An Empirical Study of Natural Language Parsing of Privacy Policy Rules Using the SPARCLE Policy Workbench

Carolyn A. Brodie
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
914-784-7237
brodiec@us.ibm.com

Clare-Marie Karat
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
914-784-7612
ckarat@us.ibm.com

John Karat
IBM T. J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
914-784-7832
jkarat@us.ibm.com

ABSTRACT
Today organizations do not have good ways of linking their written privacy policies with the implementation of those policies. To assist organizations in addressing this issue, our human-centered research has focused on understanding organizational privacy management needs, and, based on those needs, creating a usable and effective policy workbench called SPARCLE. SPARCLE will enable organizational users to enter policies in natural language, parse the policies to identify policy elements and then generate a machine readable (XML) version of the policy. In the future, SPARCLE will then enable mapping of policies to the organization's configuration and provide audit and compliance tools to ensure that the policy implementation operates as intended. In this paper, we present the strategies employed in the design and implementation of the natural language parsing capabilities that are part of the functional version of the SPARCLE authoring utility. We have created a set of grammars which execute on a shallow parser that are designed to identify the rule elements in privacy policy rules. We present empirical usability evaluation data from target organizational users of the SPARCLE system and highlight the parsing accuracy of the system with the organizations' privacy policies. The successful implementation of the parsing capabilities is an important step towards our goal of providing a usable and effective method for organizations to link the natural language version of privacy policies to their implementation, and subsequent verification through compliance auditing of the enforcement logs.

Categories and Subject Descriptors
H5.2. Information interfaces and presentation: User Interfaces. K4.1. Public policy issues: Privacy

General Terms
Management, Design, Security, Human Factors

Keywords
Policy, privacy, security, usability, social and legal issues, design.

1. INTRODUCTION
Today organizations are under increasing pressure to ensure that the personal information from their customers, patients, citizens and employees that the organization collects, uses, and stores is protected from both internal and external threats. Both new legislation and social pressures caused by the ever growing number of reports of phishing attacks, identity theft, and other online crime are increasing the pressure on organizations to protect personal information against these threats. Organizations that expose data bear the additional expenses associated with notifying individuals whose data may have been exposed and helping these individuals to limit their risks using techniques ranging from changing account numbers and reissuing credentials to paying for them to enroll in credit watch services to protect against identity theft. The organizations must face these expenses whether the exposure was accidental or the result of a malicious attack. In order to protect against these threats, organizations must put in place well-understood and comprehensive sets of security and privacy policies, educate their staffs on these policies, enforce them, and then audit their enforcement to ensure compliance. These processes are currently difficult for organizations to implement successfully. To further complicate the situation, much of the existing security and privacy technology is designed for use by experts and is difficult for either end users or organizational users who are not security experts to use correctly. Further, using these mechanisms incorrectly can be worse than not using them at all. Whitten and Tygar highlighted this issue while studying the use of email encryption technology when they pointed out that "security mechanisms are only effective when used correctly" and these mechanisms are often not used correctly due to usability issues [23]. The Computing Research Association (CRA) Conference on Grand Research Challenges in Information Security and Assurance echoed this concern when they identified the ability to "give end-users security controls they can understand and privacy they can control for the dynamic, pervasive computing environments of the future" as a major research challenge [8].

One particular challenge for organizations is ensuring that the policies that they create are implemented and enforced correctly. Privacy policies are not new to organizations, however very little has been done to implement them through technology [22]. To help address this issue, the authors of this paper reported on empirical results of HCI testing with an early, Wizard-of-Oz (mid-level fidelity) prototype of the SPARCLE Policy Workbench at SOUPS 2005 [7]. The usability study covered target user evaluations of the scenario-based prototype of a policy workbench that allowed privacy policy authors to write policies in natural language. The prototype then simulated the parsing of policies to show how policy rule elements would be identified by a parser, and then reviewed and modified by the user through the visualization provided.

In this paper, we describe our subsequent research which builds on the research presented from last year with the creation of a fully functional version of the SPARCLE Policy Workbench policy authoring capabilities. The 2005 SPARCLE Policy Workbench prototype differs from the 2004 prototype in that it is fully integrated with a functional shallow parser which identifies the privacy policy elements in each rule entered as natural language. These elements are then used to create policy visualizations and the XACML [19, 20] version of the policy. We describe our research into the use of parsing technology that makes the use of natural language possible. In this paper we provide an overview of the SPARCLE Policy Workbench and provide a detailed description of the grammars we developed that run on a shallow parser to identify rule elements. Using these grammars we have achieved an average of 88% to 94% parsing precision on organizational privacy policies. Next we describe the empirical results of organizational users' evaluation of the SPARCLE workbench and its parsing accuracy. Finally we discuss our ongoing research challenges into effectively recognizing and labeling the policy rule elements in the natural language to reduce the necessary pre-processing and to generalize the applicability of the SPARCLE Policy Workbench to other policy domains.

2. Related Work

Given the growing awareness by society of identity theft and other misuse of personal information, it is not surprising that privacy remains a very active area of research for current and emerging technology design [1]. In order to address these concerns, standards for the definition of machine readable privacy policies have been developed and analyzed. P3P [9] is one of the first privacy policy languages that has been standardized by an international standards body, the W3C. P3P is an XML based language that allows organizations with Websites to create machine readable versions of their privacy policies. Generally, P3P allows organizations to specify rules that contain the definition of the data to be collected, how the data will be used, the allowed roles of users of the data, the purpose of the use, and how long the data will be retained. The use of P3P by Websites has benefits for end users in that it has facilitated the creation of tools and standards that can help them determine whether or not the site's policy is acceptable [10, 17, 24].

While the ability to quickly understand a site's privacy policy and determine if the site conforms to their preferences is helpful to end-users, it is important to understand that P3P offers no guarantee that the policy is actually implemented as specified within the organization. This fact has lead to research into how machine readable (XML schema languages) privacy policies can be used by organizations to enforce policies [15] and to the development of schemas, such as XACML with a privacy profile and EPAL, to represent the internal policies that must be enforced [5, 19, 20]. Along with the development of machine readable schemas to represent internal policies, there has also been a great deal of research into how to enforce the policies [2, 3, 4, 6, 12].

Given that there is a large body of research showing that society is concerned with protecting personal information and the fact that privacy policies are not new [22], it is troubling that organizations are not doing more to implement their privacy policies through technology. Recent studies indicate that many organizations have started to recognize that privacy is an issue for them. However, they currently do not know how to use technology to help them enforce their privacy policies. The Ponemon study [21] reported that although 98% of the companies in their survey have a privacy policy, 52% believe they do not have the resources to adequately protect privacy. We suggest that one of the reasons for this situation is the difficulty that organizations face in implementing their privacy policies and the lack of usable tools that are tailored to the skills of the intended users. The SPARCLE Policy Workbench is intended to assist organizations in this regard by providing a usable tool to create and manage their privacy policies through the use of natural language. People in policy roles are experts in policy and legislation and often do not have a technical background. They are most comfortable expressing policy in natural language. Natural language is also a very usable method for communicating the content of policies to employees of an organization and people who interact with an organization. The SPARCLE system enables the expression of policy in natural language and provides the means for transforming the natural language as needed for automated enforcement within organizations.

3. Overview of the SPARCLE Policy Workbench

To help organizations implement their privacy policies successfully, the SPARCLE Policy Workbench is designed to provide them with tools to help them create understandable policies, link their written privacy policies with the implementation of the policy across their IT configurations, and then help them to monitor the enforcement of the policy through internal compliance audits. While we plan to extend SPARCLE to support multiple policy domains such as security access control, the current prototype has been tailored for privacy policy management and for allowing organizations to author and understand their policies and then create machine readable versions of the policies.

Research has shown that organizational privacy policies are authored by individuals with a range of skills. Some policy authors have a legal and/or business background while others are more technical [14]. In order to support users with a variety of skills, SPARCLE has been designed and evaluated with two methods for authoring policies [13]. The members of organizations who are responsible for creating policies and have a legal and/or business background may be more comfortable working with natural language. Therefore, SPARCLE has been

designed so that policy authors can write policy rules in natural language using a rule guide or can import existing text policies and tailor them using the rule guide. SPARCLE then transforms the natural language into a structured format. Alternatively, policy authors can use a structured format to define the elements and rule relationships that will be directly used in the machine readable policy. SPARCLE will generate natural language for rules created using this method. SPARCLE users can use either method exclusively or move between the two methods and the tool will keep the two formats synchronized. Once the policy is in the structured format, SPARCLE provides visualizations of the policy to assist the policy creators in ensuring that the policy coverage is what was intended. Finally, when the policy author is satisfied with the policy, SPARCLE will generate an XACML version of the policy for use by an enforcement engine.

3.1 Policy Selection and Natural Language Policy Authoring

When a user logs onto the SPARCLE Policy Workbench, he is given the option of selecting an existing policy to modify, uploading an existing text file, or creating a new policy file. A new policy may be blank or can be based on a template or existing policy.

Once the policy is selected the author will be taken to the natural language policy authoring page, shown in Figure 1. The area at the top of the page shows the policy name, description, domain, and the date it was first created and last modified. The rule guide that is shown above the policy text editing area in Figure 1 has two purposes. First, it reminds authors of what elements are needed in an implementable policy rule. We define an implementable privacy rule as a rule that can be defined for automated enforcement through technology. Second, the guide defines the order in which elements in a policy rule must be placed so that the natural language parsing technology can identify them with as high a degree of accuracy as possible. The author can edit text in the policy text editing area or can cut and paste text from other files so long as the rules conform to the rule guide. However, the author does not have to use the exact wordings shown in the guide. For example, the rule "Customer Service Reps can collect and use customer name **to confirm identity**" is as acceptable as "Customer Service Reps can collect and use customer name **for the purpose of confirming identity**." Once the author has finished authoring the policy, he clicks on the "Save and Continue" button to invoke the parser. When parsing has completed, the user proceeds to the Structured Policy Authoring Method page, shown in Figure 2, to see the structured format of the policy.

3.2 Structured Policy Authoring Method

Figure 2 shows the results of parsing each policy rule. When SPARCLE parses each policy rule, it saves the policy elements (i.e. user categories, actions, data categories, purposes, and conditions and obligations) found in that rule. The elements are reconstructed into sentences and shown next to radio buttons in a list with the first rule chosen by default as shown in Figure 2. While the accuracy of the parser is very high, it is not perfect so it is important for the user to compare the text of the parsed rules with the original rule text. Experience with actual organizational policies has also shown that human errors in authoring the rules are also found during this process. For example, it is very common to find rules that have no explicit purpose. For these reasons, the original text of the selected rule is shown in a field directly above the parsed rule list so that users may compare them. An example of SPARCLE identifying a missing purpose is shown in Rule 3 in Figures 1 and 2. Figure 1 shows the text entered by the policy author, "Management can report customer transactions if required by law." This rule does not contain an explicit purpose. In Figure 2, we see that SPARCLE has added the purpose "None Selected" to highlight the fact that the purpose is missing to the policy author.

All of the rule elements in the selected rule are also shown in rule element lists that appear below the reconstructed rules and control buttons on the page. Rule element lists are initially populated with domain defaults (e.g., typical terms for healthcare policies), and are extended as elements are found in new natural language rules during parsing or are explicitly added to the lists by the author. If the user wishes to change a rule, she can select the rule by clicking on the radio button next to it, then select the elements she would like from the policy element lists, and finally click on the "Modify Rule" button. She can also add a single rule or create all the rules in the policy using this method. For each rule she would like to create, the author clicks on the "Create Rule" button. A new blank rule will be added to the list and the "Create Rule" button will become the "Save Rule" button. The author can select the elements she would like in the new rule using the rule element lists. When all the elements are selected, she simply clicks on the "Save Rule" button to save the new rule. If the user returns to the authoring page (shown in Figure 1), all the changes that she has made on the structured policy authoring method page will be reflected in the text. When the policy author is satisfied with the policy they can generate an XACML version of it by clicking on the "Save as XACML" button at the bottom of the page (not shown in the screen capture).

3.3 Understanding the Policy

During our feedback sessions, privacy policy professionals in organizations have told us that one challenge facing them is that they often find it difficult to ensure that the policies that they create are complete, able to be implemented, and consistent. Figure 3 shows a visualization created to assist users in understanding their policies. The visualization consists of a 2-dimensional table in which the user can choose one policy element type with which to label the columns and a second policy element type to label the rows. In the example shown in Figure 3, "user categories" have been chosen as the column labels and "data categories" have been chosen as row labels.

Each cell in the table contains the rest of the rule elements for each rule that has a given user category and data category on its row and column headings. For example, in the cell that is in the column with the heading "Financial Analysts" and the row with the heading "customer accounts", the other rule elements "Can use for the purpose of make loan decisions" appears. This means that users that are classified as "financial analysts" can "use" "customer accounts" for the purpose of "make loan decisions". There are no conditions or obligations associated with this rule. SPARCLE also highlights who cannot access data. SPARCLE assumes that if no rule is included that explicitly allows access, access is denied. Therefore, the lack of a rule in a cell indicates that individuals in the specified user group cannot access the specified data for any purpose.

Policy authors can look at the data in multiple ways. The user can choose to change the column or row headings at any point in time using pull-down lists as shown in Figure 3. For example, a policy author might choose to view what user categories can access data for various purposes, by changing the "Row" pull-down field to "Purpose". If the policy author decides to update a particular rule, all he needs to do is to click on the rule (or the Access Denied link) in the cell and he will be taken back to the Structured Policy Authoring page (shown in Figure 2) with the rule selected and ready to be changed. Therefore, the author can move back and forth between the pages making updates to the policy and viewing the policy coverage.

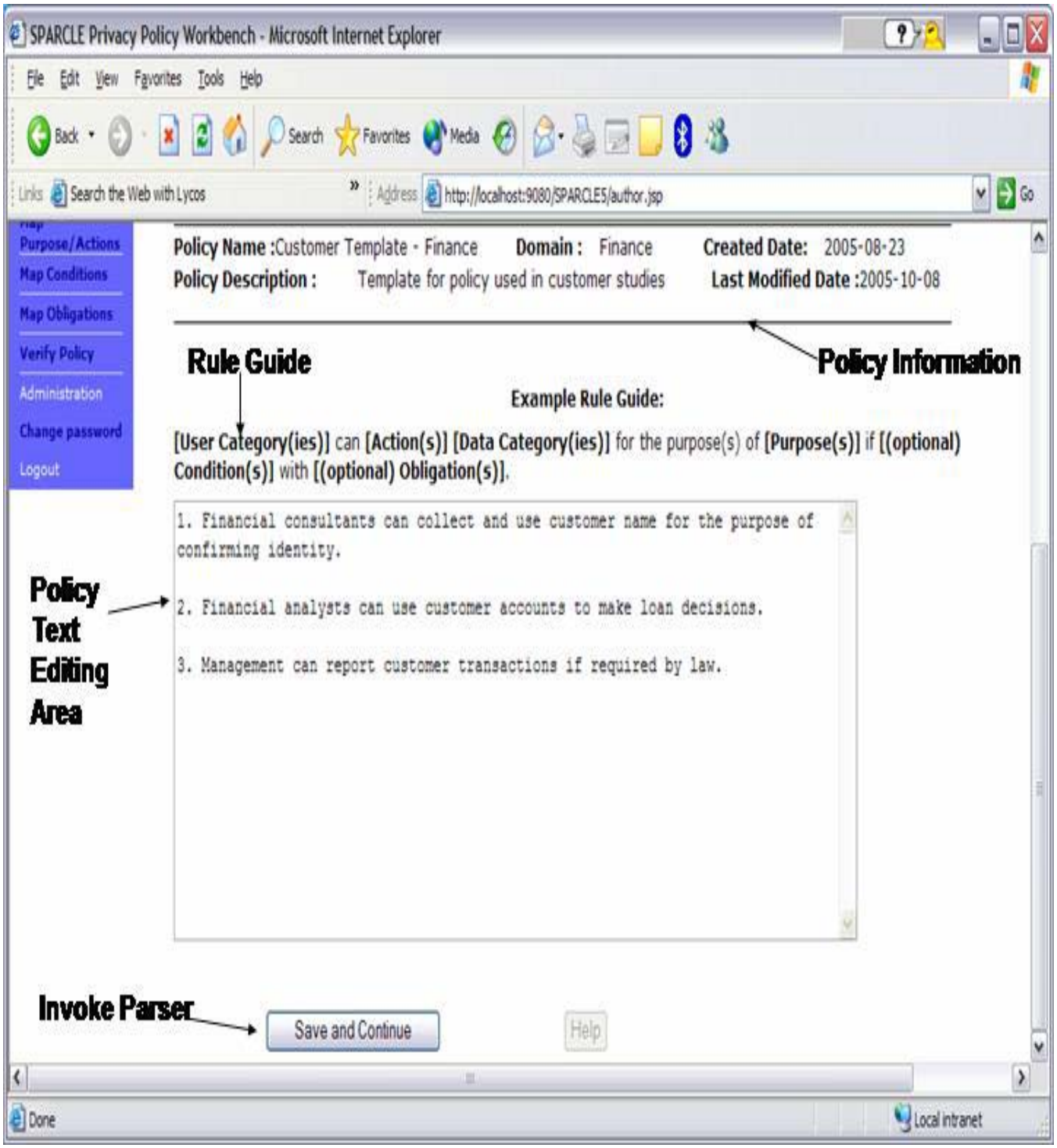

Figure 1. SPARCLE Policy Workbench natural language authoring page

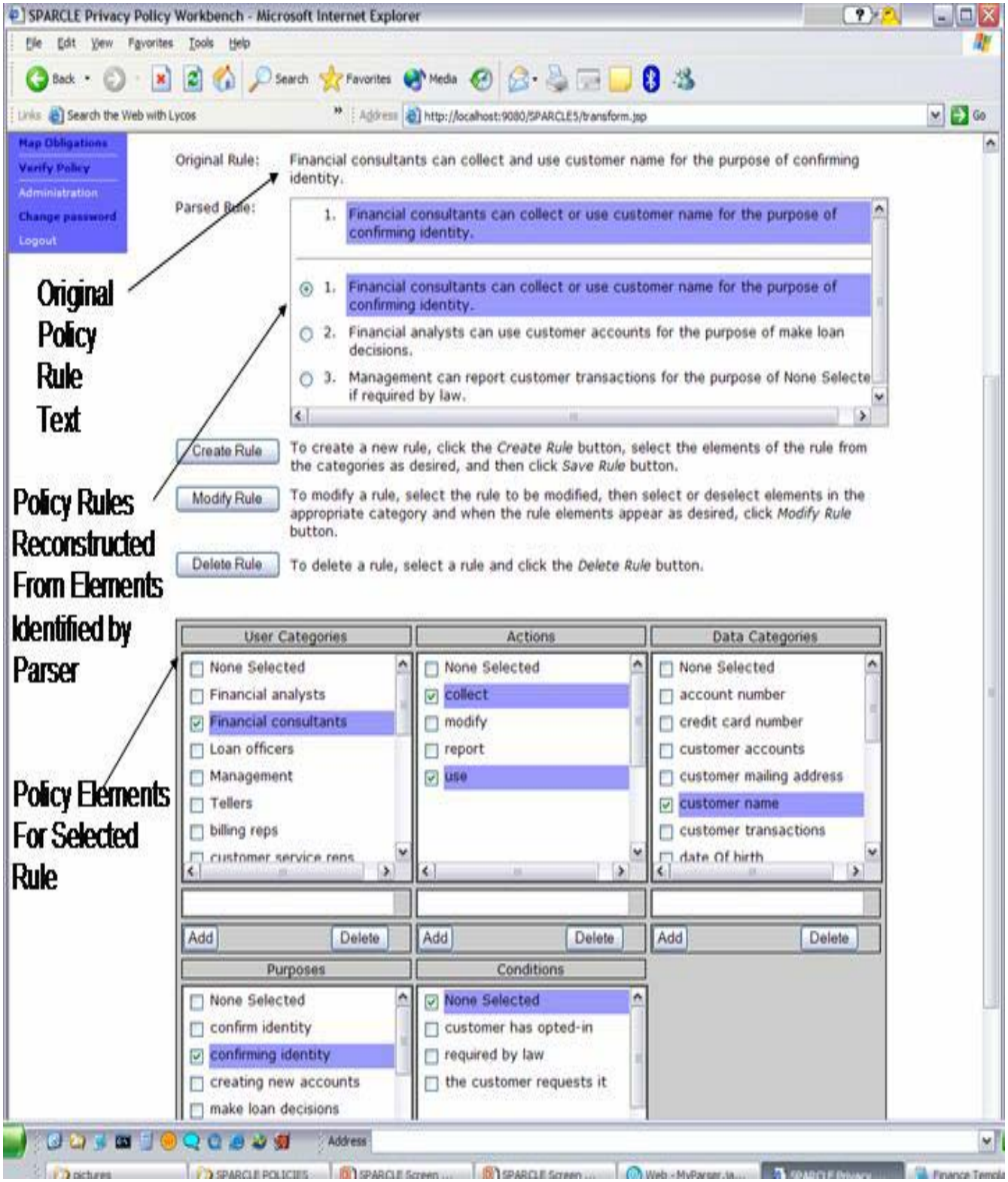

Figure 2. SPARCLE Policy Workbench structured authoring page

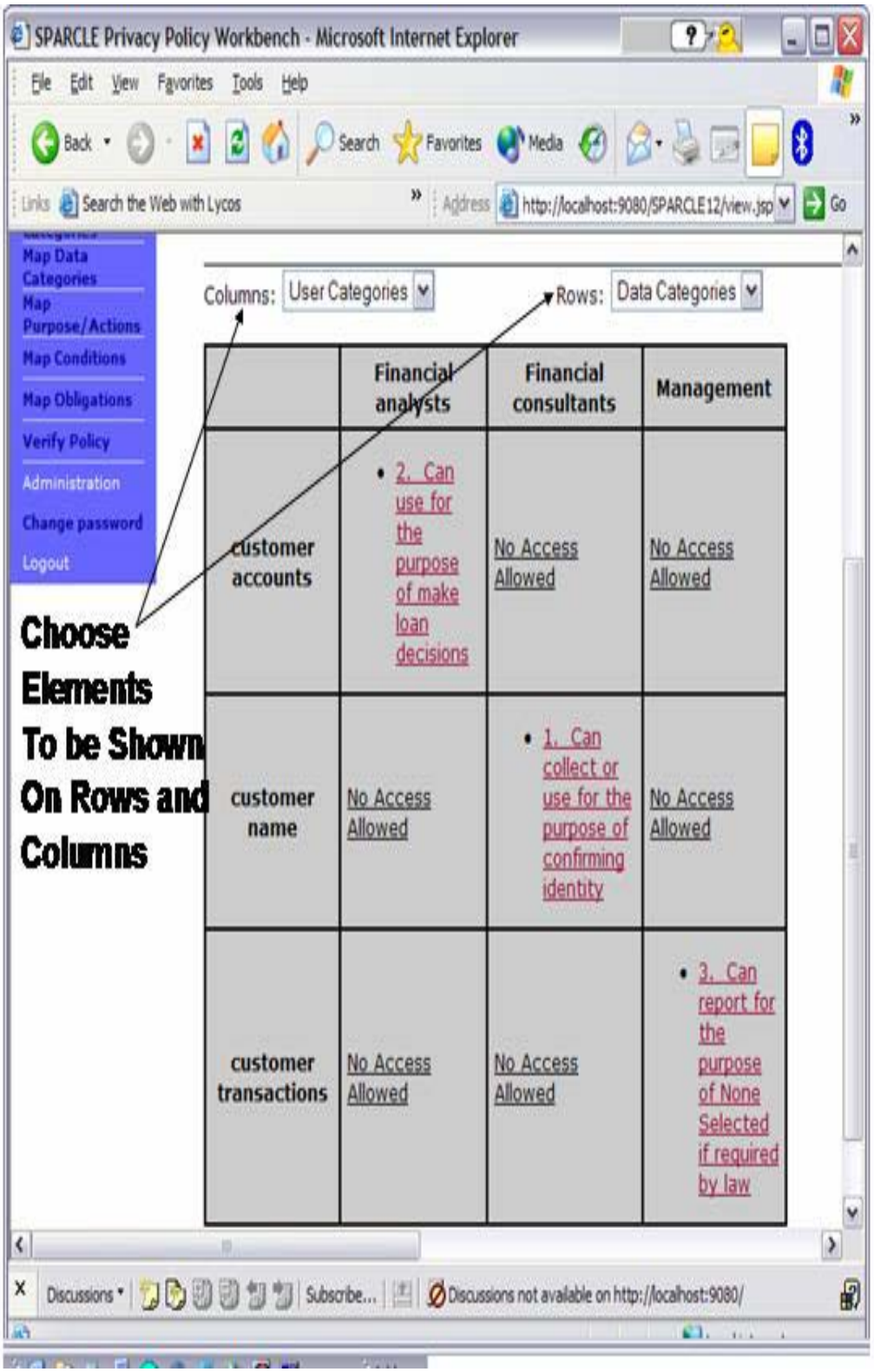

Figure 3. SPARCLE Policy Workbench policy visualization page

4. Natural Language Parsing in SPARCLE

In order to identify the elements in a policy rule, SPARCLE uses a shallow parser [18] that is based on the IBM Unstructured Information Management Architecture [11]. A shallow parser processes text in a number of stages, beginning with operations that use limited linguistic knowledge to identify syntactic structures such as nouns, noun phrases, verbs, verb groups, and modifying phrases. From the basic part of speech information, the shallow parser then uses one or more grammars to identify the desired text in a document based on patterns of parts of speech. For SPARCLE, we created a set of grammars that are designed to identify the five policy elements in each rule: user categories, actions, data categories, purposes, and conditions/obligations. User categories are roles that organizational users are assigned as part of their jobs. Data categories are the pieces of information that a user with a specific role can act on. Actions are defined as the things that a user can do with the data specified in the rule. Good examples of SPARCLE actions include those defined by the P3P specification [9] such as read, use or modify, although SPARCLE does not limit users to this list. Purposes are the allowed reasons a user can take an action on a particular kind of data. Conditions and obligations are treated as one policy element type because our analysis of privacy policies has shown that policy authors tend not to differentiate between conditions that must be true before a rule can fire and obligations that must be fulfilled after data is accessed. Together these elements make a privacy rule as shown in Figures 4 and 5.

4.1 Grammar Design

Analyzing natural language to understand policy rules is a difficult task [16]. In order for SPARCLE to be a useful and usable tool, we needed to create a set of grammars that would identify policy elements in each rule with high reliability. We started by analysing privacy policy rules that we collected (1) from organizations in the health care, banking/finance, and government domains and (2) during an empirical evaluation of the two authoring methods employed by SPARCLE [13]. The policies from organizations were provided to us for research purposes by the participants of the organizations we invited to participate in our research. The organizational policies varied in scope, length, and level of specificity. The policy rules from the empirical evaluation were written by 36 participants in a study comparing the two authoring methods used in SPARCLE. During the study the participants were given scenarios that described the privacy needs of organizations in the domains of health care, government, and banking/finance. Participants were asked to create policy rules for three scenarios using three different policy authoring methods: a control method, natural language with the guide that is shown in Figure 1, and a structured selection list method that was an earlier version of the structured authoring method shown in Figure 2 (see above). We analysed all of the policies to understand the linguistic structure of the each type of policy element in the rules. Using this analysis, the grammars were created.

In order to achieve a higher degree of parsing accuracy we decided to use constrained natural language rather than attempt to parse completely unconstrained natural language. Two basic constraints were placed on the policy rules. First, a policy rule must be written as a single sentence. This allows the parser to easily identify the scope of each rule. Our research suggests that limiting rules to a single sentence is a reasonable restriction because rules are often written as a single sentence in the organizational policies we have analyzed. Second, in order to classify policy rule elements correctly, the elements must appear in each rule in one of two orders as shown in Figures 4 and 5.

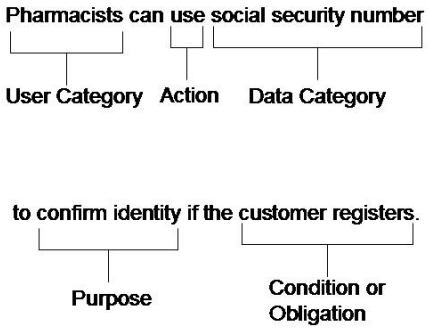

Figure 4. The first possible SPARCLE privacy rule structure.

In the first rule structure, shown in Figure 4, the policy rule element order is as follows: one or more user categories, one or more actions, one or more data categories, one or more purposes, and, optionally, one or more conditions or obligations. In the second possible rule structure, shown in Figure 5, the policy rule element order is modified so that a rule is made up of one or more data categories, one or more actions, one or more user categories, one or more purposes, and, optionally, one or more conditions or obligations. Our analysis of both the organizational policies we collected and the rules collected in the empirical evaluation indicated that these two formats were commonly used by rule authors. Thus we believed that restricting users to writing rules using these common formats would be easily learned and followed.

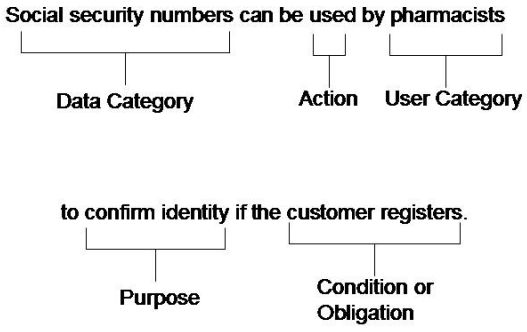

Figure 5. The second possible SPARCLE privacy rule structure.

SPARCLE uses a set of grammars that operate in a cascade [18] to add meta-data tags to the rules indicating where privacy

elements start and stop. In the cascade the grammars execute in a particular order where the meta-tags inserted by earlier grammars can be referenced by later grammars. We designed our grammars to operate by first trying to identify the most complex elements in each rule and tagging those so that the subsequent grammars can operate on the rest of the rule as they proceed to identify the more straight forward linguistic parts of the rule. For this reason conditions and obligations are identified first, followed by purposes, user categories, data categories, and finally actions. Conditions and obligations tend to be the most complex part of the rule because they can take many linguistic forms and because they are optional parts of privacy rules. While some conditions and obligation phrases have the fairly simple form of a noun followed by a verb as shown in Figures 4 and 5, others are much more linguistically complex, such as the following example, "with the condition that they will be deleted from our records in 2 years upon the closing of the account". Along with conditions and obligations, we found that purposes can also take many linguistic forms. The simplest purposes are of the form "verb noun" as shown in Figures 4 and 5, but they can be much more varied. For example, one purpose that we found in a privacy policy rule consisted of the phrase, "for the purpose of entering into transactions with the supplier, performing transactions with the supplier, and administering the relationship." Therefore, purposes are the second element tagged in our grammar cascades. User categories and data categories tend to be simple noun phrases such as "pharmacists" and "Social Security Numbers", although they too can have somewhat more complex linguistic structure such as the phrase, "database of customer names". However these have not been found to be as varied as conditions and obligations, and purposes. Finally, we have found that actions are the most straight forward policy element with terms such as "use", "collect", and "delete".

At various points in the grammar design and development process, we calculated the accuracy of the grammars using a test set of 7 privacy policies with a total of 46 rules that were selected from the policies created during the empirical study [13]. An example of one of these policies is shown in Figure 6.

1. Customer Service Reps and pharmacists will collect name, permanent address, contact information, demographic information, and personal information for the purpose of providing customer service.
2. Customer Service Reps will use SSN for the purpose of verifying identity.
3. Pharmacists will use current medication information for the purpose of checking for drug interactions.
4. The research department will use age breakdowns and gender breakdowns for the purpose of generating reports about drug use by demographic group given that the reports will be deleted in five years.
5. Marketing will provide customer name and customer address to our partner organizations for the purpose of sending special offers if the customer indicates a willingness to receive special offers.
6. The research department will use customer information for the purpose of creating reports with the condition that the reports do not contain individually identifying information.

Figure 6. Sample privacy policy used in testing the parsing grammars.

For each policy we calculated liberal and conservative values for parsing precision and recall. Parsing precision is defined in the information retrieval literature as the number of correct words or phrases retrieved divided by the total number of retrieved items. Recall is defined as the number of correct words or phrases retrieved divided by the number of correct words or phrases that exist in the text. Recall differs from precision in that if a word phrase is incorrectly identified as an item, it is included in the precision calculation but not in the recall calculation. These measures were chosen as a measure of the quality of the parsing because they are commonly used as measures by the information retrieval community. However, we found that they did not quite meet our needs in that they did not tell us how to handle retrieved items that were partially correct. In other words, we had cases where two words out of a three word phrase were returned or one or more extra words were retrieved with the correct phase. Therefore, we decided to do two sets of precision and recall calculations, one liberal and one conservative. In the liberal calculation we considered partial phrases that were identified to be correct if they were complete enough to be understood. In the conservative calculation we required an exact match on a phrase to classify it as correct. In other words, in the liberal case if one word in a three-word phrase is missing or an extra word is included, it was counted as being found and in the conservative case it was counted as missed. For example, one word in the action element was missed in the following rule, "Customer service reps can externally share customer addresses with third party affiliates for the purpose of fulfilling customer requests". In this rule SPARCLE identified the action as "share" rather than "externally share", and therefore this element was counted as partially correct. In another example, an extra few words were labeled as part of the data element in the rule "Bank managers can give customer information to third parties for the purpose of marketing if the customer agrees to it. " In this case the phrase "customer information to third parties" was labeled as the data element. The data elements should have consisted of only the words "customer information". The other three words are actually defining the recipient of the data, a policy element that SPARCLE currently does not identify, but will in the future. In both of these cases, we believe that the tagged phrase is understandable to the user, but is not completely correct. These cases would be considered missed elements in our conservative calculations and found elements in the liberal calculations. On average across these 7 policies, SPARCLE achieves a conservative precision of 86%, a conservative recall of 88%, a liberal precision of 95% and a liberal recall of 97%. Once we felt that SPARCLE was functioning well with this test data, we wanted to evaluate SPARCLE using actual organizational policies.

5. 2005 SPARCLE Policy Workbench Usability Evaluation

Twenty-six participants in North America were recruited for the usability evaluation of SPARCLE. Potential participant organizations were identified through their demonstrated interest at professional conferences on privacy and also through peer referrals in the international privacy community. The participants who volunteered to be part of the usability evaluation represented large, well known international health care and banking/finance organizations and US government organizations. Within their

respective organizations the recruited participants had responsibility for the creation, implementation, and/or auditing of privacy policies. The backgrounds of the participants included specialization in law, public and organizational policy, operational business management, auditing, compliance, and human resources. Participants were promised confidentiality regarding their feedback. They received no payment other than a copy of the SPARCLE Policy Workbench evaluation report in thanks for their participation in the evaluations. The evaluation report included a description and analysis of the aggregated de-identified data across all evaluation sessions.

The evaluation sessions were conducted on site at each of the participants' organizations. The evaluation sessions had two parts. First, 90-minute scenario-based usability walkthroughs of the SPARCLE prototype were conducted with small groups who discussed and provided ratings of the tool's functionality, its precision and performance. A user scenario that illustrated core user tasks and exercised key functionality was used to guide the walkthrough. A version of the user scenario was tailored to each of the three domains (healthcare, banking/finance, and government). In preparation for each evaluation session, the SPARCLE team asked for and received a portion of the organization's privacy policies which we prepared and analyzed using the SPARCLE Policy Workbench before the sessions (more detail on the preparation process is provided below). Then the policies were loaded into SPARCLE and parsed using the workbench. The results of the policy analysis were presented and discussed with the participants during the sessions. During the course of the 90 minute sessions, we gathered verbal and written feedback on the usability, design, acceptability of the parsing and value of the privacy tool.

In the second part of the evaluation session, participants had the opportunity to gain some hands-on experience with the fully-functioning prototype. The participants were invited to work with SPARCLE one at a time and could explore the functionality in authoring privacy policy rules using one or both methods in the tool. They could parse their newly created rules and see the accuracy of the parser. They provided verbal feedback on their hands-on experiences. All user evaluation data was analyzed. Quantitative data were analyzed for statistical differences across the new 2005 evaluation data and the 2004 data reported previously [7] and no statistical differences were found. Content analysis of the qualitative data identified summary themes and relationships between the data elements and guides ongoing design activity.

Statistical analysis of the quantitative data showed that the functional version of the SPARCLE Policy Workbench was evaluated very highly by the participants. During the walkthroughs, participants (26 participants in 5 sessions) rated the prototype very positively. The average rating for the 21 features presented was 5.73 on a 7-point scale with 7 indicating that the feature has "Highest Value" to the organization and 1 indicating "No Value". These features included policy authoring features (e.g. the ability to author rules in natural language), as well as policy visualization features (e.g. the ability to see policy coverage in a two-dimensional table). Figure 7 shows the participant ratings of different features collected during the 2005 SPARCLE evaluations. One of the features, "Enter rules with NL and Guide" appears twice in the results. This is because we collected ratings of this feature first by itself and then in conjunction with the use of structured method for creating rules. The structured method is called "Enter rules with lists" in Figure 7. The results show very strong and positive support for the 8 features of the SPARCLE policy workbench for authoring and transformation.

Given the anchors on the scale run from "No Value" to "Highest Value", the fact that the majority of items are rated as very high value is compelling. It is notable that the ratings for the key design features range between 5.42 and 6.15 indicating high and very high value to intended users of the system.

For the purposes of this paper we are specifically interested in the parsing accuracy. Along with the results of the SPARCLE parsing of each organization's policies, we discussed SPARCLE's parsing accuracy and the amount of time necessary to prepare a policy to be parsed. We wanted to know if participants felt the preparation time (an average of about 90 minutes per policy) and the accuracy were acceptable to them. Therefore, we asked participants specifically about these characteristics. Encouragingly, participants rated the parsing accuracy favorably, rating it 5.54 on the same 7 point scale. They were less favorable about the preparation time, rating this at a moderately favorable level (4.63). Participants generally expressed a desire to have "no additional work" needed to input policy rules. This user requirement is a valuable goal. Given the capability of present and emerging technology, it is likely that some human pre-processing component will be required in the short term. User reaction to this pre-processing may improve in practice as they experience the benefits of having policy rules in a standard readable format. We will continue to track user reaction to rule preparation in future work and will explore every possibility for further automating this step in the process.

5.1 Parsing Results using SPARCLE

In addition to the participants' perceptions regarding the parsing accuracy, it is important to look at the empirical parsing results across the organizational policies. Therefore, we calculated the liberal and conservative parsing values for precision and recall for the organizational policies we collected. As stated previously, each organization that participated in the walkthroughs provided the research team either a portion or all of an external privacy policy or an internal privacy policy. Each of these policies contained a range of 7-18 rules. In order to parse the policies using SPARCLE, the text required some initial pre-processing. This task was performed by team members who had not participated in the grammar writing in order to gauge the difficulty of the task for individuals not familiar with the grammars. This activity was necessary because the content and format of organizational privacy policies varies widely and generally includes non-rule text. The pre-processing included a two step process. First the explanatory text and heading format information were removed from the text. Then rules were rewritten using the restricted natural language guide shown in Figure 1. Using this guide we ensured that a rule was represented as a single sentence which indicated who was to use the data, what actions were to be taken, the name of the data element, the purpose of the use of the data and any conditions that further constrained the use of the data. In general this re-writing just consisted of re-structuring existing sentences so that elements were in the order easiest for the parser to handle.

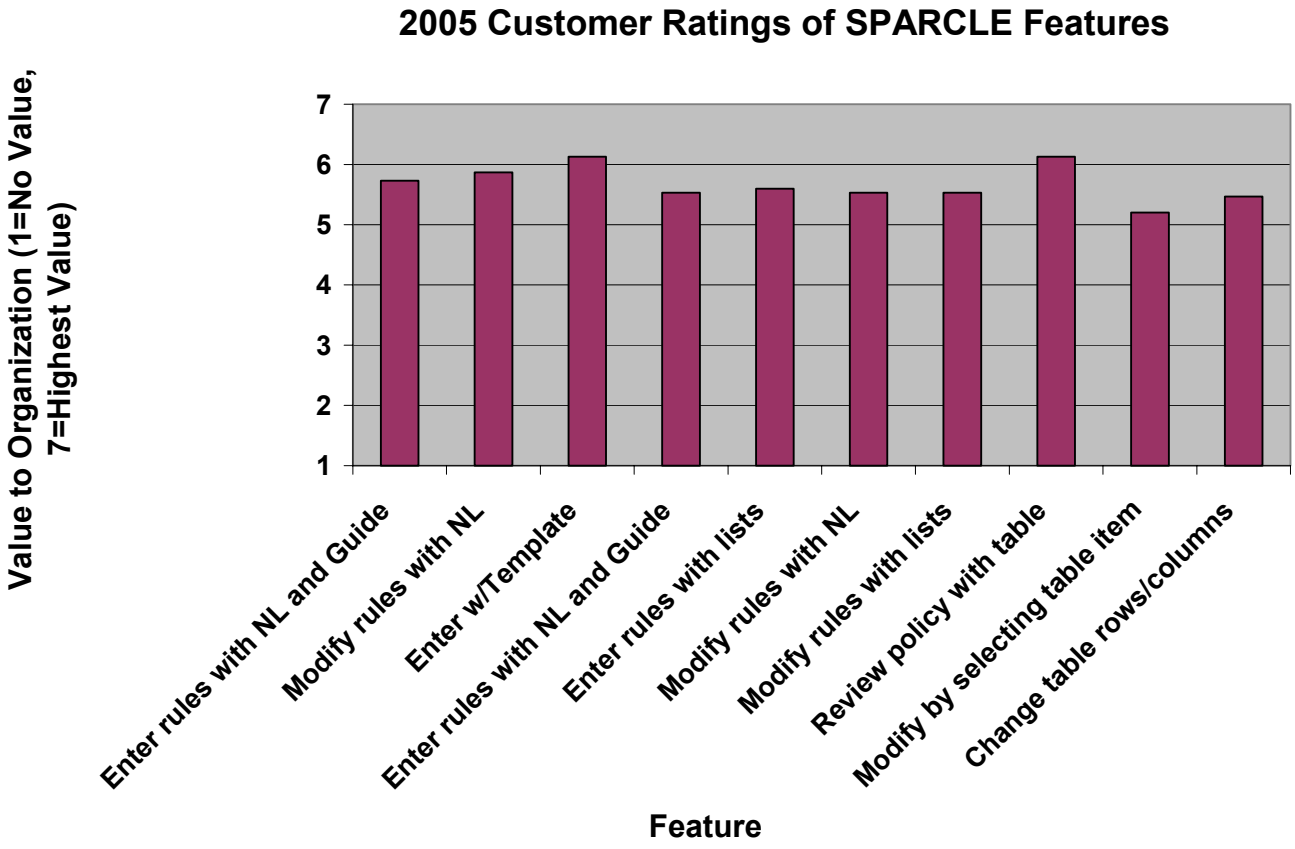

Figure 7. 2005 participant ratings of key features of the SPARCLE Policy Workbench

The organizational policy text was then uploaded into SPARCLE and parsed. For each policy that was parsed, we calculated a liberal and conservative precision and recall value. The results obtained from parsing the customer policies were quite promising, as shown in Table 1. In addition to the policies collected in preparation for the evaluations, we also analyzed the privacy policy from a high tech company. The results are included in the table. The parsing precision ranged from 82-100%, with an average liberal precision of about 94% and a liberal recall of 99%.

6. Future SPARCLE Parsing Research

The initial SPARCLE Policy Workbench functional prototype was very well received by the participants of the design walkthroughs, due in part to the high accuracy of the parsing. However, it is also important to understand how we can increase the parsing accuracy and reduce the amount of pre-processing necessary for inputting a policy into SPARCLE. Therefore, we will now discuss the next steps in the parsing research to further improve the usability of the workbench. Finally, we explore the generalization of the workbench techniques to other domains and levels of policy detail.

Table 1.Parsing results using organizations privacy policies

Organization	Conserv. Precision	Conserv. Recall	Liberal Precision	Liberal Recall
Government1	0.86	0.90	0.94	0.99
Finance1	0.89	0.95	0.93	1.00
High Tech1	0.89	0.97	0.91	1.00
Health Care1	0.91	0.96	0.94	1.00
Finance2	0.82	0.87	0.91	0.97
Government2	0.92	0.92	1.0	1.0
Average	0.88	0.93	0.94	0.99

6.1 Research areas involving SPARCLE Parsing

In order to improve SPARCLE's parsing accuracy, we need to look at the types of errors that were encountered when parsing organizational privacy policies. An analysis of these errors

showed that the vast majority of problems we encountered were of two basic types: 1) rules that refer to a data recipient or data source that was different from the data user and 2) words that have uses in multiple parts of speech and that were linguistically misclassified by the early stages of the parser.

Some privacy rules refer both to a data user and a data recipient or source. An example of such a rule is as follows: "Customer service reps can share customer mailing address with business partners for the purpose of expediting shipping." In this rule the data user is the "Customer service reps", the action is "share", the data category is "customer mailing address" and the purpose is "expediting shipping". However, there is also a data recipient mentioned that does not clearly fit into any of the categories. Therefore, a future version of SPARCLE will include optional data recipient and data source element types and the grammars will be tailored to identify these elements.

The other type of parsing error that we commonly encountered was caused by misclassified words. In general these are words that can be correctly used as multiple parts in speech. For example, "account" can be a noun (e.g. "bank account") or a verb (e.g. "He must account for his actions."). For this reason, these terms can be easily misclassified by a general purpose parser. However, within a given domain these words are often more likely to have a particular use. A future research issue is to build and incorporate domain specific dictionaries that allow us to specify word use for a particular domain.

6.2 Research areas involving SPARCLE Pre-processing

Currently SPARCLE users must perform pre-processing to prepare existing policies to be analyzed by the SPARCLE Policy Workbench. As explained earlier, this pre-processing takes two forms. First, non-rule text that exists in policies (e.g. introductions and definitions) must be removed so that the policy only consists of the rules and then the user must ensure that rules at least loosely conform to the basic structures shown in Figures 4 and 5. While participant's in our design walkthroughs of the current SPARCLE Policy Workbench indicated that this was acceptable, a future research direction is to tackle the issue of automatically identifying and processing definitions and other introductory policy information. We also plan to continue work to increase the grammars' ability to parse rules in less constrained formats than they can currently handle well. In particular we have a long-term goal of allowing the policy elements within a rule to be in any order that the author might desire.

6.3 SPARCLE Parsing Generalization

Currently the SPARCLE Policy Workbench supports the creation and management of privacy policies, however, we believe that the approach of combining natural language processing with a structured authoring and review method, and policy visualizations can be generalized into other policy domains. A future research effort is to modify the grammars and the associated interfaces to support other policy domains such as security policies, network management policies, and system management policies. Another interesting research issues concerns generalizing the SPARCLE Policy Workbench to support other languages. Currently SPARCLE operates on policies in English, however practical issues involving the international use of policies requires that the workbench be available in multiple national languages. The existence of parser and translation techniques for many other languages make this a viable area of future research.

In addition to extending the SPARCLE workbench to support other policy domains and languages, it would also be interesting to consider the need for generating a version of a policy for individuals who share data with an organization, but are not part of the organization. For example, while we have concentrated on the creation of privacy policies to be used within an organization, there is also a need to produce a version of privacy policies for the organization's customers, clients, and patients. We have found that organizations generally want their external policies to be generalizations of the internal, more detailed policies. Research is needed to explore to what degree internal policies need to be modified and generalized to produce a high quality external policy that is consistent with the internal policy.

7. Conclusions and Discussion

In this paper we have presented the progress that we have made on the SPARCLE Policy Workbench in the last year through the design, implementation, and evaluation of functional natural language parsing capabilities on constrained natural language policies. This research builds on an earlier, mid-level fidelity SPARCLE Policy Workbench prototype that we used to identify and confirm user requirements for a policy workbench [7]. The current SPARCLE Policy Workbench is designed to assist policy authors by allowing them to write privacy policies in natural language, parse the policies, visualize them and then produce a machine readable (XACML) form of the policy. In this paper we have described the parsing strategies that SPARCLE employs and the results that we have achieved on actual organizational privacy policies. We have explained how we created a set of grammars that execute within a shallow parser that achieves 94% liberal precision when parsing the policies we obtained from large healthcare, banking/finance, and government organizations that participated in the usability walkthrough sessions with the workbench.

During these evaluation sessions, the participants reported that they found the ability to parse policies, as SPARCLE is now implemented, to be a very favorable feature and this has encouraged us to continue research along this line. With this in mind, we have analyzed both the cases in which the parsing has returned incorrect results and the current pre-processing that is necessary. We have identified future research efforts to increase the SPARCLE parsing accuracy and lessen the amount of pre-processing necessary, as well as to generalize the workbench to include other policy domains, languages, and audiences.

In addition to the future research into the parsing of policies, we believe there is a great deal of research in policy management yet to be done. Regarding the current SPARCLE Policy Workbench prototype, there are at least two additional areas of future research. First, there is a need for a policy critic utility which will allow policy authors to identify inconsistencies within and across their policies. Second we plan to work with more complete sets of organizational policies in order to address scaling issues. Research issues downstream from the parsing include mapping the policy elements identified during the authoring step to the organization's implementation and providing internal audit and compliance tools to ensure that the policy is being enforced as intended and to allow organizations to answer inquiries about how

a particular customer, patient, or constituent's data has been used and for what purposes.

Currently, there is growing legal and societal pressure for organizations to be more vigilant at protecting the personal information that they collect and use in their business processes. However, many organizations currently use manual procedures rather than technology to enforce their policies and currently have no way to ensure that the policy enforcement implementation is what was intended by the written policy. The SPARCLE Policy Workbench, when it is complete, will provide users with the ability to link the natural language policies with the implemented policy and audit that policy to make sure that it is being enforced as intended. The SPARCLE Policy Workbench is a good example of how the application of HCI methods can facilitate the development of the usable privacy and security tools that are necessary for both organizations and individuals to protect personal information.

8. REFERENCES

[1] Ackerman, M., & Mainwaring, S. (2005). Privacy issues in human-computer interaction. In L. Cranor & S. Garfinkel (Eds.) Security and Usability: Designing Secure Systems That People Can Use, Sebastopol, CA: O'Reilly, 381-400.

[2] Anderson R. J. A (1996). Security Policy Model for Clinical Information Systems. *In the Proceedings of the 1996 IEEE Symposium on Security and Privacy*, 30-43.

[3] Anderson R. J. (2000). Privacy Technology Lessons from Healthcare. In the Proceedings of the 2000 IEEE Symposium on Security and Privacy.

[4] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2003). Implementing P3P Using Database Technology. *Proceedings of the 19th International Conference on Data Engineering*, Bangalore, India.

[5] Ashley, P., Hada, S., Karjoth, G., Powers, C., and Schunter, M. (2003). *Enterprise Privacy Architecture Language (EPAL 1.2).* W3C Member Submission. http://www.w3.org/Submission/EPAL/

[6] Bohrer, K., Levy, S., Liu, X., and Schonberg, E. (2003). *Individual Privacy Policy Based Access Control. In Proceedings of the 6th International Conference on Electronic Commerce Research (ICECR-6).*

[7] Brodie, C., Karat, C., and Karat, J. (2005). Usable Security and Privacy: A Case Study of Developing Privacy Management Tools. *Proceedings of the Symposium on Usable Privacy and Security, (SOUPS '05)*, ACM Digital Library.

[8] CRA Conference on "Grand Research Challenges in Information Security and Assurance". http://www.cra.org/Activities/grand.challenges/security/. November 16-19, 2003.

[9] Cranor, L. (2002). *Web Privacy with P3P.* Cambridge: O'Reilly.

[10] Cranor, L. (2005). Privacy policies and privacy preferences. In L. Cranor & S. Garfinkel (Eds.) *Security and Usability: Designing Secure Systems That People Can Use,* Sebastopol, CA: O'Reilly, 447-472.

[11] IBM Research UIMA(2005) http://www.research.ibm.com/UIMA/

[12] IBM Tivoli Privacy Manager for eBusiness (2004). http://www-306.ibm.com/software/tivoli/products/privacy-mgr-e-bus/.

[13] Karat, C., Karat, J., Brodie, C., and Feng, J. (2006). Evaluating interfaces for privacy policy rule authoring. *Proceedings of the Conference on Human Factors in Computing Systems – CHI 2006*, ACM Press, 83-92.

[14] Karat, J., Karat, C., Brodie, C., and Feng, J. (2005). Privacy in information technology: Designing to enable privacy policy management in organizations. *International Journal of Human Computer Studies*, 63, 1-2, 153-174.

[15] Karjoth, G. and Schunter, M.(2002) A Privacy Policy Model for Enterprises. *Proceedings of the 15th IEEE Computer Security Foundations Workshop*, 271-281.

[16] Michael, J.B., Ong V. L. , and Rowe N. C, (2001) "Natural-language processing support for developing policy-governed software systems", *39th International Conference on Technology for Object-Oriented Languages and Systems,* IEEE Computer Society Press, pp. 263-274.

[17] Microsoft Internet Explorer (2004). Help Safeguard your privacy on the web. http://www.microsoft.com/windows/ie/using/howto/privacy/config.mspx

[18] Neff, M., Byrd, R., and Boguraev, B. (2003) The Talent system: TEX-TRACT architecture and data model. In Proceedings of HLT-NAACL *Workshop on Software Engineering and Architectures of Language Technology Systems*, Edmonton, Alberta, Canada.

[19] OASIS (2005). eXtensible Access Control Markup Language Version 2.0. http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf.

[20] OASIS (2005). Privacy Policy Profile of XACML v2.0. http://docs.oasis-open.org/xacml/2.0/PRIVACY-PROFILE/access_control-xacml-2.0-privacy_profile-spec-os.pdf.

[21] Ponemon Institute and IAPP. (2004). 2003 Benchmark Study of Corporate Privacy Practices.

[22] Smith, J. (1993). Privacy policies and practices: Inside the organizational maze. *Communications of the ACM*, 36, 12, 105-122.

[23] Whitten, A. and Tygar J. D. (1999) Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. *In Proceedings of the 9th USENIX Security Symposium*, August, 1999.

[24] W3C (2002) A P3P Preference Exchange Language 1.0 (APPEL 1.0). http://www.w3.org/TR/P3P-preferences/