure 5 lists the most frequent verbs (in their stemmed word form) in our dataset. The top 10 stemmed verbs in the condensed dataset are make, get, buy, wed, is, find, live, play, use, write. While this list is interesting from a goal-oriented perspective and largely reasonable, it also highlights some of the limitations of our simplified approach, for example “wed” is the result of mistakenly POS-tagging “wedding” as VBG rather than the result of the verb “wed” occurring in the dataset very often (as we were able to confirm by evaluating occurrences of wed vs. wedding in the dataset). Another question we were interested is whether a minority of users is responsible for issuing explicit intentional queries, or whether a larger set of users issues such queries. This would have implications for the broader relevance of different degrees of intentional explicitness in search queries.

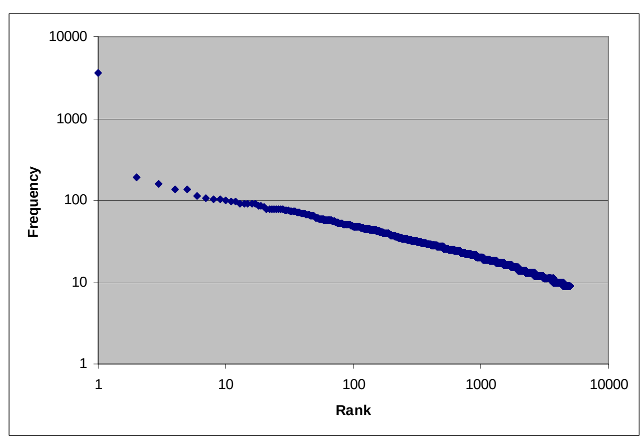


Figure 6. Number of queries per user: rank/frequency plot

In the above figure 6, users are ranked based on their number of queries in the condensed set, whereas only the first 5000 ranks are shown. Frequency corresponds to the number of queries. While the absolute number of explicit intentional queries in the AOL search query log has been estimated to lie between 1.69% and 3.01% [25], the proportion of *users* in our condensed dataset is significantly higher: 14.37% of the users from the entire dataset appear in the condensed dataset as well. As the data points approximately follow a line on a logarithmic scale, the rank frequency distribution appears to represent a power law - a distribution that is often found in systems that contain traces of social activities or interactions.

THREATS TO VALIDITY

In the following, we describe threats to validity according to [28]:

Construct validity: The constructs we intended to investigate in our study are *explicit* and *implicit intentional queries*. Being aware of a broad spectrum of different degrees of explicitness of goals in search queries, we have introduced a simplified distinction for practical purposes. While this distinction enabled us to explore the relevance of different degrees of explicitness, it might be an oversimplification of the underlying phenomenon. However, by defining different degrees of intentional explicitness as a continuous spectrum we hint towards more elaborated future approaches. In addition, relying on part-of-speech tagging and involving expert judgment to distinguish between explicit and implicit intentional queries also puts certain limitations on the generality of our approach. By providing a definition for goals we aimed to objectify our process to a certain extent.

Internal validity: The experts involved in labeling the training set of queries were two of the authors of this paper, which might introduce a potential bias to our results. We tried to mitigate this bias by requiring the experts to reach consensus on the judgment made, and by involving more than one expert. The decision to exclude shorter queries ($n \leq 2$) prohibits us to make statements about a large part of the AOL dataset (~60%). However, our decision was motivated by the inherent difficulty of part-of-speech tagging one or two word English queries correctly, and by the fact that search engine vendors report increasing average query length over the past years[4].

External validity: While we are referring to established theories and definitions on goals from different research areas including human-computer interaction, goal-oriented requirements engineering and search query analysis, our work is biased towards the data available in the AOL search dataset (2006). Investigating other search query logs with respect to different degrees of intentional explicitness is something we are interested in.

Reliability: We have documented and described our experimental classification approach, and built on existing toolkits such as the WEKA toolkit [27], so that reproducing our results is possible within the given limits.

OUTLOOK

In future work, it would be interesting to identify more fine-grained degrees of intentional explicitness and more precise criteria for distinguishing between them. Mining relations between explicit and implicit intentional queries would be another interesting stream of research, as this could allow for search engines to interactively support goal refinement or goal generalization activities. We have identified a number of seemingly suitable web corpora, such as 43things.com, ehow.com, medhelp.org and others, that could be used in related future research efforts. Another promising field of future work seems to be the development of more precise classification approaches. In order to advance in this direction, approaches could, for example, take context or domain knowledge into account to increase the quality of classification (e.g. eliminating movie titles or queries related to song lyrics). Categorization of explicit intentional queries into taxonomies of human goals [6] would be another interesting endeavor that could yield fruitful insights into the goals users pursue on the web. Investigating how our results translate to other contexts, such as the 43things.com website – a website that encourages users to share their goals - is another stream of future research we are interested in.

SUMMARY & CONCLUSIONS

This paper introduced a novel perspective on analyzing search query logs: different degrees of intentional explicitness. We have argued that these degrees represent a *continuous dimension*, and we have shown by example that they are *orthogonal* to existing query categories, such as transactional or informational queries. In an effort to make this novel dimension amenable to analysis, we have introduced two simplified degrees of intentional explicitness, and applied it to the AOL search database. Our analysis demonstrated the principle reasonability of our concepts, and highlighted a series of potentials and challenges when studying different degrees of intentional explicitness in search query logs. Learning about different degrees can be considered essential for leveraging the full analytical potential of "databases of intentions" - and for understanding their limitations. In addition, considering different degrees of intentional explicitness appears critical for search engine vendors to better assess the level of service they can or should provide for different user queries. We have presented a *theoretical* elaboration of different degrees of intentional explicitness and preliminary *empirical* evidence for the principle reasonability of these concepts. More robust techniques to understand a search query's degree of intentional explicitness could have a significant impact on narrowing the cognitive gap between a user's goals and the query she formulates. Finally, our findings could have a broader impact on web search

---

4 http://blogs.zdnet.com/micro-markets/index.php?p=27, last accessed Nov 21, 2007

research, as well as behavioral and social studies of motivation on the web.

ACKNOWLEDGMENTS

We thank Anwar Us Saeed for providing support in implementing parts of the experimental classification approach and Mark Kröll for very helpful comments and criticism. The research of this contribution is funded in part by the Austrian Competence Center program Kplus.

REFERENCES

1. Allan, J., & Raghavan, H., Using part-of-speech patterns to reduce query ambiguity. *Proc. SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, ACM Press (2002), 307--314.
2. Baeza-Yates, R.; Calderón-Benavides, L. & González-Caro, C., The Intention Behind Web Queries, *Proc. SPIRE 2006*, Springer (2006), 98-109.
3. Beitzel, S. M., Jensen, E. C., Frieder, O., Grossman, D., Lewis, D. D., Chowdhury, A., Kolcz, A., Automatic web query classification using labeled and unlabeled training data, *Proc. SIGIR 2005*,. New York, NY, USA, ACM Press (2005), 581-582.
4. Broder, A., A taxonomy of web search, *SIGIR Forum* **36**(2), (2002), 3-10
5. Cavnar, W. B., Trenkle, J. M., N-gram-based text categorization, *Proc. SDAIR 1994*, 161-175.
6. Chulef, A. S.; Read, S. J. & Walsh, D. A., A hierarchical taxonomy of human goals, *Motivation and Emotion* **25** (3), (2001), 191-232.
7. Domingos, P., Pazzani, M. J., On the optimality of the simple bayesian classifier under zero-one loss, *Machine Learning*, vol. 29, no. 2-3 (1997), 103-130
8. Faaborg, A. & Lieberman, H., A goal-oriented web browser, *Proc. CHI 2006*, ACM Press (2006), 751-760.
9. Fürnkranz, J., A study using n-gram features for text categorization, Tech rep., Austrian Institute for Artificial Intelligence (1998).
10. Greene, K., The future of search. http://www.technologyreview.com/Biztech/19050/, last accessed on July 18th, 2007, MIT Technology Review, July 16 (2007).
11. Grobelnik, M., Mladenic, D., Efficient text categorization, *ECML-98 Workshop on Text Mining*, Chemnitz, Germany (1998).
12. Horvitz, E.; Breese, J.; Heckerman, D.; Hovel, D., Rommelse, K., The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users, *Proc. UAI 1998*, (1998), 256-265.
13. Jansen, B. & Spink, A., How are we searching the World Wide Web? A comparison of nine search engine transaction logs, *Information Processing and Management* **42**(1), (2006), 248-263.
14. Jurafsky, D., Martin, J. H., Speech and Language Processing: An introduction to natural language processing, Computational Linguistics and Speech Recognition (International Edition), Prentice Hall (2000).
15. Kirsh, D., When is information explicitly represented?, UBC Press (1990), 340-365.
16. Lee, U., Liu U., & Cho J., Automatic identification of user goals in Web search. *Proc. WWW '05*, New York, NY, USA, ACM Press (2005), 391—400.
17. Liu, H.; Lieberman, H. & Selker, T., GOOSE: A goal-oriented search engine with wommonsense, *Proc. AH 2002*, Springer-Verlag, London, UK (2002), 253-263.
18. Loper, E.. Bird, S., NLTK: The Natural Language Toolkit, (2002).
19. Norman, D., The design of everyday things, (1988).
20. Pass, G., Chowdhury, A., Torgeson, C., A picture of search, *Proc. InfoScale 2006, Hong Kong*, ACM Press (2006).
21. Regev, G. & Wegmann, A., Where do goals come from: the underlying principles of goal-oriented requirements engineering, *Proc. RE 2005*, Washington, DC, USA, IEEE Computer Society (2005), 253-362.
22. Rose, D., Levinson, D., Understanding user goals in web search, *Proc. WWW 2004*, New York, USA (2004).
23. Ryu, H. & Monk, A., Analysing interaction problems with cyclic interaction theory: Low-level interaction walkthrough, *PsychNology Journal* **2**(3), (2004), 304-330.
24. Sebastiani, F., Machine learning in automated text categorization, *ACM Computing Surveys* , vol. 34, no. 1, (2002), 1-47.
25. Strohmaier, M.; Lux, M.; Granitzer, M.; Scheir, P.; Liaskos, S. & Yu, E., How do users express goals on the web? - An exploration of intentional structures in web search, *in* 'We Know'07 International Workshop on Collaborative Knowledge Management for Web Information Systems, in conjunction with WISE'07, Nancy, France, (2007).
26. Strzalkowski, T. & Carballo, J., Natural language information retrieval: TREC-5 report, *in* Text REtrieval Conference, (1998), 164-173.
27. Witten, I. H., Frank, E., Data mining: practical machine learning tools and techniques, *Morgan Kaufmann Series in Data Management Systems*, 2nd edn. Morgan Kaufmann, (2005).
28. Yin, R. K., Case study research: design and methods (Applied Social Research Methods), SAGE Publications, (2002).
29. Yu, E., Modelling strategic relationships for process reengineering, PhD thesis, Department of Computer Science, University of Toronto, (1995).