

Laura L. Christopherson. Use of an ontology-based note-taking tool to improve communication between analysts and their clients. A Master's Paper for the M.S. in I.S. degree. November, 2005. 90 pages. Advisor: Jeffrey P. Pomerantz

Every systems design project starts with a conversation between clients and systems analysts/developers. Eliciting knowledge via communication is fraught with difficulty; yet it is required if requirements are to be discerned. When the success of the project rests on a wildcard like communication, it is no surprise that many IT projects go over schedule and budget and are replete with errors. By deliberately examining clients' natural language expressions of the problem space and then classifying knowledge contained therein, it was possible to create an ontology-based note-taking tool. This paper investigates and demonstrates success in the tool's ability to aid analysts in aligning their views of the problem space with their clients' views. Although the reliability of the tool requires further investigation, the raw numbers suggested that consistency of use is possible. Furthermore, the tool was considered to provide complete coverage of concepts by those who used it. It has potential and further work is needed.

#### Headings:

Systems Analysis

Communication

Terminology

Ontology

Requirements Engineering

USE OF AN ONTOLOGY-BASED NOTE-TAKING TOOL TO IMPROVE  
COMMUNICATION BETWEEN ANALYSTS AND THEIR CLIENTS

by  
Laura L. Christopherson

A Master's paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Information Science.

Chapel Hill, North Carolina

November 2005

Approved by

---

Jeffrey P. Pomerantz

## TABLE OF CONTENTS

INTRODUCTION.....	2
RESEARCH QUESTION .....	4
LITERATURE REVIEW .....	5
The Communication Problem .....	7
The Effect Communication Problems Have on Systems Design.....	12
Analyzing Language.....	14
Language-Oriented Tools and Methods .....	14
Why an Ontology .....	16
Ontology Engineering Guidelines .....	19
METHODOLOGY .....	20
FINDINGS.....	25
The Information Seeking Ontology.....	26
The Note-Taking Tool and Consistency.....	29
Coder Reactions and Completeness .....	37
Follow Up Interviews and Accuracy .....	39
Further Research Needed.....	39
CONCLUSION .....	41
REFERENCES.....	43
APPENDIX A: UML DIAGRAM.....	47
APPENDIX B: DATA DICTIONARY.....	48
APPENDIX C: SCREENSHOT OF THE NOTE-TAKING TOOL.....	65
APPENDIX D: CODING INSTRUCTIONS .....	67
APPENDIX E: CODER QUESTIONNAIRE .....	76
APPENDIX F: HOW CODINGS WERE ANALYZED.....	77
APPENDIX G: CALCULATION OF THE KAPPA.....	86

## INTRODUCTION

Every systems design project starts with a conversation, of some sort, between the clients (or “users,” recipients of the end product) and representatives (or systems analysts) from the group of developers charged with fulfilling the needs of the clients. This conversation and subsequent conversations are critical to the success of the project. Almost like two partners learning to dance with each other for the first time, these conversations are organic, fluid, variegated experiences happening at multiple levels. Each participant brings his own experiences and knowledge to bear on the situation but is, as yet, unversed in the language of the other. This level of unfamiliarity requires each participant to make conscious thoughts, feelings, ideas, and learned experiences that he may have neatly tucked away in the back of his mind, in the dwelling where only tacit knowledge resides; and sometimes there are stubborn tidbits that refuse to bubble up to the surface, remaining a mystery and only later identified when the finished product reveals their omission. Similarly, there will be notions that simply are not remembered and therefore never contributed to the conversation.

Each participant, like a dancing partner, must interpret nonverbal cues, inflection, emphasis, masked emotion, rhythm, and nuance. Who will lead and who will follow? Social constructs such as these serve to constrain the conversation alluding to issues of power and control, revealing bias and judgment, as well as affording an arena for organizational politics to erupt and gnaw at the edges of circumstance.

And then there is the conversation's everlasting focus on the problem that often manifests itself in what the clients *don't* want, in the negative, in limitation. Conversely, the discussion could take a more fantastical turn and veer in the direction of what the client *might* want in the future, opening the team up to the impossible and the potentially unused or unfinished product.

Eliciting knowledge via communication is necessarily fraught with difficulty; yet it is required if a systems analyst wishes to learn about the client, his needs and behaviors. This knowledge is what permits the analyst to determine system requirements and specifications. By knowing the end user, the analyst can know what will work for the end user. So when a wildcard like communication is that upon which the success of the project rests, it is no surprise that many IT projects go over schedule, go over budget, and are replete with errors which require more time and more money to correct. When presented with a product that in no way resembles what they had expected or that does not accurately mirror their needs and behaviors, clients feel disconnected from it. There is no sense of ownership, only a feeling of futility and wasted effort by all.

So what is to be done about this? Perhaps an analysis of the stuff of communication – natural language: words, phrases, sentences – is where the solution resides; and to go a bit further, perhaps bringing form and structure to natural language expressions to address issues of context and culture, different backgrounds and experience, tensions between tacit and explicit knowledge, memory problems, fantastical speculation, and limited scope. By deliberately examining the problem space and

classifying knowledge contained therein, it might be possible to produce a tool that helps systems analysts align their views of the problem space with their clients' views. This tool could be as simple as a form upon which a systems analyst could take notes during client interviews.

One possible way to classify this knowledge and then fashion it into a note-taking tool is through ontology engineering. Ontologies are “explicit formal specifications of the terms in the domain and relationships among them” (Noy & McGuinness, n.d., section 1, para 1). They are structured vocabularies; and it is this structure that might bring order to an otherwise messy situation by providing a consistent, dependable framework for guiding client interviews. Ontologies foster a “shared understanding of some domain of interest” (Uschold & Grunniger, 1996, p. 5). “Such an understanding can serve as the basis for communication” (Jiang, Zhang, & Wang, 2005, p.34). They are a “guarantee of consistency” (Gruber, 1995, p. 909). An ontology-based note-taking tool may help reduce misunderstanding, missed information, and help to overcome some of the barriers that thwart successful acquisition of requirements.

## **RESEARCH QUESTION**

The purpose of this paper is to investigate an ontology-based note-taking tool's ability to improve communication between systems analysts/developers and their clients (or users). This paper will evaluate the ontology-based note-taking tool on three levels: consistency, completeness, and accuracy. Specifically, will the tool help provide more consistent understanding of client utterances across multiple listeners? Will it aid the systems analyst in providing more complete representations of client utterances by

providing a framework of cues to prompt the client for information that might ordinarily be missed? And will the tool help the systems analyst more accurately match his perception of the problem domain to the client's perception by promoting *shared* understanding through the use of an agreed upon vocabulary?

Previous research on analyst-client communication has focused more on the communication itself without proposition of tools or aids for improving communication. If tools or aids have been proposed, very little empirical research has been done on their ability to improve communication. Since it is a relatively new area of knowledge, ontology research has primarily focused more on the act of engineering ontologies; or it has been explored more for use in domains other than requirements elicitation. Some work has been done in the use of ontologies for requirements elicitation but usually it has not centered on guiding and improving the verbal communication that occurs between analyst and clients during interviews. More, it has concentrated on automatic means of requirements gathering. Thus, this appears to be the first study to investigate an ontology-based tool's ability to improve understanding in client interviews.

Explicitly stated, the research question for this work is: Will an ontology-based note-taking tool improve a systems analyst/developer's ability to match his perceptions of the problem domain to the client/user's perceptions?

## **LITERATURE REVIEW**

Any system or solution, in whatever form, that is designed to solve a user's problem must have evolved from a set of requirements that constituted a design blueprint or, to use more formal terminology, functional specification. These requirements are most

often elicited from the user (or client) by the systems analyst (or designer/developer) through the use of some type of “communicative act” (Alvarez, 2002, p. 1), usually an interview. The days of programmers simply sitting down and writing code and then serving it up to clients are gone (Goguen & Linde, 1993). Al-Rawas and Easterbrook (1996) concluded that 77% of the users they interviewed confirmed that some type of verbal communication method was employed during interactions with systems analysts. Specifically 30% indicated formal meetings, 12% indicated telephone conversations, and 35% indicated some other type of face-to-face discussion (p. 9). Many researchers agree that successful requirements engineering and knowledge acquisition relies on effective communication (Alvarez, 2002; Coughlan & Macredie, 2002; Agarwal & Tanniru, 1990; Al-Rawas & Easterbrook, 1996; Jin, Bell, Wilkie, & Leahy, 2003). Communication of this type is geared towards “developing a shared understanding of an ambiguous situation” (Coughlan & Macredie, 2002, p. 53). “The root of the requirements problems lies in the common ground between the user and the designer, which can only be discovered through communication activities that facilitate a sharing of information” (Coughlan & Macredie, 2002, p. 53).

In order to successfully acquire user requirements, a systems analyst must build rapport with the user, understand the user’s motivations (Alvarez, 2002), negotiate issues of culture, politics, and organizational context which impinge on the development process (Alvarez & Urla, 2002). They must “map informal understanding into a formal information architecture” (Byrd, Cossick, & Zmud, 1992, p. 120). It is a collaborative act where “both interviewer and client are engaged in creating the meaning of the question



and answers that constitute the narrative as they negotiate understanding through language” (Alvarez & Urla, 2002, p. 40).

The meaning that is constructed is a representation of the problem domain. It is this discussion of the problem or “deficit subject” that opens the door for intervention (Alvarez, 2001, p. 8); and ecumenical understanding and analysis of this domain leads the analyst to formal specifications for how the problem should be solved (Kaindl, 1993), specifications that hopefully ensure successful implementation and use of the system. Yet the process of achieving salient requirements through communication is rather thorny. Many researchers agree that systems design is often impeded by the inherent complexities in discursive elicitation (Alvarez, 2002; Coughlan & Macredie, 2002; Alcázar, 2000).

### **The Communication Problem**

Suppose there is a spectrum or continuum that reflects the measurement of how successful or easy communicative interactions are. When two close friends get together, they begin their interaction from a positive point on this spectrum. They are “in the black” so to speak. They know what to expect from each other, for the most part. More than likely, they share common experiences, vocabulary, and background (Kelsey, 1996); and this simplifies and eases communication between the two. Systems analysts and clients typically start from a negative point on the spectrum; they are “in the red.” Often, they do not hold much in common; and often, they are meeting each other for the first time. To complicate matters further, there are many factors which influence or comprise communication, making it a complex, multi-layered phenomenon. Not only must the

basic elements of communication such as pitch, tone, inflexion, gesture, facial expression, and diction be interpreted (Alvarez, 2002), there are a slew of other factors that converge to make communication a singularly perplexing act. Social context, background experiences, vocabulary, bias, power, control, politics, role presentment, tacit knowledge, explicit knowledge, knowledge subject to the erosion of memory, ill-formed ideas, uncertainty, subject matter constraints, and capriciousness head the list. Starting in the negative only compounds these issues.

Communication is multi-modal. Not only are there a variety of signals (pitch, tone, etc.) that constitute any given message, information can be obtained through variety of mediums such as documents, observation, and interviewing (Marakas & Elam, 1998). Messages flowing through channels back and forth must be interpreted and responded to. The nature of these interpretations and responses shape and re-shape and shape again, with each new utterance, the meaning that is created in the dialogue.

Interpretation of signals requires an understanding of the social underpinnings of the situation and the participants involved. Alvarez (2002) describes requirements as “socially mediated;” Flynn & Jazi (1998) proffer the phrase “socially constructed.” Many researchers concur that requirements elicitation and systems development is characterized by a variety of socially-oriented attributes (Kilker & Gay, 1998; Hands, Peiris, & Gregor, 2004; Sutton, 2000; Goguen, 1993; Antón & Earp, 2004; Jiang, Zhang, & Wang, 2005; Robinson & Bannon, 1991). Whether conscious or not, participants bring unique perspectives, motivations, agendas, emotions, and world views to the table. Each participant is situated in various social contexts and has been ascribed with different determinations of status and role depending upon the milieu. Individualized past

experiences have personally shaped each participant's attitudes, ideas, and responses, giving way to vastly different mental models. Sutton (2000) speaks of participants as having a private world that must be transformed into public languages. Each participant has his own level of knowledge and expertise, skills, and requisite vocabulary. What means one thing to one person may have a completely different meaning or a highly ambiguous meaning to another; and when these concepts are represented formally (such as being captured in functional specifications documents), they tend to shift meaning when passed between different semantic communities. Robinson and Bannon (1991) refer to this as "ontological drift." Hence, assuming that there exists one objective meaning (Galliers & Swan, 2000), "complete or permanent" (Gerson & Star, 1986) of any particular notion could be perilous.

Additionally cultural influences must be exposed and will give further dimension to the nature of such communication challenges. For instance, clients "may feel that they do not have the same level of technical knowledge as the developer and may be inhibited or intimidated during a face-to-face interview with the developer, as they do not wish to appear foolish" (Hands, Peiris, & Gregor, 2004, p. 502). Conversely systems analysts/developers "tend to lack domain or business knowledge and consequently tend to misunderstand or ignore some requirements and their social context" (Coughlan & Macredie, 2002, p. 52). Flynn and Jazi (1998) refer to this as the "user-developer culture gap" (p. 53). This gap leads to power struggles between participants. Clients often feel that they are at a disadvantage or in a lesser position of power; and this is often reinforced by the fact that the systems analyst often directs or controls the flow of dialogue. As Alvarez (2002) discovered, the client is more likely to fall into a personal narrative or

story of events; and the systems analyst reframes what the client says in more succinct terms or with more pointed questions due to a need for specificity and brevity. Clients view these interactions more personally while analysts tend to view them more professionally. This constant interruption and re-interpretation of what the client says necessarily elevates the analyst above the client. Consequently, this leads to further confusion about the roles each participant should assume. Cooperation is negotiated through constantly shifting roles (Coughlan & Macredie, 2002) as defined by moment-to-moment interpretations of what the other participant seems to be signaling as acceptable behavior.

In addition to power tensions between an individual systems analyst and an individual client, the requirements elicitation process is overshadowed by organizational politics. Other stakeholders can have tremendous impact on the process. For instance, even the constitution of the development team (i.e. representatives from both sides) is subject to politics. Al-Rawas and Easterbrook (1996) found that the selection of team members from client organizations was not always made wisely. Some team members were chosen because they could be spared from regular work duties instead of being selected based on subject matter expertise. Sometimes team members were primarily management, thus systems were developed that almost completely ignored those who would actually use the system. Furthermore, pluralism in terms of participants not necessarily holding with the organization's mission or objectives, political manipulations by participants (Galliers & Swan, 2000), and the swiftly changing, "highly fragmented and differentiated" constitution of client organizations (Clegg et al. 1997, p. 858) add further cultural complications.

Then there is the nature of the information communicated to add yet another dimension. Many researchers agree that users have difficulty specifying what they need (Hands, Peiris, & Gregor, 2004; Sutton, 2000; Coughlan & Macredie, 2002; Jin et al. 2003; Goguen & Linde, 1993). Often this is due to the inability to express tacit knowledge (Polanyi, 1966; Sutton, 2000; Alvarez, 2002; Coughlan & Macredie, 2002; Galliers & Swan, 2000). In other words, “we can know more than we can tell” (Polanyi, 1966, p. 4). There is a gap between what one knows unconsciously, deep in the recesses of one’s mind, and what one is able to talk about explicitly. A classic example is to ask someone to write instructions on how to tie a shoe. This is something done almost every day by a wide number of people. It is a task that has become so ingrained that it is no longer conscious (as it was when it was first being learned as a child) and therefore not easily expressed. So when a systems analyst asks a user to speak specifically about a task, a task that should be modeled in the system, the analyst is asking the user to reconstruct from memory the detailed facets of the task. Because the user is probably not within the context of the task at the time he is being asked about it, he is less likely to provide a meaningful description. Moreover, the process of recall is further constrained by the “limited storage capacity” of one’s memory (Agarwal & Tanniru, 1990, p. 124). This produces yet another undesired effect: users will cease to speak from a place of authenticity. Instead they will speculate and make fantastical assumptions about how they *might* perform a task or what they *might* want in a system. Since these ideas cannot be verified with any facts grounded in reality, they could potentially lead to impossible requirements, unfinished systems, or, simply, unused systems.

Typically, what comprises the focus of such discussions is “the problem,” that which is negative. According to Sutton (2000), users are better equipped to talk about what they do *not* want rather than what they *do* want. This focus on that which is deviant can unfortunately narrow or limit the discussion thereby obscuring relevant requirements that are not directly associated with problems per se, yet have significant value and impact and should be represented in the resulting system (Alvarez, 2001).

Lastly, the iterative, unpredictable disposition of communication and specifically requirements elicitation interviews further aggravates the situation. Requirements are rarely fixed (Goguen & Linde, 1993). They are constantly emerging, fluctuating, being modified and honed (Galliers & Swan, 2000). The mercurial proclivity of requirements means that the process of obtaining them is also erratic and therefore requires successive iterations (Alcázar & Monzón, 2000). Cycling through multiple interviews and attempting to sort through compound sets of rich data can be tedious and time-consuming. The stamina and attention to detail required is enormous; and important requirements are bound to be missed.

### **The Effect Communication Problems Have on Systems Design**

Researchers comment that imprecise requirements are frequently the cause of poorly developed and underused systems or projects that simply fail (Curtis, Krasner, & Iscoe, 1988; Marakas & Elam, 1998). They expand on this by concluding that poorly formulated requirements are often the result of the communication problems discussed above; and it is also believed that system successes are attributable to the careful identification of impeccable requirements.

The research conducted by Clegg et al. (1997) reveals that information technology projects are vital to organizations and that “many organizations cannot function successfully with them”; yet “80-90% of investments in new technology fail to meet all their objectives” (p. 855). Ironically, the “most common single reason [to implement IT projects] is focused on cost reduction” (p. 858). Many researchers cite projects going over budget, delivered late, fettered with errors, and resistant to adoption by users as the consequences of poorly specified requirements (Hands, Peiris, & Gregor, 2004; Marakas & Elam, 1998; Alvarez, 2001; Clegg et al. 1997; Sutton, 2000; Coughlan & Macredie, 2002; Jenkins, Naumann, & Wetherbe, 1984). Specifically, Jenkins et al. (1984) discovered that the most commonly used information gathering technique was the interview (both individual and group interviews). They found that 78% of project leaders said that requirements were discovered after the approval period and that 65% attributed this to “faulty or incomplete analysis” (p. 78 –79). “Most projects are completed significantly over budget, require more effort than estimated, and are completed late” (Jenkins et al. 1984, p. 81). Because, as was mentioned earlier, some type of verbal communication, usually in the form of interviews, is most often employed to elicit requirements and because unsatisfactory requirements are at the root of development problems such as cost and late delivery, it is reasonable to assume that the intrinsic complications of discourse contribute significantly to these development problems. Therefore, to solve these problems, that which comprises the root of communication, and therefore the root of the problem, must be examined.

## **Analyzing Language**

Communication is fundamentally linguistic. Systems analysts and clients “negotiate understanding through language” (Alvarez & Urla, 2002, p. 40). Interviews allow the systems analyst to probe the client for more details, reflect back to the client and receive feedback on discussed concepts. This type of communication is collaborative and requires both parties to be fully engaged in the communicative act. What a client actually says “provides a window for understanding clients’ vocabularies and conceptualizations” as well as alluding to cultural constructs (Alvarez & Urla, 2002, p. 47-48). Because communication is grounded in language, an analysis of the expressions used by clients may provide a clue as to how to ameliorate misunderstanding and misinterpretation that ultimately leads to unsatisfactory requirements.

## **Language-Oriented Tools and Methods**

There are an abundance of tools and methods to support the requirements engineering process: object-oriented approaches, prototyping, participative design, ethnographic methods, scenarios, use cases, contextual inquiry, etc. Each tool targets one or more facets of the development process and was developed to distill certain traits of the client and his needs. Conversely, Westrup (1999) explains that the tool or method of choice also influences the analyst; and while it shapes the analyst’s approach to the situation in favor of eliciting certain types of data during certain phases of development, it also shapes the very substance of information collected. Thus, a concordance of many tools is often preferred to develop a comprehensive picture of the client-organization’s needs.



Hands, Peiris, and Gregor (2004) propose a computer-based interviewing tool that walks the user through a series of questions so as to reduce any discomfort a user might feel in an initial face-to-face interview. However they assert that a face-to-face interview is imminent and can be facilitated by a print-out of the data collected via the computerized tool.

Marakas and Elam (1998) conducted research on the efficacy of semantically structured interviews. Similar to ideas conveyed in Pomerantz's (2005) review of question taxonomies, Marakas and Elam believe that interviews should be a confederation of various types of questions such as those that seek to verify information or those that seek to define concepts or those that seek to determine causes, etc. They concluded that semantic structuring improves the accuracy of accumulated information and reduces errors. It appears to strengthen the abilities of new analysts; and resultant logical representations tend to be more closely associated with the original discourse.

In an attempt to address the "user-developer culture gap," Flynn and Jazi (1998) present user-led requirements construction where users are trained in modeling requirements and asked to build their own models. While time-consuming and potentially confusing for users, this approach could make users feel that they are on more equal footing with analysts/developers.

Kaindl (1993) asserts the use of object-oriented classes that classify the concepts acquired during elicitation. Manipulation of the data is achieved through a hypertext interface. Unlike the work in this paper, Kaindl's work focuses more on what happens *after* the information is collected from clients, specifically in translating that data into formal requirements.

Alvarez's (2002) work centers on analyzing the actual discourse between clients and analysts to help expose social substratum and make analysts more aware of important clues that might ordinarily be missed if they do not encourage the narrative frame from which clients tend to operate.

Travers and Haas (2003) aim to reduce ambiguity of concepts expressed in hospital emergency rooms through the use of natural language processing. By computationally whittling terms down to atomic properties, they are able to formulate a concise, consistent terminology that allows for cross-referencing of terms so that a multiplicity of user vocabularies may be aligned.

Barbosa et al. (2005) build a lexicon to aid communication. The lexicon is substantially more informal than an ontology (i.e. not drawing upon the inferential capabilities of an ontology). They provide a hypertext interface for the lexicon and propose its use in conjunction with scenarios.

### **Why an Ontology**

Jin, Bell, Wilkie, and Leahy (2003) state that "ontology intends to address the problem of lack of a shared understanding and the consequent poor communication among people, which is precisely the problem in requirements elicitation" (p. 3). Because ontologies are designed to capture natural language descriptions of domains of interest, provide more consistent representations of such domains, reduce ambiguity inherent in communication, reduce error via the structuring of knowledge, and provide ways to extend and specialize captured domain knowledge, an ontology is proposed as the

support structure for a note-taking tool that could be used during client interviews, to hopefully improve communication between participants.

Gruber's (1995, p. 908) definition of an ontology is "an explicit specification of a conceptualization." Although rather vague, it appears to be the most cited definition. Guarino (1997) attempts to refine this definition: "An ontology is a logical theory that constrains the intended models of logical language" (p. 298). Noy and McGuinness (n.d.) provide a definition that, although somewhat general, is less technical and therefore more accessible to laymen: "An ontology defines a common vocabulary for researchers who need to share information" (section 1, para. 2). Uschold and Gruninger (1996) proffer this definition:

'Ontology' is the term used to refer to the shared understanding of some domain of interest which may be used as a unifying framework to solve ... problems [of poor communication, specifications of systems, limited inter-operability, and the need for re-use and sharing]....

An ontology necessarily entails or embodies some sort of world view with respect to a given domain. The world view is often conceived as a set of concepts (e.g. entities, attributes, processes), their definitions and their inter-relationships; this is referred to as a *conceptualization* (p. 5).

Sowa (2000) asserts that an ontology "is the study of existence, of all the kinds of entities – abstract and concrete – that make up the world. It supplies the predicates of predicate calculus and the labels that fill the boxes and circles of conceptual graphs. The two sources of ontological categories are observation and reasoning" (p. 51).

I offer the following definition: An ontology is a vocabulary that describes a particular domain of interest through the identification of classes (or entities) and their properties (or relations to one another). It can be expressed in first-order predicate logic and thus has inferential capabilities not present in taxonomies.

Ontologies foster “shared understanding” (Uschold & Gruninger, 1996, p. 1) between disparate players (Mizoguchi & Ikeda, 1997) by “reduc[ing] and eliminat[ing] conceptual and terminological confusion” (Jiang, Zhang, & Wang, 2005, p. 34). In short, they ensure consistency (Gruber, 1995) which is critical to the requirements engineering process. Consistent understanding of the domain of discourse reduces ambiguity, lessens the impact of contextual/cultural differences between players, and grounds discussion in current reality.

They “support the visualization and editing of charts which show snapshots of the information space” (Heflin, 2004, section 2.4). Because they are structured, they can improve accuracy and match logical representations more closely to the originally expressed information. As Agarwal and Tanniru (1990) discovered, structuring communication is also advantageous due to its ability to bolster the interviewing skills of novices. Additionally it provides a set of cues for prompting clients so that memory problems or the dilemma of breaking through the tacit knowledge barrier is reduced. The capriciousness of the process could become more ordered because of the reliability of ontologies.

Moreover, ontologies are uniquely positioned to be conducive to both clients and programmers. There is a natural language exterior which is apprehensible for clients and a logical, object-oriented interior which is so for programmers. They are re-usable and

extendable to a variety of domains and therefore can be specialized for the needs of each new client (Uschold & Gruninger, 1996), capturing the individuality of different social groups and quantifying it into a cohesive framework.

Ontologies are being used in a variety of ways. Those that are similar to the work of this paper include Carreno et al. (2000). They are using an ontology-based approach to knowledge acquisition from text through the use of natural language recognition. Uschold et al. (1998) have constructed the Enterprise Ontology to aid developers in taking an “enterprise-wide view of an organisation” which can then be used to aid in decision making, requirements specification, and communicating and sharing knowledge across an organization (p. 31). Jin et al. (2003) build off of Uschold’s work by proposing a three-tiered ontology to specifically guide the requirements elicitation process. Their approach is intended to automate both interactions with clients and the development of application models.

Using ontologies to shape the requirements engineering process is clearly not a new idea. However, the work outlined above is not specifically addressing the issue of improving natural language communication between stakeholders in an interview in order to achieve, down the road, more polished requirements. This paper will investigate this idea; thus requiring a transition to a discussion of how ontologies should be crafted.

### **Ontology Engineering Guidelines**

“Representations of reality actively construct rather than passively reflect the world” (Bloomfield & Vurdubakis, 1994, p. 10). Therefore, it is incumbent upon the ontology engineer to adhere to certain principles that will ensure the integrity of the

world modeled in the ontology. Noy and McGuinness (n.d.) recommend making a clear determination of scope and domain at the outset of ontology development. Gruber (1995) contends that definitions that describe classes should be objective and they should be documented in natural language. He goes on to say (and others agree with him), that ontologies should be extendable and capable of being specialized for other domains (Uschold & Gruninger, 1996; Heflin, 2004). The W3C (Heflin, 2004) furthers these guidelines by suggesting that ontologies be highly expressive and easy to learn and use. What's more, they and others require consistent business rules as evidenced by the satisfaction of competency questions (Uschold & Gruninger, 1996; Noy & McGuinness, n.d.). Gruber (1995, p. 910) calls for "minimal encoding bias" thereby requiring compatibility with other tools and languages (Heflin, 2004). And lastly, Uschold and Gruninger conclude, "If an ontology is a framework for communication among people, then the representation of the ontology can be informal, as long as it is precise and captures everyone's intuitions" (p. 27).

Based these guidelines, an ontology has been developed to describe the problem space as communicated to the researcher. The ontology has then been translated into a tool to assist systems analysts in shaping communication with clients. Essentially, this tool, a tool for inscribing notes from client interviews, is the user-interface for the ontology.

## **METHODOLOGY**

This study examines the natural language answers given to the interview question: What difficulties or frustrations have you experienced in trying to find the information

you need to do your job? Forty-three employees of two North Carolina state government agencies were interviewed. Interviewees were chosen via a combination of purposive and convenience sampling. Specifically, these state employees were originally interviewed as a part of another study that focused on determining NC state employees' information needs and uses for the purpose of making recommendations to the State Library of North Carolina so that they might improve their services to state employees. Interviewees were selected based on contact lists received by the State Library. Any potential interviewees that participated in the State Library study and were able to provide an answer to the above question were considered for this Master's Paper research. A total of 50 answers (or problem expressions) were gathered from the 43 state employees. Thus, some state employees provided multiple examples of information-seeking difficulties/frustrations. Additionally 3 of the 50 problem expressions were gathered from online surveys that comprised second-round sampling for the State Library study.

These expressions were audio-recorded during interviews and the transcripts of these served as the source material for the development of the ontology. After careful analysis of these transcripts, salient features were identified for inclusion in the ontology which was then created using Protégé (<http://protege.stanford.edu>), a free-ware application offered by Stanford University. The OWL plugin was used so that the ontology would meet the specifications of the W3C's OWL Web Ontology Language. A Protégé OWL tutorial composed by Matthew Horridge, et al. (Horridge, 2004) was used as a guideline for the ontology's creation.

After undergoing multiple revisions, the final version of the ontology was checked for concept consistency using various features of Protégé in conjunction with the

Racer Pro 1.8.1 reasoner (also free-ware) offered by Racer Systems (<http://racer-systems.com>). The ontology was then visually inspected using the OntoClean methodology devised by Nicola Guarino and Christopher Welty (2004).<sup>1</sup> A UML model and data dictionary view of the ontology are located in Appendices A and B respectively. A CD-Rom is also enclosed that contains the electronic Protégé files and HTML files.

From the ontology a note-taking device was engineered. This tool was composed in HTML as an online form. A screenshot of the tool is located in Appendix C. This tool could have easily been drafted in Microsoft Word but the researcher felt that her greater adeptness at formatting content in HTML rather than in Word would produce a more usable tool for coders other than herself.

To test the reliability of the tool, the afore-mentioned problem expressions were coded into the tool by a total of 6 coders (or raters), one being the researcher. Upon submission of codings, the tool emailed results to the researcher who then tallied ratings in a Microsoft Excel spreadsheet for later percent-of-agreement (or Kappa) calculation. The tool's intended purpose is to simulate a real-world experience where an analyst would sit with a client during a client interview and enter appropriate interview information into the tool. Conversely, because coders worked from already captured problem expressions with no opportunity for probing clients, their ability to interpret the textual material was limited.

Each coder, with the exception of the researcher, coded 20 problem expressions.

The researcher coded all 50 expressions. This means that any given problem expression

---

<sup>1</sup> Natasha Noy at Stanford was very helpful in attempting to assist me with an automated way of evaluating the ontology using an OntoClean ontology that could be imported into Protégé. The original RDF file of the OWL version of the OntoClean ontology had errors; and Ms. Noy fixed these errors and provided alternative help information about its use in OWL (the help information on the Protégé site was incorrect for the OWL plugin). Unfortunately, I was never able to get the file to work because Protégé kept locking up in the middle of the process or it would corrupt the file; but I greatly appreciated Ms. Noy's help all the same.



was coded by a total of 3 coders. Coders were given a set of detailed instructions that, although quite lengthy, was necessarily so to ensure consistent understanding of the concepts involved. This set of instructions was reviewed by one coder and the advisor to this Master's Paper before presentation to all coders. These instructions are located in Appendix D. Pop-up help information was also available on various form items on the online note-taking tool. Coders could mouse-over links and information would pop up to provide additional help to them. Additionally, the author made herself available via phone and email for questions. Inter-rater reliability was evaluated using the formula outlined by Fleiss (1971) because the traditionally used calculation does not make allowances for more than 2 coders on a given subject. See Appendix G for specifics on how this calculation was performed.

To ensure validity of the tool's ability to foster and improve communication, the researcher re-interviewed 15 (or 35%) of the original interviewees. One of these interviewees had offered 2 examples of experienced difficulties, making these follow-ups 32% of the original sample size of 50 expressions. Invitations for follow-ups were sent to all original interviewees who indicated, during the State Library interview, that they would be open to a follow up. All those that responded to these email invitations were included in this set of second-round interviews unless timing and other constraints prevented inclusion.<sup>2</sup>

Each second-round interviewee was emailed a document that contained the specifics of the study, a transcript of their answer to the question stated at the beginning

---

<sup>2</sup> For instance, the researcher had scheduled an interview with one person who was not available at the scheduled time. When this person was approached again to re-schedule, he asked that the researcher "keep trying." After several phone calls, the researcher gave up.

of the Methodology section, and two follow-up questions. They were asked to review the transcript and questions prior to the scheduled phone interview.

The phone interview was initiated by asking the interviewee if they had any further questions about the study that might not have been answered in the document that was emailed to them. Then they were asked if they had read the transcript. If not, they were asked to do so. Then, referring both to the transcript and the print-out of my coding of the problem expression, I explained to them what I had understood them to say. I illuminated all the features touched upon in the note-taking tool and then asked them to assess the accuracy of my understanding with the following two questions:

1. How accurately did I understand your problem?
  - 0 – Not at all. That's not at all what I meant.
  - 1 – You understood me somewhat. Some parts are true but most points were omitted or inaccurate.
  - 2 – Your assessment of what I said was pretty good but still lacking.
  - 3 – Your assessment of what I said was completely accurate. You nailed it!
2. What facets of your problem did I misunderstand or leave out?

If a participant selected anything other than “3 Your assessment of what I said was completely accurate. You nailed it!” for the first question, then they were asked the second question. Otherwise, the second question was unnecessary. Results from these follow-up interviews were tabulated; and they and other findings are discussed in the next section.

## FINDINGS

Tom Gruber asserts that an ontology is a “systematic account of Existence” (Gruber, 1995, p.908). In order for an ontology to be logically rigorous, thereby allowing maximum inferential capabilities, an ontology ought to be able to clearly delineate concepts from one another. Mutual exclusivity would be preferred and alternative interpretations discouraged by the supporting axioms. An ontology should be able to say something objective about the real world. For instance, it is true that people exist. It is also equally undisputable what the qualities are that constitute a human being. Human beings are mammals; they have the capacity for speech and higher thinking. These qualities, and probably some others, clearly demark a class called HumanBeing. It would be obvious what types of individuals belong to this class.

However, most of the real world is highly subjective. Buddha says, “All we are is a result of what we have thought.” Thus, representing any real world situation via an ontology is an exercise in resolving tension between the ability to clearly define objective concepts and the inherent subjectivity of those concepts as they exist in the real world. The information seeking ontology this paper reports on is an exaggerated case of this because the essence of this ontology is how the information seeking process is *perceived* by the interviewees. Almost all classes in this ontology are representations of attitudes about given facets of the search task. What is expensive to one person may not necessarily be so to another, for example.

## The Information Seeking Ontology

Gruber's (1995) design criteria for ontologies insists on clarity. "An ontology should effectively communicate the intended meaning of the defined terms. Definitions should be objective" (p. 909). "Definitions should restrict the possible interpretations of terms" (p. 910). Making classes objective is not completely feasible with value judgments, that which comprises the information seeking ontology.

"An ontology should make as few claims as possible about the world being modeled, allowing the parties committed to the ontology freedom to specialize and instantiate the ontology as needed" (Gruber, 1995, p. 911). This "weak theory" is in conflict with the need to make classes objective and singularly interpretable; but Gruber acknowledges this conflict. The information seeking ontology uses this weak theory in that it allows for different assessments of *Timeliness*, *Existence*, *Worth*, etc., by different interviewees. Each client can assess the same piece of *Information* differently. One can say that its *Pliability* was *UsableAsIs* while another can say that it *RequiresManipulation*. As long as the definitions are clearly articulated alongside these loose commitments, Gruber says this meets both the clarity and minimal ontological commitment criteria. The data dictionary that accompanies this ontology is very strict while the instructions for the use of the tool clearly state that coders/analysts should approach each expression *from the user's perspective*. The commitment to this ontology is defined as: agents should draw from the client/user's perspective and attempt to find the corresponding rigid class and requisite individual/instance that addresses the user's opinion most accurately while bearing in mind that what is, for example, *Tedious* to one user may not be so for another.

Furthermore, because the tool can influence the analyst, neutral terms were chosen for classes so as to pre-empt bias and prejudice. For instance, instead of the word “Barrier” which makes a negative value judgment, the term *AccessRequirement* was used. If it can be inferred that an *AccessRequirement* is acting as a barrier, then it is a matter of interpretation by later reviewers of the gathered data rather than something inherent in the ontology itself that could provoke bias at the outset. In other words, because the ontology is objective and the term *AccessRequirement* is used, the analyst is not being predisposed to viewing all access requirements as potential barriers.

Individuals/instances, however, do contain more judgmental terms; but because these concepts (such as *Tedious*) are especially crafted to extract opinion, they must include negative assertions. Individuals are intended to reflect the client’s opinions and the classes are intended to guide the analyst. The neutrality at the class level reduces bias, while the range of emotion and opinion expressed at the individual level accommodates the client.

Gruber (1995) asserts that ontologies should be coherent; “it should sanction inferences that are consistent with the definitions” (p. 909). Since this ontology passed the consistency checks made with Protégé and Racer as well as the visual inspection via the OntoClean methodology, the information seeking ontology will be considered coherent.<sup>3</sup>

“One should be able to define new terms for special uses on the existing vocabulary, in a way that does not require the revision of the existing definitions” (Gruber, 1995, p. 910). Easily new *InformationCharacteristics* or

---

<sup>3</sup> The information seeking ontology was also visually inspected by Dr. Jeffrey Pomerantz and Dr. Stephanie Haas, at the School of Information & Library Science at the University of North Carolina at Chapel Hill.

*SearchTaskCharacteristics* or *HumanEntities*, etc. may be added to the information seeking ontology without restructuring the hierarchy or relationships between classes. Thus, the information seeking ontology is extendible. It has been crafted to address the issues outlined in the Literature Review. Systems analysts are charged with researching and understanding not only the problems their clients/users face but also their tasks, their cultural environment, constraints or aids that thwart or assist in task completion, their goals and end results, and the artifacts of their work. There are also features which are generalizable to other domains. The concept of actor/agent is captured in the *HumanEntity* class. Culturally oriented relationships between *HumanEntities* are captured in the *hasInfluenceOver* (and its inverse) and *isCooperativeWith* properties. Monetary and financial concepts are captured in *InformationCharacteristic:Cost*, *InformationCharacteristic:Worth*, and *AccessRequirement:Expenditure*. The concept of task is captured in *SearchTask* and could be easily adapted to suit other types of tasks. End results or goals are modeled in the class *Information* (the object of all *SearchTasks*) and *SearchResult*. Other end results could easily be added or swapped out for these. Artifacts are modeled in the *Container* class. Attitudes are captured most specifically in *InformationCharacteristic*, *SearchTaskCharacteristic*, and *ContainerCharacteristic*. *Knowledge* and *Learning* might also be needed in describing other domains.

The classes that might be more specialized to the information seeking domain would be *Surrogate* (which could easily be deleted if necessary) and *AccessRequirement*. *AccessRequirement*, however, helps to symbolize cultural constraints such as power over resources. Although its adaptation to suit other domains could potentially violate Gruber's criteria of extendibility (because it might require significant alteration of both

classes and relationships), it can serve as a model for including such cultural constraints in adaptations of this ontology.

Lastly the information seeking ontology satisfies Gruber’s criteria of “minimal encoding bias.” It is not dependent on symbol-level encoding as evidenced by the coders ability to use it without being presented with the logical inferences or mathematics behind it.

### **The Note-Taking Tool and Consistency**

In the Research Question section, it was stated that this research would evaluate the ontology-based note-taking tool on three levels: consistency, completeness, and accuracy. How frequently the coders agreed in their codings of problem expressions aims at a determination of consistency. Before this can be discussed, a few notes about the how the ontology was translated into the HTML tool must be made.

On the tool, the fields for Analyst, Client, and Search Task Title were added largely to help distinguish one coding from another. Coders were asked to enter their first names into the Analyst field<sup>4</sup> and the problem expression ID<sup>5</sup> into the Client field. (The Client field could correspond to the *FirstPerson* class in *HumanEntity*.) In a real world scenario, the full names of both parties would be entered.

The Search Task Title field is a place where both coders and real world analysts would enter a simple catch phrase to help delineate one problem expression/interview from another. (This corresponds to the *SearchTask* class.) Some other type of ID number would need to be added to help establish the record’s uniqueness in a real world situation.

---

<sup>4</sup> All coders had different first names; so last names were not needed to distinguish one from another.

<sup>5</sup> Each coder was given a list of problem expressions that were distinguished from one another by a unique identification number.

The **Date** field is acting as a place marker for real world use where such information would be important for project tracking needs.

In the ontology, the class *Container* is represented by the term **Location** in the note-taking tool. Although *Container* is a better representation of all the dimensions of this concept (as indicated in the data dictionary), coders did not need to be aware of such details and so the word **Location** was chosen as a term that would be more familiar to them. The **Location** option, **Another Person**, represents the transitive relationship of *HumanEntity isKeeperOf Container isKeeperOf Information* where it can be inferred that *HumanEntity isKeeperOf Information* and therefore can serve as a *Container/Location*.

*Journal* has been referenced as **Journal/Magazine** in the tool. Although the definition for *Journal* in the data dictionary clearly distinguishes it from a regular magazine, this extra step of adding another class called **Magazine** in the ontology and separating the two in the tool did not seem necessary at this time. If more problem expressions were evaluated where such a distinction seemed necessary (in that, for example, the two types of resources posed different search challenges for users and therefore should be analyzed separately), then a **Magazine** class could be added.

Lastly, there is an **Unspecified Article** listed under **Location** on the tool. This did not need to be included in the ontology since it represents a case where an analyst's probing of the client would help elicit a more specific *Container/Location* in a real world situation. The researcher did very little probing of interviewees in obtaining problem expressions in the hopes of obtaining bare bones descriptions of experienced problems – bare bones in the sense of not already interpreted by the researcher. The only times that probing occurred was when there was linguistic confusion over a term or phrase used,



never for more complete information. The reason for this is because the manner in which the researcher would probe a client might be very different in how another would probe. Thus any information gained from such probing could be viewed as interpreted in some way. For instance, if an interviewee said, “I had problems searching the website” and was then asked, “Did you actually use a search engine or did you simply navigate the site,” the answer to this question could be seen as evidence of the researcher’s interpretation that “searching the website” did not necessarily mean using a search engine. Another analyst might have decided it did and would never have asked the follow-up question. Even this minimum of interpretation was not desired so as to fully exploit the coders’ interpretation abilities in using the tool. Thus, if analysts were using the tool in a real world situation, Unspecified Article would not be present in the tool since the analyst would be provided with the opportunity to probe the client/user for more details about the “article” mentioned. Was it a journal article or a newspaper article? Thus, it is not included in the ontology, only as an aid in the tool for coders.

In calculating the percent of agreement between coders, some adjustments were made preliminarily to the data. The data for Purpose and Search Task Result were thrown out since these were text fields allowing narrative data that would make assignments of agreement open to interpretation.

Although the instructions to the coders explicitly said that if an idea represented by a form item was not mentioned in the problem expression, then it should be left untouched, coders did interact with form items, possibly out of a need to be thorough. (This will be discussed in more detail later on.) The only instances where it can be safely assumed that thoroughness was the motivation, instead of a legitimate feeling that the

concept was mentioned, is in the case of the *Knowledge* question (Did the client feel that there was some knowledge or data s/he needed to perform the search task?) and the *Learning* question (Did the client take any trainings to assist him/her with performing the Search Task?). If a coder checked No, then this was converted from an N for No to a B for Blank.

Whenever form items were left untouched, they received a B for Blank rating. Otherwise they were assigned a Y for Yes or an N for No. See Appendix F for more information on how results were tallied.

Unlike Fleiss' example, coders for this research were allowed to select multiple categories for any subject, where categories are form items (ontology classes) and subjects are problem expressions. A total of 102 possible categories existed for each problem expression, making a total of 5100 (50 problem expressions X 102 possible categories) judgments made. Even if all three coders left a form item blank, thus yielding a BBB rating, this rating has meaning. It shows that all three coders agreed that the concept illustrated by the form item was not referenced in the problem expression. There were a total of 4525 judgments where all three coders agreed (mostly BBBs), either leading to a YYY, NNN, or BBB assignation. This means that 88.72% of the time all three coders agreed. There was a total of 564 judgments where only two coders agreed, either a YYB, YYN, YNN, YBB, NBB, or NNB: 11.05% of the time two coders agreed. And lastly, only 11 times (or .21%) did all three coders disagree, yielding a YNB rating.

The Kappa as calculated using Fleiss' (1971) formula<sup>6</sup> was .43 or 43%.

Unfortunately, this is not strong evidence that the agreement between coders is

---

<sup>6</sup> Dr. Michael Schlessinger, Professor Emeritus, of the Department of Mathematics at the University of North Carolina at Chapel Hill assisted the researcher in understanding Fleiss' work.

significant beyond chance. However, it must be noted that these calculations are affected by certain circumstances, namely that coders had no opportunity for probing. Coders had no way to follow-up with interviewees to ask them questions such as “Did you actually use a search engine or did you click through the site?” Questions such as these would have reduced the need to assume and make guesses. The tool could have been constructed in such a way as to reduce the need for assumption but it was felt that that would be cheating in a way. The tool was constructed as if it were to be used in a real world situation. So instead of having the *InformationCharacteristic:Existence* class modeled with the label Exists? and two radio buttons, one for Yes and the other for No, to model a real world situation where the analyst could take her cues off of the form item and explicitly ask the client, “Does the information actually exist?”, the researcher could have simply put a checkbox with a label “Information does not appear to exist.” This checkbox would have been more pertinent to the coders’ situation but not to a real world analyst because it assumes that the only reason you would interact with the form item is because you have clear indication that the information did/does not exist. There would be no opportunity for representing the existence of information with this type of form item. Therefore in a real world situation, an analyst would be unable to make such determinations; but the coder would be more assured that she should leave the form item untouched if existence was not mentioned explicitly and agreement would go up.

The Exists form item is a good example of how some coders attempted to be thorough. If the essence of the problem expressed by the interviewee was that the information they desired was not even in existence so they couldn’t obtain it, then the No button should have been checked. Otherwise the form item should have been left

untouched because the existence of the information was incidental to the central problem; but some coders customarily checked Yes. A case in point: “And then with the journal articles, there’s always the issue of timely access and using interlibrary loan.” For this problem expression, two coders checked Yes for Exists but this is not really mentioned in the problem expression. There is no real way to be certain if all the articles sought by the interviewee actually existed or not. The main thrust of the expression is the timeliness, or implied slowness, of inter-library loan. Existence is incidental. Therefore, according to the instructions, Exists should have been left untouched.

Another potential confusion for coders is that problem expressions ranged over many different levels of specificity and time span. Some interviewees spoke of general frustrations (as in the example above) and some spoke of specific incidents. Some interviewees spoke of incidents in the past and some spoke of ongoing incidents or recurring incidents. If a problem expression spoke of how costly a certain article was, it may not have included information as to whether or not the article was deemed worthy of the cost and so purchased. Because the researcher did not probe the interviewees for further information, coders would have to make assumptions about Worth, for example, with little supporting evidence.

It was discovered that some concepts overlap too much to make a black and white judgment about them. In particular, the *Container* class as represented by Location proved to be a repeated source of disagreement. In reviewing the instructions, the researcher assumed this would be self-evident to graduate students and professionals with Masters’ in Information and Library Science and therefore did not include the rigid definitions found in the data dictionary. This was a mistake since the possibility of

overlap is so great no matter what the level of expertise. For instance, some people would qualify NCLive as a *Database*, others as a *Digital Library*. Other people would say that it is both. Some would say that it is a *Website*; and all would be right. To complicate matters further, sometimes people (clients/users who do not work in technology in particular) are not positive what the resource actually is. How do you know if there is a database driving the website you are perusing? So even if the analyst had the data dictionary before him when talking with clients, there is no way for either party to be completely certain that a resource falls under only *Website* and not *Database* or *Digital Library* without being able to ask the developer of the resource. Therefore, it seems that the most important information to be gleaned from *Container* is whether it is an electronic/virtual or tangible resource (because the difference between the two translates into a different sets of challenges); and this should guide the constitution of the *Container* class. Perhaps it should be broken down into only two subclasses: VirtualContainer (containing virtual information) or TangibleContainer (containing information you can touch). VirtualContainer could contain individuals such as *Website*, *Database*, etc.; and Tangible Container could contain *Library* and print versions of *Book*, *Journal*, etc.

Sometimes there was confusion between Available and Exists. One interviewee said, “the data is not there” which caused the researcher to select No for Exists and leave Available blank since availability is not discussed. However, one coder selected Yes for Exists and No for Available. This means that the line between the two concepts is fuzzy. Perhaps this concept needs further clarification in the data dictionary and the tool. Additionally, under Characteristics of the Search Task, coders were often confused by the distinction between Difficulty Level and Level of Involvement. First of all, it is

clear that Difficulty Level can be removed. All the tasks are difficult in some way or they wouldn't have been offered as an answer to the interview question. It is a moot point. Secondly, Level of Involvement is not representing the exact nuance of meaning the researcher was aiming for. *LevelOfInvolvement* is closely tied with *NumberOfSteps*. It is supposed to represent the idea that a task can be tedious if the actor has to complete what s/he perceives to be too many steps in order to achieve the end goal. At its core is the idea of simplicity or lack thereof. This was not made completely clear in the tool, nor in its instructions. It is a difficult concept to articulate in a systematized, objective way. Simplicity is often seen as an antonym for difficulty even though this is not how this class is intended to work.

The same overlap in meaning is found in “difficult” and “frustrating.” Some coders followed the instructions that stressed not assuming an answer for Difficulty Level if the word “difficult” or “hard” or “challenging” or some other clear-cut synonym was used. Some felt that “frustrating” was an equal synonym; and given the interview question, it was. This produced disagreement.

Some interesting interpretations were made by coders that produced disagreement but could be said to be sound judgments. One coder assumed that if *Knowledge* was needed, then skills were required as well (Required Skill Level). One coder decided that if an interviewee said that a website was difficult to use, an Outside Organization was involved and that organization was being Uncooperative. This is not unreasonable. The makers of the website are an outside organization; and because they are perceived as not making their website usable, they are, in a way, being uncooperative. It was also assumed

that if the interviewee needed to know if a website existed that would satisfy their query, they also needed to know the website's URL; again, not a bad assumption!

However, there were some instances where the researcher could not understand why the expression was interpreted the way it was. For instance, this particular problem expression was obtained via an online survey: "Sales tax revenue, gross and net sales, employment, data etc. in a timely manner." One coder decided that the Available Format was Electronic, the Location was a Computer, and the Urgency Level was Medium. None of these are represented in the problem expression. The interviewee could have easily taken a trip to the library to seek this information. The coder made a note to the effect that she assumed these things because this problem expression was gathered using an online survey. This assumption is stretching things quite a bit and is probably a result of not being able to probe the interviewee and wanting to be thorough.

### **Coder Reactions and Completeness**

Each coder was emailed a simple questionnaire (See Appendix E) upon completion of assigned codings. All five coders agreed that the tool was somewhat difficult to use. Potentially this could be improved in a real world situation where opportunities for probing would reduce the ambiguity of making assumptions.

Four of the five said that they did not encounter problem expression features that were not represented on the tool. The one coder who did said she did so between 3 and 4 times. Therefore, the majority of experience with the tool attests to its completeness.

Two coders commented that the tool seemed daunting at first; but once the learning curve was surmounted, it worked well. One coder felt that some of the

expressions didn't fit precisely into the tool due to their generality. Instead of focusing on a definitive search task anchored in a specific point in time, indiscriminate statements about problems produced difficulty. An example of this might be: "It often times will involve my being able to connect with the right people." The essence of this expression is best modeled in *ContainerCharacteristic:Pertinence*, where the "right person" is acting as a *Container* of information. However, the choices offered by the Pertinence form item assume a specific task rather than a general frustration. The correct choice on the tool would be Incorrect location to find desired information; but this assumes that the seeker completed a specific task and was unable to find the right person. This particular problem expression is more general in nature. It is possible that the interviewee sometimes finds the right person sometimes not. This could be resolved with a follow-up question to the interviewee, something to the effect of, "So there are times when you are unable to find the right person?"

Another coder commented on the idiosyncrasies of the *Container* subclasses, as was discussed above. This particular coder felt uncomfortable when problem expressions did not reference form items at all and should therefore be left untouched. Another coder agreed with her by expressing a desire for a "Not Applicable" option. If used in real world situation, this discomfort would potentially subside because the analyst could use the form items as cues to guide the discussion, thereby ensuring that all form items (classes) were addressed.



### **Follow Up Interviews and Accuracy**

The accuracy or validity of the tool was evaluated by re-interviewing 35% of the original interviewees to ask for their assessment of the researcher's ability to completely understand their problem as a result of structuring the problem expression into the tool. Fourteen out the 15 people that were followed-up with gave the researcher a 3 – "Your assessment of what I said was completely accurate. You nailed it!" One person assigned the researcher a 2 because she felt the researcher placed too much emphasis on one facet of the problem – how she communicated with the library about the potential acquisition of some resources. The interviewee said she only asked a question of the library whereas the researcher implied that a something more akin to a conversation took place. Thus, complete accuracy/validity (a 3 rating) was achieved 93.3% of the time, in terms of interviewees, and 87.5%, in terms of problem expressions analyzed.

It must be admitted that these extremely positive results may be attributed to the cordiality of the people with whom the follow-ups were made. Perhaps they rated the researcher so high because they are nice people who wanted the researcher to feel successful in her completion of her Master's work or maybe they would have felt uncomfortable offering a lower rating. These points must be conceded.

### **Further Research Needed**

Because coders' were not presented with an opportunity to probe interviewees and because this was probably the greatest factor in achieving an unsatisfactory kappa, further research is needed. Specifically, it would be beneficial to replicate the calculation

of codings conducted in a more authentic analyst-client scenario. The experimental setting could include two analysts interviewing a single client about his information seeking difficulties. Each analyst could use the note-taking tool as a guide to help frame the interview, using form items as conversational cues and probes. The interview could be inputted into the tool in real time; and the resulting kappa would prove to be more representative of a faithful scenario, and therefore have greater integrity of meaning.

As was mentioned earlier, the *Container* class needs to be reevaluated.

*DifficultyLevel* can probably be stripped from the ontology. *LevelOfInvolvement* needs greater clarification; and *Knowledge* and *SkillLevel* should also be reconsidered. The role of *OutsideOrganization* should be cleaned up.

It would also be prudent to attempt different adaptations of the ontology and tool for different domains. This would provide the best check on Gruber's (1995) call for extendibility where new concepts can be added without extensive restructuring of the ontology.

And finally, the issue of focusing solely on the "deficit subject," which sometimes leads to limited information gathering, has not been fully addressed by this ontology. It is, in fact, geared towards extracting the deficit subject. Hence, it should be noted that the note-taking tool should be used in conjunction with other tools which help illuminate other, non-negative aspects of the situation under examination. Perhaps further investigation would yield solutions for the inclusion of new classes which could help extract other nuances. If this ontology were, in some way, combined with the ontologies created by Uschold et al. and Jin et al., this might help to augment its ability to provide a

more comprehensive, less deficit-focused picture of the needs of the client and his organization.

## **CONCLUSION**

The ontology-based note-taking tool has been created to help frame client interviews by systematically accounting for all features possible in a discussion of information-seeking problems; and as such, it provides measures for reducing communication problems. For instance, analysts may allow the client to narrate freely with little interruption because the tool provides a structured set of cues and probes possible. Because the analyst does not have to interject lest he forget an important point, the narrative frame that clients naturally ascribe to can be allowed its due. The analyst can safely rely on the tool to remind him of key points to be covered; and when there is a natural pause in the flow of narrative, he can refer to the tool for guidance on probing questions he might now ask.

Problems with tacit knowledge and memory may be lessened. The pre-defined set of features provides structure to an otherwise haphazard, unpredictable conversation. These features, as represented by form items, may help jar the client's memory, helping him to recall tidbits previously forgotten or bring hidden tidbits to the surface. When new concepts are discovered, the extendibility of the ontology will easily allow their inclusion.

Additionally, the tool is grounded in experience, not future desires and possibilities, thereby limiting discussions to the here and now. It explores social context and culture because of its focus on the perceptions of the client. Because the tool is based

on an ontology and can therefore be expressed logically, it can act as a knowledge base or database, making the collection and storage of information less overwhelming as with ordinary qualitative measures. It can also be more easily translated into programmatic specifications and graphic models to suit the needs of developers.

Finally, this study has shown that an ontology-based note-taking tool does aid a systems analyst in her ability to match her perceptions of the problem space with her clients' perceptions. Although, at this time, the reliability of the tool requires further investigation, the raw numbers suggest that consistency of use is possible. Furthermore, the tool is considered by most coders to provide complete coverage of concepts; and should new concepts arise, their inclusion can only improve the tool. The information seeking ontology has potential but requires more work. It is the hope of the researcher that this line of research will continue so that clients/users may be the benefactors of processes, solutions, and systems of higher integrity and quality.

## REFERENCES

- Agarwal, R. & Tanniru, M.R. (1990 Summer). Knowledge acquisition using structured interviewing: An empirical investigation. *Journal of Management Information Systems*, 7(1), 123-140.
- Alcázar, E.G. & Monzón, A. (2000 June). A process framework for requirements analysis and specification. In Proceedings of the 4<sup>th</sup> International Conference on Requirements Engineering, by the IEEE Computer Society.
- Al-Rawas, A. & Easterbrook, S. (1996 February). Communication problems in requirements engineering: A field study. In Proceedings of the 1<sup>st</sup> Westminster Conference on Professional Awareness in Software Engineering, Royal Society, London, England.
- Alvarez, R. (2001 June). There's more to the story: Using narrative analysis to examine requirements engineering as a social process. Paper presented at the 7<sup>th</sup> International Workshop on Requirements Engineering, Interlaken, Switzerland.
- Alvarez, R. (2002 January). Discourse analysis of requirements and knowledge elicitation interviews. In Proceedings of the 35<sup>th</sup> Hawaii International Conference on System Sciences (HICSS'02), IEEE Computer Society.
- Alvarez, R. & Urla, J. (2002 Winter). Tell me a good story: Using narrative analysis to examine information requirements interviews during an ERP implementation. *The DATA BASE for Advances in Information Systems*, 33(1), 38-52.
- Antón, A.I. & Earp, J.B. (2004 August). A requirements taxonomy for reducing Web site privacy vulnerabilities. *Requirements Engineering*, 9(3), 169-185.
- Barbosa, S.D.J., Silveira, M.S., Greco de Paula, M.G., & Breitman, K.K. (2005 June) Supporting a shared understanding of communication-oriented concerns in human-computer interaction: A lexicon-based approach. *Lecture Notes in Computer Science*, 3425, 271-288.
- Byrd, T.A., Cossick, K.L., & Zmud, R.W. (1992 March). A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS Quarterly*, 16(1), 117-138.

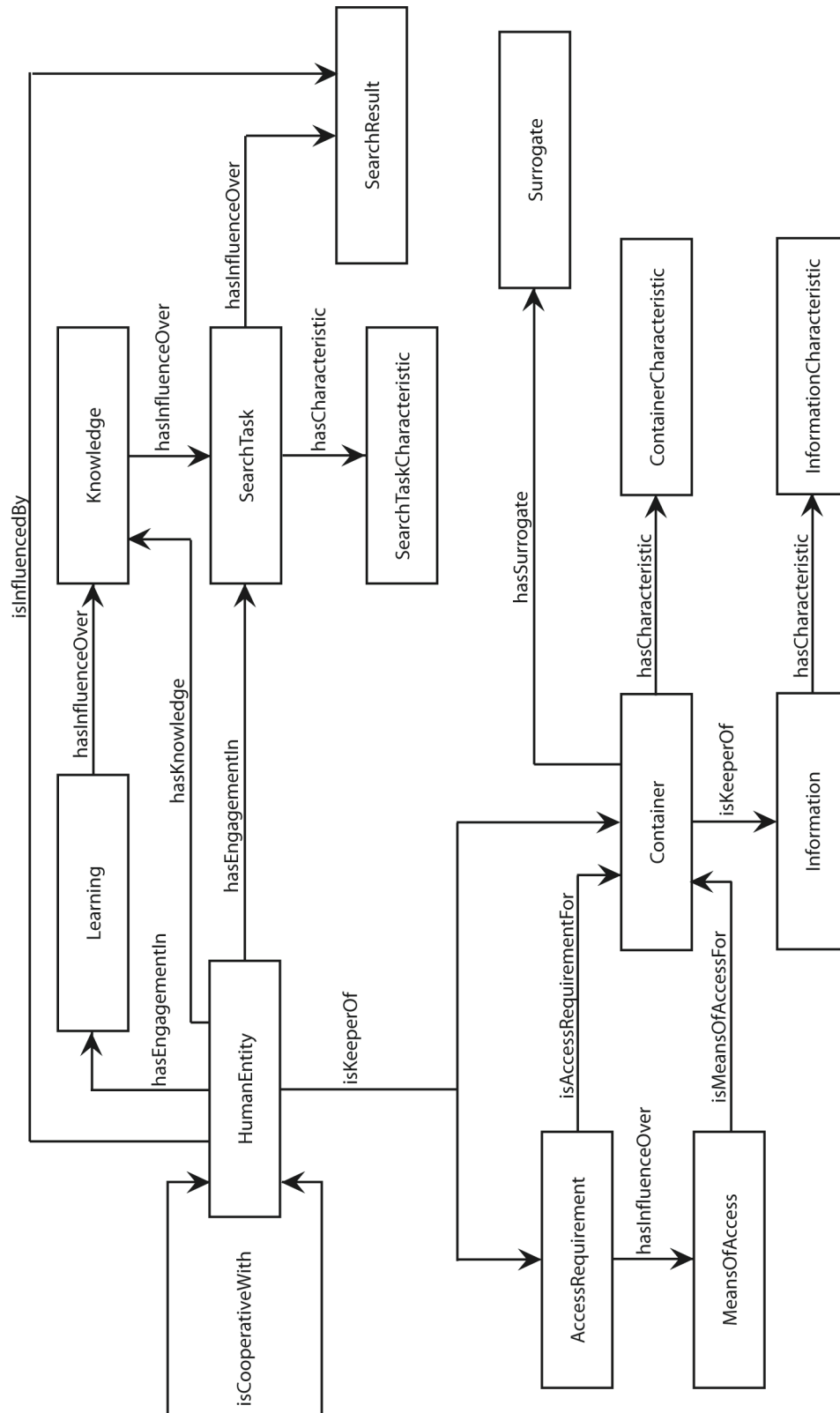
- Carreño, R.S., Breis, J.F., Béjar, R.M., & Gómez, P.C. (2000). An ontology-based approach to knowledge acquisition from text. *Cuadernos de Filología Inglesa*, 9(1), 191-212.
- Clegg, C., Axtell, C., Damodaran, L., Farbey, B., Hull, R., Lloyd-Jones, R., Nicholls, J., Sell, R., & Tomlinson, C. (1997 September). Information technology: A study of performance and the role of human and organizational factors. *Ergonomics*, 40(9), 851-871.
- Coughlan, J. & Macredie, R.D. (2002 June). Effective communication in requirements elicitation: A comparison of methodologies. *Requirements Engineering*, 7(2), 47-60.
- Curtis, B., Krasner, H., & Iscoe, N. (1988 November). A field study of the software design process for large systems. *Communications of the ACM*, 31(11), 1268-1287.
- Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin*, 76(5), 378-382.
- Flynn, D.J. & Jazi, M.D. (1998 January). Constructing user requirements: A social process for a social context. *Information Systems Journal*, 8(1), 53-63.
- Galliers, R.D. & Swan, J.A., (2000 September). There's more to information systems development than structured approaches: Information requirements analysis as a socially mediated process. *Requirements Engineering*, 5(2), 74-82.
- Gerson, E.M. & Star, S.L. (1986 July). Analyzing due process in the work place. *ACM Transactions on Office Information Systems*, 4(3), 257-270.
- Goguen, J.A. (1993). Social issues in requirements engineering. In Proceedings of the IEEE International Symposium of Requirements Engineering, by the IEEE Computer Society, 194-195.
- Goguen, J.A. & Linde, C. (1993 January). Techniques for requirements elicitation. In Proceedings of IEEE International Symposium on Requirements Engineering, by the IEEE Computer Society, 152-164.
- Gruber, T.R. (1995 November). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5/6), 907-928.
- Guarino, N. (1997 February) Understanding, building, and using ontologies. *International Journal of Human-Computer Studies*, 46(2/3), 293-310.

- Guarino, N. & Welty, C.A. (2004). An overview of OntoClean. In S.Staab & R.Studer, (Eds.), *Handbook on Ontologies*, 151-171. Berlin, Heidelberg, New York: Springer-Verlag.
- Hands, K., Peiris, D., & Gregor, P. (2004 January). Development of a computer-based interviewing tool to enhance the requirements gathering process. *Requirements Engineering*, 9(3), 204-216.
- Heflin, J. ed. (2004 February 10) OWL web ontology language use cases and requirements. Retrieved September 10, 2005 from the W3C website: <http://www.w3.org/TR/webont-req/>.
- Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004 Augus). A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools, Edition 1.0. Retrieved September 16, 2005 from the Protégé website: <http://protege.stanford.edu/doc/users.html>.
- Jenkins, A.M., Naumann, J.D., & Wetherbe, J.C. (1984). Empirical investigation of systems development practices and results. *Information & Management*, 1, 73-82.
- Jin, Z., Bell, D.A., Wilkie, F.G., & Leahy, D.G. (2003). Automated requirements elicitation: Combining a model-driven approach with concept reuse. *International Journal of Software Engineering and Knowledge Engineering*, 13(1), 53-82.
- Jiang, D., Zhang, S. Wang, Y. (2005). Towards a formalized ontology-based requirements model. *Journal of Shanghai Jiaotong University (Science)*, 10(1), 34-39.
- Kaindl, H. (1993 April). The missing link in requirements engineering. *ACM SIGSOFT Software Engineering Notes*, 18(2), 30-39.
- Kelsey, R.B. (1996 March). Bad fixes, change specifications, and linguistic constraints on problem diagnosis. *ACM SIGSOFT Software Engineering Notes*, 21(2), 74-78.
- Kilker, J. & Gay, G. (1998 June). The social construction of a digital library: A case study examining implications for evaluation. *Information Technology and Libraries*, 17(2), 60-70.
- Marakas, G.M. & Elam, J.J. (1998 March). Semantic structuring in analyst acquisition and representation of facts in requirements analysis. *Information Systems Research*, 9(1), 37-63.
- Mizoguchi, R. & Ikeda, M. (1997) Towards ontology engineering. In Proceedings of the Joint 1997 Pacific Asian Conference on Expert Systems / Singapore International Conference on Intelligent Systems, 259-266.

- Noy, N. & McGuinness, D.L. Ontology development 101: A guide to creating your first ontology. Retrieved October 11, 2004 from <http://protege.stanford.edu/publications/ontology-development/ontology101-noy-mcguinness.html>.
- Polanyi, M. (1966). *The Tacit Dimension*. New York: Anchor Books.
- Pomerantz, J. (2005). A Linguistic Analysis of Question Taxonomies. *Journal of the American Society for Information Science and Technology*, 56(7), 715-728.
- Robinson, M. & Bannon, L. (1991 September). Questioning representations. In Proceedings of ECSCW, Amsterdam. Retrieved on September 11, 2005 from <http://www.ul.ie/~idc/library/papersreports/LiamBannon/15/QuestFin.html>.
- Sowa, J.F. (2000). *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove: Brooks/Cole.
- Sutton, D. (2000 September). Linguistic problems with requirements and knowledge elicitation. *Requirements Engineering*, 5(2), 114-124.
- Travers, D.A. & Haas, S.W. (2003 August). Using nurses' natural language entries to build a concept-oriented terminology for patients' chief complaints in the emergency department. *Journal of Biomedical Informatics*, 36(4), 260-260.
- Uschold, M. & Gruninger, M. (1996). Ontologies: Principles, methods, and applications. Technical Report AIAI-TR-191, Artificial Intelligence Application Institute, University of Ediburgh Ireland. Retrieved August 21, 2005 from <http://www.aiai.ed.ac.uk/project/enterprise/ontology.html>.
- Uschold, M., King, M., Moralee, S., & Zorgios, Y. (1998 March). The Enterprise Ontology. *Knowledge Engineering Review*, 13(1), 31-89.
- Westrup, C. (1999 January). Knowledge, legitimacy and progress? Requirements as inscriptions in information systems development. *Information Systems Journal*, 9(1), 35-54.



## APPENDIX A: UML DIAGRAM



## APPENDIX B: DATA DICTIONARY

### DOMAIN

This ontology applies to the domain of information seeking tasks and the difficulties experienced in trying to locate information.

### SCOPE

The ontology largely focuses on how the information seeker *perceives* the search task. It touches on various aspects of the information seeking task such as the task behavior and characteristics, actors/agents, cultural environment, characteristics of the information sought and end results, artifacts and their characteristics, avenues of access, requirements and barriers to achieving the end goal, and aids to the search task (such as knowledge, learning, and surrogates).

### CLASSES

Please note:

- All subclasses inherit properties from their superclass. Any properties a subclass holds, other than those that are inherited from the superclass, will be noted.
- Almost any class could be expanded to include more subclasses or to have more instances/individuals. For instance the SingleIndividual subclass of HumanEntity could be expanded to include Friend or FamilyMember.
- Any classes that would not have subclasses but only instances/individuals (i.e. – terminal classes) are noted with an asterisk (\*) and possibly examples of instances/individuals.
- Defined instances/individuals are shown in curly braces {}.
- OntoClean Key:
  - Rigidity:
    - Rigid: +R
    - Semi-Rigid: -R
    - Anti-Rigid: ~R
  - Identity:
    - Identity: +I
    - No Identity: -I
    - Anti-Identity: ~I
  - Unity:
    - Unity: +U
    - No Unity: -U
    - Anti-Unity: ~U

### AccessRequirement (-R-I-U)

**DEFINITION:** A requirement placed on a Container which restricts the Container's access. Not all Containers have AccessRequirements. AccessRequirements must be kept/owned by some governing HumanEntity but a given owning HumanEntity may not be an instance/individual in this ontology.

#### NECESSARY & SUFFICIENT PROPERTIES:

- ∃ hasInfluenceOver MeansOfAccess
- ∃ hasKeeper HumanEntity
- ∃ isAccessRequirementFor Container

#### NECESSARY PROPERTIES:

owl:Thing

**SUBCLASSES:**

- **Expenditure (-R-I-U)**  
*DEFINITION:* A monetary payment, ultimately made to the HumanEntity that controls the Container; however the HumanEntity making the payment may never know exactly who will be receiving the payment (especially if the Container is a website that allows for online payments).
  - **DepartmentalExpense (-R-I-U) \***  
*DEFINITION:* A monetary payment the WorkDepartment makes.
  - **PersonalExpense (-R-I-U) \***  
*DEFINITION:* A monetary payment a SingleIndividual makes.
- **Membership (-R-I-U)**  
*DEFINITION:* A means of access whereby the Container collects information about the HumanEntity wishing to gain access; sometimes requires payment (see Expenditure).
  - **LibraryCard (+R-I-U) \***  
*DEFINITION:* A means of access specifically for a Library.  
 Possible Instances: a specific library card number
  - **Registration (+R-I-U) \***  
*SYNONYM:* Subscription  
*DEFINITION:* A means of access that may require an Expenditure and, if electronic, will definitely require a user name and password.
- **Permission (+R-I-U)**  
*DEFINITION:* A means of access whereby the HumanEntity wishing to access the Container must obtain permission from another HumanEntity to do so. In the case of FirewallAdmittance, the requesting HumanEntity may never know who the HumanEntity is that controls the firewall and therefore grants or denies passage.
  - **ConfidentialityWaiver (+R-I-U) \***  
*DEFINITION:* A document (required by the HumanEntity that controls the Container) that must be formally signed by the HumanEntity wishing to use the Container. The signature indicates the using HumanEntity's agreement to keep the Information contained in the Container private.
  - **FirewallAdmittance (+R-I-U) \***  
*DEFINITION:* A system construct placed on a Computer or Server to provide security to HumanEntities using said system. Whereas other Permissions are one-sided in that they are required to *receive* Information in the Container, FirewallAdmittance includes *getting out past* a Container, such as a Computer, to get to Information.
  - **ProprietaryAgreement (+R-I-U) \***  
*DEFINITION:* A document that requires the using HumanEntity to sign in confirmation of rights that were negotiated with another HumanEntity who owns the ProprietaryRights to Information.

## Container (~R-I-U)

**DEFINITION:** A physical or virtual object that contains Information. All Containers are considered to keep Information of some kind in this ontology. All Containers have one or more MeansOfAccess. All Containers must be kept/owned by some governing HumanEntity but a given owning HumanEntity may not be an instance/individual in this ontology. *Note: The isKeeperOf property between subclasses is fluid. In other words, it is possible that a Computer could be the keeper of other Containers in addition to a Website.*

**REGARDING ONTOCLEAN ASSIGNMENT:** Because Containers are interchangeable and can contain other Containers, it would be incorrect to assume that any one instance must appear in only one subclass; thus ~R. Additionally, instances would be distinguishable in some but not all cases; thus -I. For instance, one Container could exist in multiple Containers as well as be divided into multiple parts (such as Journal articles and Book chapters). Because any given Container can contain multiple containers or be housed in multiple Containers or have its parts housed separately, -U is required.

### NECESSARY & SUFFICIENT PROPERTIES:

- ∃ hasKeeper HumanEntity
- ∃ hasMeansOfAccess MeansOfAccess
- hasMeansOfAccess ≥ 1
- ∃ isKeeperOf Information

### NECESSARY PROPERTIES:

- owl:Thing
- ∃ hasAccessRequirement AccessRequirement
- ∀ hasCharacteristic ContainerCharacteristic
- hasCharacteristic ≥ 0
- ∃ hasSurrogate Surrogate
- hasSurrogate ≥ 0

### SUBCLASSES:

- **Archive (~R-I-U) \***

*DEFINITION:* “Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control; permanent records.” (Retrieved from The Society for American Archivists Glossary: [http://www.archivists.org/glossary/term\\_details.asp?DefinitionKey=156](http://www.archivists.org/glossary/term_details.asp?DefinitionKey=156) on 10/9/05.)

#### NECESSARY PROPERTIES:

- ∃ isKeeperOf Newspaper

- **Book (~R-I-U) \***

*DEFINITION:* A type of Container that contains prose or narrative Information. All content in a Book is centered on one topic.

*POSSIBLE INDIVIDUALS:* a book’s title like “Harry Potter and the Goblet of Fire”

*REGARDING ONTOCLEAN ASSIGNMENT:* The ~R is self-evident. A Book could be in a Digital Library. The -I would account for instances where you might have a full book in a Digital Library but only a chapter of the same book in another Container. The -U also alludes to the same situation mentioned in support of -I.

- **Computer** ( $\sim R-I-U$ ) \*

*DEFINITION:* A type of Container that contains virtual Information in a variety of forms.

*NECESSARY PROPERTIES:*

$\exists$  isKeeperOf Website

*REGARDING ONTOCLEAN ASSIGNMENT:* The  $\sim R$  and  $-I$  are fairly obvious in that a Computer could be considered a Server. The  $-U$  would indicate a situation where a Computer might have a Digital Library housed on it as well as a Database or some other Container.

- **Database** ( $\sim R-I-U$ ) \*

*DEFINITION:* A type of Container that provides Information in a structured way that takes advantage of relationships between items in the Database.

*NECESSARY PROPERTIES:*

$\exists$  isKeeperOf Newspaper

$\exists$  isKeeperOf Journal

- **DigitalLibrary** ( $\sim R-I-U$ ) \*

*DEFINITION:* A type of Container, and sometimes a type of Database, that provides Information in a structured way that usually includes some type of user-services function and search capabilities.

*NECESSARY PROPERTIES:*

$\exists$  isKeeperOf Newspaper

$\exists$  isKeeperOf Book

$\exists$  isKeeperOf Journal

- **Internet** ( $\sim R-I-U$ ) \*

*DEFINITION:* A type of Container that can contain a variety of virtual Information and is most often a Container of Containers.

*NECESSARY PROPERTIES:*

$\exists$  isKeeperOf Website

$\exists$  isKeeperOf Archive

$\exists$  isKeeperOf Digital Library

$\exists$  isKeeperOf Database

- **Journal** ( $\sim R-I-U$ ) \*

*DEFINITION:* A type of Container, similar to a Book, yet different from a Book in that it has an attribute of seriality that a Book may or may not have. Also, Journals may contain articles (content) that do not agree in theme/topic. These articles are almost always research-based. The authors vary and are rarely employed by the company that owns the Journal.

- **Library** ( $\sim R-I-U$ ) \*

*DEFINITION:* A type of Container that has physical structure in the sense that HumanEntities may actually enter and dwell in it. The Library differs significantly from other containers in that it is the Container most situated to be a Container of all other Containers. For the purposes of this ontology, I have decided that a Library, at a minimum, has Newspapers, Books, Databases (at least for the catalog), and Journals.

*NECESSARY & SUFFICIENT PROPERTIES:*

- ⊃ isKeeperOf Book
- ⊃ isKeeperOf Database
- ⊃ isKeeperOf Journal
- ⊃ isKeeperOf Newspaper

*NECESSARY PROPERTIES:*

- ⊃ isKeeperOf DigitalLibrary
- ⊃ isKeeperOf Archive
- ⊃ isKeeperOf VideoTape
- ⊃ isKeeperOf Database

- **Newspaper** (~R-I-U) \*

*DEFINITION:* A type of Container, like a Journal, whose goal is to provide serial Information; yet unlike a Journal, provides new Information on a daily basis. Its focus is to provide Information on current events or news. It may contain some research-oriented articles; but for the most part does not and simply reports current facts. The authors of a Newspaper are typically a group of in-residence employees employed by the company owning the Newspaper.

- **Server** (~R-I-U) \*

*DEFINITION:* A type of Container, like a Computer, but with substantially more processing power.

## Necessary Properties:

- ⊃ isKeeperOf Website
- ⊃ isKeeperOf Archive
- ⊃ isKeeperOf Digital Library
- ⊃ isKeeperOf Database

- **VideoTape** (~R-I-U) \*

*DEFINITION:* A type of Container that contains moving image and audio Information.

- **Website** (~R-I-U) \*

*DEFINITION:* A type of Container that can contain a variety of virtual Information. A DigitalLibrary, for example, may be seen as a Website.

**ContainerCharacteristic** (+R-I-U)

**DEFINITION:** Those attributes of a Container as *perceived* by a HumanEntity – usually the FirstPerson. It is possible that a FirstPerson may not verbalize any characteristics of the Container he searches.

**REGARDING ONTOCLEAN ASSIGNMENT:** Because a finite set individuals have been defined for each subclass, implying a limitation on the individuals one may designate for a subclass, the assignment of +R (although very strong) has been made. This means that a person using this ontology must choose from one of the available instances if needing to model the class to which the instance belongs. Perhaps this ontology could be revised so that instances change. At the time of revision, the +R would have to be re-evaluated. Because the instances within a subclass are subject to the value judgment of the FirstPerson, they are not always distinguishable. Whereas one FirstPerson may describe the a given Container as SparseOrLimited, another FirstPerson may describe the exact same Container as Satisfactory. Because any given ContainerCharacteristic can be combined with another to formulate a more detailed picture of the Container, none of the ContainerCharacteristics can be said to be entities unto themselves; thus –U.

**NECESSARY & SUFFICIENT PROPERTIES:**

$\forall$  isCharacteristicOf Container

**NECESSARY PROPERTIES:**

owl:Thing

**SUBCLASSES:**

- **Address (+R-I-U) \***

*DEFINITION:* A determination of if the address of the Container is perceived as being static or not; specifically, answering the question: did the Container change location and become difficult to find?

*INDIVIDUALS:* {Static Changing}

- **AmountOfInformation (+R-I-U) \***

*DEFINITION:* The perceived amount of information the Container contains.

*INDIVIDUALS:* {SparseOrLimited Satisfactory Overwhelming}

- **InterfaceUsability (+R-I-U) \***

*DEFINITION:* The level of ease in using the Container. This can include ease of navigating the Container.

*INDIVIDUALS:* {EasyToUse DifficultToUse}

*NECESSARY PROPERTIES:*

$\exists$  isInfluencedBy Organization

- **Organization (+R-I-U) \***

*DEFINITION:* The quality how organized the information in the Container is.

*INDIVIDUALS:* {Organized Disorganized}

*NECESSARY PROPERTIES:*

$\exists$  hasInfluenceOver InterfaceUsability

- **Pertinence (+R-I-U) \***

*DEFINITION:* The HumanEntity's determination of relevancy for the Container; specifically, an assessment of whether the Container satisfies the HumanEntity's query.

*INDIVIDUALS:* {CorrectContainerToSearch IncorrectContainerToSearch}

- **Vitality (+R-I-U) \***

*DEFINITION:* The Container's state of activity; specifically whether the Container is "live" or "up" as opposed to "down" and inaccessible.

*INDIVIDUALS:* {Active Inactive}

**HumanEntity (+R-I+U)**

**DEFINITION:** One or more human beings (person/people).

**REGARDING ONTOCLEAN ASSIGNMENT:** For a person or group of people to exist in this ontology, they must be in one or more of the following subclasses. Individuals are not always distinguishable in the sense that sometimes a FirstPerson, for instance, in recounting a search task performed in the past, could mention another SingleIndividual as a WorkColleague who might later be mentioned as a

WorkSupervisor by a completely different FirstPerson. Similarly, the people who comprise a WorkDepartment may also comprise a WorkOrganization; thus making the two entities only partially distinguishable. Although one might argue that this also implies that a WorkDepartment, or for that matter, ProfessionalsInField, is not a clearly demarcated whole; for the purposes of this ontology they are complete entities unto themselves. WorkDepartments are made up of SingleIndividuals that have a common purpose. This common purpose delineates them as a whole. ProfessionalsInField share a common focus; and this focus forms the boundary around this group making it a whole.

#### NECESSARY PROPERTIES:

owl:Thing  
 $\exists$  hasEngagementIn SearchTask  
 $\exists$  hasEngagementIn Learning  
 $\forall$  hasKnowledge Knowledge  
hasKnowledge  $\geq 0$   
 $\exists$  isCooperativeWith HumanEntity  
isCooperativeWith  $\geq 0$   
 $\exists$  isKeeperOf AccessRequirement  
 $\exists$  isKeeperOf Container

#### SUBCLASSES:

- **GroupOfIndividuals** (-R-I-U)  
*DEFINITION:* A group of people.

##### NECESSARY PROPERTIES:

$\exists$  hasInfluenceOver SingleIndividual

##### SUBCLASSES:

- **WorkDepartment** (-R-I-U) \*  
*SYNONYMS:* WorkDivision, WorkOffice  
*DEFINITION:* A group of people who form a unit within a WorkOrganization. Specifically, a FirstPerson's home department within his/her place of employment.  
  
*POSSIBLE INDIVIDUALS:* Division of Medical Assistance in the Department of Health and Human Services
- **WorkOrganization** (-R-I-U) \*  
*SYNONYMS:* Business, Company, WorkInstitution  
*DEFINITION:* A group of people who form a unit geared towards satisfying a common work mission; can be further divided into WorkDepartments. Specifically, a FirstPerson's place of employment.  
  
*POSSIBLE INDIVIDUALS:* the company's name such as SAS, The University of North Carolina at Chapel Hill
- **OutsideOrganization** (-R-I-U) \*  
*DEFINITION:* A group of people who form a unit outside the WorkOrganization. The FirstPerson will not be a part of this group but may have business or conduct collaborative activities with this group.
- **ProfessionalsInField** (-R-I-U) \*  
*DEFINITION:* A group of people who hold a professional focus in common. They could also constitute members of the WorkOrganization, OutsideOrganization, friends, family, etc.



- **SingleIndividual (+R-I+U)**  
*DEFINITION:* A single person.

*NECESSARY PROPERTIES:*

$\exists$  isInfluencedBy GroupOfIndividuals

*SUBCLASSES:*

- **FirstPerson (+R-I+U) \***  
*DEFINITION:* The specific client/user sharing his/her search experiences which this ontology is meant to capture.  
  
*POSSIBLE INDIVIDUALS:* the person's name such as Laura Christopherson
- **Other Professional (+R-I+U) \***  
*DEFINITION:* Another professional that the FirstPerson interacts with. This person does not work for the same WorkOrganization as the FirstPerson and therefore cannot be a WorkColleague, WorkSubordinate, or WorkSupervisor.
- **WorkColleague (+R-I+U) \***  
*DEFINITION:* The working colleague of the specific client/user sharing his/her search experiences which this ontology is meant to capture.
- **WorkSubordinate (+R-I+U) \***  
*DEFINITION:* The working employee of the specific client/user sharing his/her search experiences which this ontology is meant to capture.
- **WorkSupervisor (+R-I+U) \***  
*DEFINITION:* The working supervisor of the specific client/user sharing his/her search experiences which this ontology is meant to capture.

## **Information (-R-I-U) \***

**DEFINITION:** Any atomic datum or grouping of data that has meaning or is intended to impart meaning to a HumanEntity. It is the object of the SearchTask. Possible instances include: how many people aged 65 or older who are on Medicare in the state of North Carolina; a prose treatment on the nature of federal government's response to Hurricane Katrina; etc.

**NECESSARY PROPERTIES:**

owl:Thing

$\forall$  hasCharacteristic InformationCharacteristic

hasCharacteristic  $\geq 0$

## **InformationCharacteristic (+R-I-U)**

**DEFINITION:** Those attributes of a desired piece of Information as *perceived* by a HumanEntity – usually the FirstPerson. It is possible that a FirstPerson may not verbalize any characteristics of the information he/she seeks.

**NECESSARY & SUFFICIENT PROPERTIES:**

$\forall$  isCharacteristicOf Information

**NECESSARY PROPERTIES:**

owl:Thing

**SUBCLASSES:**

- **Availability** (+R-I-U) \*

*DEFINITION:* The quality of being available or unavailable to the HumanEntity.

*INDIVIDUALS:* {Available Unavailable}

*NECESSARY PROPERTIES:*

$\exists$  isInfluencedBy Existence

- **Centrality** (+R-I-U) \*

*DEFINITION:* The quality of how cohesive the Information is in terms of where it is located. This specifically refers to whether the FirstPerson may obtain an answer to his/her information question in one Container or whether s/he must search different Containers for pieces of his/her answer.

*INDIVIDUALS:* {LocatedInOnePlace ScatteredAcrossLocations}

- **Completeness** (+R-I-U) \*

*DEFINITION:* The quality of being complete or incomplete as perceived by the HumanEntity. For instance, if a Container provides only an Abstract and not the full-text article, a HumanEntity may judge this Information to be incomplete.

*INDIVIDUALS:* {Complete Incomplete}

- **Confidentiality** (+R-I-U) \*

*DEFINITION:* The quality that describes whether the Information contained in the Container is available to anyone or whether a select group of HumanEntities may view it. Specifically, this refers to the controlling HumanEntity's desire to keep this Information secret. Usually they are unwilling to share this Information.

*INDIVIDUALS:* {Confidential Public}

- **Cost** (+R-I-U) \*

*DEFINITION:* The monetary value of the Information; what is being charged to the HumanEntity so that he/she may purchase the Information.

*INDIVIDUALS:* {OneDollar TwoDollars [etc]}

*NECESSARY PROPERTIES:*

$\exists$  hasInfluenceOver Worth

- **Currency** (+R-I-U) \*

*DEFINITION:* The quality of being current and up-to-date or being old as perceived by the HumanEntity.

*INDIVIDUALS:* {Current OutOfDate}

*NECESSARY PROPERTIES:*

$\exists$  hasInfluenceOver Worth

- **Existence** (+R-I-U) \*

*DEFINITION:* The quality of being in existence somewhere, anywhere, regardless of availability. For instance, an answer to the question: does the Information exist, has it come into being?

*INDIVIDUALS:* {Exists Absent}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver Availability

- **Format (+R-I-U) \***

*DEFINITION:* Whether the Information is available electronically or in print or both. This characteristic can provide further dimension to Availability in that sometimes Availability can be perceived by the FirstPerson as available in a preferred format, such as electronic.

*INDIVIDUALS:* {Electronic Print}, {Electronic} U {Print}

- **Pliability (+R-I-U) \***

*DEFINITION:* The quality of how usable the Information is in its current form.

*INDIVIDUALS:* {UsableAsIs RequiresManipulation}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver Worth

- **ProprietaryRights (+R-I-U) \***

*DEFINITION:* The quality of being proprietary and requiring negotiated rights to access. This is different from Confidentiality in that the controlling HumanEntities are usually willing to share this Information but they wish to ensure some type of control over its use by the petitioning HumanEntity. The controlling HumanEntity continues to own the Information. Often, an expenditure is required on the part of the petitioning HumanEntity to gain usage of the Information.

*INDIVIDUALS:* {Proprietary NotProprietary}

- **Relevance (+R-I-U) \***

*DEFINITION:* The quality of how well the Information matches the HumanEntity's needs.

*INDIVIDUALS:* {NotRelevant SomewhatRelevant HighlyRelevant}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver Worth

- **Worth (+R-I-U) \***

*DEFINITION:* The measure of how valuable the Information is to the HumanEntity with respect to its Cost and some other InformationCharacteristics such as Relevance, Currency, and Pliability.

*INDIVIDUALS:* {WorthItExpensive NotWorthItExpensive WorthItInexpensive NotWorthItInexpensive}

*NECESSARY PROPERTIES:*

⊃ isInfluencedBy Relevance

⊃ isInfluencedBy Currency

⊃ isInfluencedBy Pliability

⊃ isInfluencedBy Cost

## **Knowledge (-R-I-U)**

**DEFINITION:** An understanding of some sort held by a HumanEntity that affects the HumanEntity's ability to perform the SearchTask. For instance, if a FirstPerson has specific knowledge of the

KeyWords: “Adobe Photoshop Clone Tool,” he/she will have better success in locating Information on how to use the Clone Tool in the computer application, Adobe Photoshop. Not all HumanEntities in this ontology will have Knowledge as defined here.

**NECESSARY & SUFFICIENT PROPERTIES:**

∃ isKnowledgeHeldBy HumanEntity

**NECESSARY PROPERTIES:**

owl:Thing

∃ hasInfluenceOver SearchTask

∃ isInfluencedBy Learning

**SUBCLASSES:**

- **KeyWords (-R-I-U) \***

*DEFINITION:* Search terms (which are not evaluated by a librarian or other information professional as Surrogates are) conceived of by a HumanEntity, which are entered into a search engine to search a Container for Information.

- **ContainerAddress (-R-I-U) \***

*DEFINITION:* Knowledge of the actual location signifier in order to visit a Container.

- **ContainerExistence (-R-I-U) \***

*DEFINITION:* Knowledge that a certain Container exists that contains the Information the HumanEntity needs.

*INDIVIDUALS:* {AwareOfContainer UnawareOfContainer}

- **RegistrationInformation (-R-I-U) \***

*DEFINITION:* Knowledge/remembrance of the actual username and password to obtain access to a Container.

## **Learning (+R-I-U)**

**DEFINITION:** A method engaged in by a HumanEntity to assist him/her in improving Knowledge so that he/she may improve SearchTask performance. Not all HumanEntities in this ontology will engage in Learning.

**NECESSARY PROPERTIES:**

owl:Thing

∃ hasInfluenceOver Knowledge

∃ isEngagedInBy HumanEntity

**SUBCLASSES:**

- **InPersonWorkshop \***

*DEFINITION:* A type of learning situation where a HumanEntity must be physically present with other students and a teacher(s) to participate in the learning situation. Duration is irrelevant; thus an instance could be a semester-long class or a one-day training. The distinguishing characteristic between this subclass and OnlineWorkshop is whether the HumanEntity must be physically present or not.

*POSSIBLE INDIVIDUALS:* THE name of the workshop such as NCLive Training

- **OnlineWorkshop \***

*DEFINITION:* A type of learning situation where a HumanEntity may participate virtually, via a computer.

## MeansOfAccess (+R+I-U)

**DEFINITION:** A form of admittance to a Container for the purpose of obtaining Information. A Container may have more than one MeansOfAccess. It is influenced, or more specifically, obstructed, by AccessRequirements. Not all MeansOfAccess are influenced by an AccessRequirement.

**REGARDING ONTOCLEAN ASSIGNMENT:** The way in which this ontology defines or uses the concept of Borrowing, LoggingIn, etc. necessarily make them separate and distinguishable subclasses. Additionally, each instance of these subclasses is set apart from another by, at the very least, time. A FirstPerson may navigate many or the same website multiple times during a SearchTask; but each instance of Navigation is fundamentally different due to the fact that they occur at different times usually to accomplish different nuances of the same search. For instance a FirstPerson may navigate the same website twice in the exact same manner: clicking through the exact same links in the exact same order but will do this at different times during the SearchTask and probably will be seeking slightly different information. S/he may not have retrieved a certain detail on the first go-round and may want to revisit the webpage to re-review it for that detail. A HumanEntity may borrow different books at different times in order to resolve a single search objective. Lastly, because a given MeansOfAccess may not be the only MeansOfAccess attempted to resolve a search objective, it is reasonable to assume that a single MeansOfAccess may or may not be a complete entity or action in and of itself.

### NECESSARY & SUFFICIENT PROPERTIES:

$\forall$  isMeansOfAccessFor Container  
isMeansOfAccessFor  $\geq 1$

### NECESSARY PROPERTIES:

owl:Thing  
 $\exists$  isInfluencedBy AccessRequirement

### SUBCLASSES:

- **Borrowing (+R+I-U) \***

*DEFINITION:* A means of access specific to a Library where a HumanEntity borrows a Container or Information. This is directly associated with the Access Requirement: Membership: Library Card.

#### NECESSARY & SUFFICIENT PROPERTIES:

$\forall$  isMeansOfAccessFor Library

- **LoggingIn (+R+I-U) \***

*DEFINITION:* A means of access specific to any Container that can be seen as virtually situated. It requires the use of a user name and password. This is directly associated with the Access Requirement: Membership, usually Registration; but sometimes having a Library Card means that a HumanEntity is also given online access to resources.

- **Navigation (+R+I-U) \***

*DEFINITION:* A means of access specific to any Container that can be seen as virtually situated; but unlike LoggingIn, does not require the transmission of personal identifiers of the HumanEntity. Unlike the SearchEngine, Navigation assumes travel through a virtual domain by a series of clicks on hyperlinks.

- **Searching (+R-I-U) \***

*DEFINITION:* For the purposes of this ontology, this always involves the use of a search engine. It is a means of access specific to any Container that can be seen as virtually situated; but unlike LoggingIn, does not require the transmission of personal identifiers of the HumanEntity. Unlike Navigation, Searching assumes the use of KeyWords to obtain a list of potential websites to navigate.

- **VerbalCommunication (+R-I-U) \***

*DEFINITION:* A means of access whereby the searching HumanEntity may verbally communicate with another HumanEntity (acting as a Container of Information).

- **DownloadingFiles (+R-I-U) \***

*DEFINITION:* A means of access whereby the searching HumanEntity may pull files from a server via a file transfer program such as WS\_FTP.

### **SearchResult (+R-I-U) \***

**DEFINITION:** The outcome of the SearchTask having an impact on the HumanEntity who engaged in the SearchTask. Possible instances could include: information located and found to be useful, some information located that must be assembled together to produce meaning, etc.

**NECESSARY & SUFFICIENT PROPERTIES:**

$\exists$  hasInfluenceOver HumanEntity

$\forall$  isInfluencedBy SearchTask

**NECESSARY PROPERTIES:**

owl:Thing

### **SearchTask (+R-I-U) \***

**DEFINITION:** The operation a HumanEntity performs to obtain Information to satisfy a need. Its outcome is the SearchResult and its object is Information. It is influenced by Knowledge held by the HumanEntity. The FirstPerson may or may not specify characteristics of the SearchTask.

**NECESSARY & SUFFICIENT PROPERTIES:**

$\exists$  hasInfluenceOver SearchResult

$\exists$  isEngagedInBy HumanEntity

**NECESSARY PROPERTIES:**

owl:Thing

$\forall$  hasCharacteristic SearchTaskCharacteristic

hasCharacteristic  $\geq 0$

$\exists$  isInfluencedBy Knowledge

### **SearchTaskCharacteristic (+R-I-U)**

**DEFINITION:** Those attributes of a specific SearchTask as *perceived* by the HumanEntity (usually the FirstPerson) who engaged in the SearchTask. It is possible that a FirstPerson may not verbalize any characteristics of the SearchTask he/she performed.

**NECESSARY & SUFFICIENT PROPERTIES:**

$\forall$  isCharacteristicOf SearchTask

**NECESSARY PROPERTIES:**

owl:Thing

**SUBCLASSES:**

- **DifficultyLevel (+R-I-U) \***

*DEFINITION:* The quality of arduousness required.

*INDIVIDUALS:* {Easy Difficult Challenging}

*NECESSARY PROPERTIES:*

- ⊃ hasInfluenceOver LevelOfInvolvement
- ⊃ hasInfluenceOver TimeDuration
- ⊃ hasInfluenceOver Timeliness
- ⊃ isInfluencedBy SkillLevel
- ⊃ isInfluencedBy NoveltyLevel
- ⊃ isInfluencedBy Plan

- **LevelOfInvolvement (+R-I-U) \***

*DEFINITION:* The degree of simplicity or tedium involved in the SearchTask.

*INDIVIDUALS:* {Simple SomewhatTedious Tedious}

*NECESSARY PROPERTIES:*

- ⊃ hasInfluenceOver DifficultyLevel
- ⊃ hasInfluenceOver TimeDuration
- ⊃ hasInfluenceOver Timeliness
- ⊃ isInfluencedBy SkillLevel
- ⊃ isInfluencedBy NumberOfSteps
- ⊃ isInfluencedBy Plan

- **NoveltyLevel (+R-I-U) \***

*DEFINITION:* The quality of familiarity with the specific SearchTask.

*INDIVIDUALS:* {New Familiar Routine}

*NECESSARY PROPERTIES:*

- ⊃ hasInfluenceOver DifficultyLevel
- ⊃ hasInfluenceOver TimeDuration
- ⊃ hasInfluenceOver Timeliness

- **NumberOfSteps (+R-I-U) \***

*DEFINITION:* The quality of involvement, in the sense of how deeply a HumanEntity must proceed in order to complete the SearchTask. This could change over time. For instance, if a HumanEntity has performed a SearchTask more than once, he/she may find that some steps may be circumvented, thereby reducing the number of steps required.

*INDIVIDUALS:* {OneStep TwoSteps [etc]}

*NECESSARY PROPERTIES:*

- ⊃ hasInfluenceOver LevelOfInvolvement
- ⊃ hasInfluenceOver TimeDuration
- ⊃ hasInfluenceOver Timeliness

- **OriginOfNeed (+R-I-U) \***

*DEFINITION:* The derivation of the desire for the Information sought in the SearchTask. LaterallyMotivated implies the need originates with a WorkColleague. FromBelowMotivated implies the need originates with a WorkSubordinate. FromAboveMotivated implies that the need originates with a WorkSupervisor or possibly a WorkDepartment or WorkOrganization.

*INDIVIDUALS:* {PersonallyMotivated LaterallyMotivated FromBelowMotivated FromAboveMotivated}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver UrgencyLevel

- **Plan (+R-I-U) \***

*DEFINITION:* The answer to the question: how extensively did the HumanEntity think strategically about how he/she would perform the SearchTask prior to engaging in the SearchTask?

*INDIVIDUALS:* {NoPlan PartiallyFormedPlan FullyFormedPlan}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver DifficultyLevel  
 ⊃ hasInfluenceOver TimeDuration  
 ⊃ hasInfluenceOver Timeliness  
 ⊃ isInfluencedBy LevelOfInvolvement

- **Purpose (+R-I-U) \***

*DEFINITION:* The reason why the HumanEntity undertook the SearchTask. Possible instances could include: I needed this Information to help me complete a routine work task; I needed this information to assist me in making a decision as to how to proceed with an idea I have; I needed this information to learn more about a particular topic; etc.

- **SkillLevel (+R-I-U) \***

*DEFINITION:* The level of technical savvy perceived to be required for the SearchTask.

*INDIVIDUALS:* {NoMoreOrNewSkillsRequired MoreOrNewSkillsRequired}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver DifficultyLevel  
 ⊃ hasInfluenceOver TimeDuration  
 ⊃ hasInfluenceOver Timeliness  
 ⊃ isInfluencedBy LevelOfInvolvement

- **TimeDuration (+R-I-U) \***

*DEFINITION:* The span of time taken to complete the SearchTask.

*INDIVIDUALS:* {OneMinute FiveMinutes OneHour [etc]}

*NECESSARY PROPERTIES:*

⊃ hasInfluenceOver DifficultyLevel  
 ⊃ hasInfluenceOver Plan  
 ⊃ hasInfluenceOver Timeliness  
 ⊃ isInfluencedBy SkillLevel  
 ⊃ isInfluencedBy NumberOfSteps  
 ⊃ isInfluencedBy NoveltyLevel  
 ⊃ hasInfluenceOver LevelOfInvolvement

- **Timeliness (+R-I-U) \***

*DEFINITION:* The pace of the SearchTask.

*INDIVIDUALS:* {VerySlow Slow Quick}



*NECESSARY PROPERTIES:*

- ⊃ hasInfluenceOver LevelOfInvolvement
- ⊃ hasInfluenceOver TimeDuration
- ⊃ hasInfluenceOver SkillLevel
- ⊃ isInfluencedBy Plan
- ⊃ isInfluencedBy NoveltyLevel
- ⊃ isInfluencedBy NumberOfSteps

- **UrgencyLevel (+R-I-U) \***

*DEFINITION:* The level of importance of the task in respect to how quickly it must be undertaken and completed. It would be assumed that the higher the urgency level the more quickly the task must be initiated and completed to the desired degree of completion.

*INDIVIDUALS:* {Low Medium High}

*NECESSARY PROPERTIES:*

- ⊃ isInfluencedBy OriginOfNeed

**Surrogate (+R+I+U)**

**DEFINITION:** A descriptor of a Container that acts as both a finding aid for the Container as well as a tool for judging relevance of the Information contained in the Container. Not all Containers will have Surrogates.

**REGARDING ONTOCLEAN ASSIGNMENT:** Although +I and +U are very strong assignments in this particular ontology where so much is subjective in that it reflects what the FirstPerson perceives to be true, Surrogate is probably the most objective class in the ontology. One abstract for one article is always distinguishable from another abstract for a different article. Although a Citation for a single journal article could appear in APA format whereas another author could cite this same journal article in MLA format, it is essentially the same citation because it points to the same journal article and provides the same identifiers such as the same title, author, etc. One book could be listed under multiple SubjectHeadings and one SubjectHeading could be used for multiple books; but each SubjectHeading implies one topic and therefore assigns at least one topic per Container. Because each subclass and its individuals could meet its obligation as a Surrogate on its own merit, each subclass and individual has unity. It does not require other parts or wholes to define it as a Surrogate.

**NECESSARY PROPERTIES:**

- owl:Thing
- ⊃ isSurrogateOf Container

**SUBCLASSES:**

- **Abstract (+R+I+U) \***

*DEFINITION:* A short blurb about the Information contained in the Container. The Abstract may be searched and it provides a preview to determine relevance.

- **Citation (+R+I+U) \***

*DEFINITION:* Specifically, information about a Container including author, title, published date, volume and issue numbers if a serial Container, page numbers if Journal/Newspaper/Book, URL if Website/DigitalLibrary, etc.

- **SubjectHeading (+R+I+U) \***

*DEFINITION:* A topic or group of topics assigned to a Container by a subject matter expert of some kind (like a librarian). Its intention is to convey what the Information is *about* and

therefore acts as not only a finding aid but also an aid to relevance judgments.

- **TechnicalMetadata** (+R+I+U) \*  
*DEFINITION:* Information about a Container stated in RDF or XML and usually adhering to some sort of standard such as the Dublin Core. This may include Citation information or SubjectHeadings, etc.; and it is often specifically included to make location of the Container on the Internet more effective.
- **VideoClip** (+R+I+U) \*  
*DEFINITION:* A short excerpt from a video that is supposed to be representative of the Information contained in the audio/video of a VideoTape. It helps to guide relevance decisions.

## PROPERTIES

hasAccessRequirement ↔ isAccessRequirementFor

hasCharacteristic ↔ isCharacteristicOf

hasEngagementIn ↔ isEngagedInBy

hasKeeper ↔ isKeeperOf

- Transitive

hasKnowledge ↔ isKnowledgeHeldBy

hasMeansOfAccess ↔ isMeansOfAccessFor

hasSurrogate ↔ isSurrogateOf

isCooperativeWith

- Symmetric

## APPENDIX C: SCREENSHOT OF THE NOTE-TAKING TOOL

Note-Taking Tool

http://www.unc.edu/~llchrist/nttool/

Getting Started Latest Headlines NY Times Washington Post Letters of Marque: ...

Google Note-Taking Tool

### Note-Taking Tool

**General Instructions:** All fields are optional except Analyst, Client, Search Task Title; they have a red asterisk. If you would like help with an item, **rollover** the link. Here is a link to the full instructions: [Coding Instructions PDF](#). **Remember: Approach every form object from the stance of: how is the client perceiving this...**

Analyst: \*  Date:

Client: \*  Search Task Title: \*

#### Characteristics Of Information Sought (or Found)

As perceived by the client...

Exists? ☐ Yes ☐ No Usable As Is? ☐ Yes ☐ No Cost: \$

Available? ☐ Yes ☐ No Available Format: ☐ Electronic ☐ Print Worth:

Complete? ☐ Yes ☐ No Confidentiality Issues Involved? ☐ Yes ☐ No Centrality:

Current? ☐ Yes ☐ No Proprietary Rights Required? ☐ Yes ☐ No How well did the information match the client's needs?

#### Location Of Information

Where was the information located? Check all that apply.

☐ Computer ☐ Book ☐ Website ☐ Internet (or "the web")

☐ Server ☐ Journal/Magazine (including articles) ☐ Database ☐ Unspecified article

☐ Library ☐ Newspaper (including articles) ☐ Digital Library

☐ Another person ☐ Video Tape ☐ Archive

#### Location Characteristics

As perceived by the client...

Amount of Information Contained:  Active? ☐ Yes ☐ No

Interface Usability:  Address: ☐ Static ☐ Changing

Pertinence:  Organized? ☐ Yes ☐ No

#### Means Of Access

What types of access methods did the client use or feel was needed? Check all that apply.

☐ Borrowing ☐ Used Search Engine

☐ Logging In ☐ Communication (verbal: in person, phone, email)

☐ Navigation (just clicking through links; web surfing) ☐ Uploading/downloading files

#### Surrogates Used Or Involved In The Search Task

What surrogates did the client use or feel was needed? Check all that apply.

☐ Abstract ☐ Metadata (actual XML or RDF coding)

☐ Citation ☐ Video Clip

☐ Subject Heading

#### Any Access Requirements Encountered

What restrictions were placed on access to the information and/or location? Check all that apply.

☐ Client's department would have to or did purchase information. ☐ Library membership (via a library card) required ☐ A confidentiality waiver must be signed.

☐ Client would have to or did personally purchase information. ☐ Registration/subscription ☐ A proprietary agreement must be negotiated.

☐ Permeation of a firewall required.

#### Client's Knowledge And Learning That Might Prepare Him/Her For Search Task

Did the client feel that there was some knowledge or data s/he needed to perform the search task? ☐ Yes ☐ No

If yes, what did the client believe s/he needed to know? (Check all that apply.)

☐ Search Terms ☐ If a location exists for the desired information

☐ Registration Information (username/password) ☐ The location's address

Other knowledge needed:

Did the client take any trainings to assist him/her with performing the Search Task? ☐ Yes ☐ No

If yes, did the client feel that the training(s) was helpful? ☐ Yes ☐ No

**Characteristics Of The Search Task***As perceived by the client...*

<a href="#">Required Skill Level:</a>	Select	<a href="#">Level of Involvement:</a>	Select	Urgency Level:	Select
Difficulty Level:	Select	<a href="#">Number of Steps:</a>		Origin of Need:	Select
Level of Familiarity:	Select	<a href="#">Duration:</a>	Minute(s)	<a href="#">Purpose:</a>	
Search Plan Devised?	Select	Timeliness:	Select		

**Search Result****Cultural Factors***Were any of the following types of people involved; and if so, what was the nature of their involvement?*

<input type="checkbox"/> Work Colleague	Select	<input type="checkbox"/> <a href="#">Work Department</a>	Select
<input type="checkbox"/> Work Subordinate	Select	<input type="checkbox"/> <a href="#">Overarching Work Organization</a>	Select
<input type="checkbox"/> Work Supervisor	Select	<input type="checkbox"/> <a href="#">Outside Organization</a>	Select
<input type="checkbox"/> Other Professional	Select	<input type="checkbox"/> <a href="#">Professional Community</a>	Select

Done

## APPENDIX D: CODING INSTRUCTIONS

### Overview of Project:

For my research assistantship, I interviewed people who work for NC state government, specifically the Department of Health & Human Services and the Department of Commerce. The data from one of the questions from this interview serves as the fodder for my Master's Paper research. The question was: "What difficulties/frustrations have you experienced in trying to get the information you need to do your job?" From these "problem expressions" (their answers to this question), I developed an ontology to describe the problem space. Then I translated the ontology into a note-taking tool (which I've made available in HTML) that, theoretically, could be used during interviews between systems analysts and clients. My hope is that this will help structure interviews so that communication between analysts and clients is improved. To establish reliability, I have asked you to replicate my coding efforts and I am extremely grateful for your assistance!!

**Tool Location:** <http://www.unc.edu/~llchrist/nttool/>

### About the Tool:

The note-taking tool is a web form that will email me the results when you click the submit button. All fields are optional except the Analyst, Client, and Search Task Title field. This means that there will be times when you simply do not run across something in a problem expression that matches a form object. For instance, one form object is Exists? You may read problem expressions that never discuss the "existence" of information. So you would leave this form object untouched. Help information is available by mouse rollovers (*not clicks*) on links.

### Overall Process:

1. Please complete the form for each problem expression. Since you will receive 20 expressions, you will complete the form 20 times.
2. For the Analyst field, please put your name there. Your first name is sufficient. I just need to tell you all apart. (I will not print your names in the paper. Your confidentiality will be protected.)
3. For the Client field, please put the ID number for the problem expression.
4. You may leave the Date field blank. Date is to represent the date in which the real-life systems analyst would sit down to talk to the real-life client. So we don't need to worry about Date for coding. It's just a place marker.
5. For Search Task Title, please make up a name for the problem expression that you feel is appropriate. There is no right or wrong answer here. For example, if the problem expression talks about finding only abstracts and not the full-text article and how this is frustrating, you could put "Abstract Only Frustration" in this field. This does not need to be unique. The Client ID will suffice for establishing unique records.
6. Then fill in the other form fields and click Submit when you are done.
7. You will then be presented with a confirmation/thank-you page which will allow you to link back to a freshly reset form for inputting another problem expression.

## Instructions per Form Object:

**Note:** Problem expressions can focus on many different types of scenarios. Some interviewees spoke of specific information-seeking instances such as information they sought in the past or information they still need but can't find. Some spoke of searching in general with no mention of a specific task they performed. No matter how they approached the interview question, consider what the underlying meaning is and don't worry about the temporal slant. For instance, if they said something like "I'm looking for information on people age 65 and older who receive Medicare. I don't know if this information exists," then go ahead and click "No" for the **Exists?** form object; because at the time they were saying this to me, they didn't have proof of the information's existence. So *for them*, it did NOT exist. This could be interpreted as "I was frustrated because I can't find this information. It may not exist and I need it. Needing something that doesn't exist in the form you need it (or pulled together in the way you need it) is exasperating." So modeling existence is critical here.

Also, you may be confused when you see that I have things like **Required Skill Level** under the header **Characteristics of the Search Task** instead of being associated with people. Ordinarily we are taught that "skill" is an attribute of a person, not a task. Ontologies work differently from databases in this way. Also, characteristics (or attributes) are often those qualities as *perceived* by the interviewee. **Approach every form object from the stance of: how is the client *perceiving* this as it relates to his/her needs?** So for instance, **Required Skill Level** is what the interviewee *thought* was needed to accomplish the task. She might have said, "I just don't know if I had the right skills. I don't know if it was a problem with me or with the system." So for this, you would select "More /new Skills Required" from the **Required Skill Level** drop down. She suspects this may be a problem. Because the idea of not having the right skill set is a possibility in her mind, go ahead and account for that. Her perception, even if she's unsure but suspects, is worth modeling.

Also, assume the most neutral drop-down menu item, if you are unsure. For instance, the drop down for Timeliness has options: Quick, Slow, Very Slow. If you are unsure whether the client perceived the process to be slow or very slow, then it is fine to pick Slow.

Only work from actual mentions of these concepts. Try not to assume unless a definition of a concept allows for assumption. The example at the end of this document will explain this better.

## Characteristics of Information Sought (or Found)

- **Exists?** Does the client think the information exists? If the existence is in question (as in the example above), assume "No."
- **Available?** Does the client perceive the information to be available to him/her, not just available in general. The client may believe the information exists but feel that s/he can't access it for whatever reason.

- **Complete?** Does the client consider the information to be complete or is something lacking? Just receiving an abstract and not the full-text when the full-text is desired would require a “No.” So complete should be evaluated in terms of if the information is complete for the client’s perceived needs.
- **Current?** Is the information to the desired level of currency (up-to-date-ness) for the client? Remember, most of this is a matter of perception. What is current to one client may not be current to another.
- **Usable As Is?** This question requires a mention of actual *use* of the information. Is the information usable as is or does it require some type of manipulation? For instance, does the client have to further extract or collate data to get the exact meaning they need? Is it in the right format or do they have to hand-code the data into another application in order to use it?
- **Available Format:** This refers to the information being electronic or print rather than file format. Obviously journal articles come in both formats; but if they are talking about finding the article on a website, then assume only electronic. Only check the Available Format(s) they specifically mention or allude to in a way that can be corroborated with some other feature of the expression as in the case of finding an article on a “website.”
- **Confidentiality Issues Involved?** Does the client mention that confidentiality is involved - confidentiality of either the information they seek or the information they need to communicate in order to get more information? Just because something is proprietary or requires payment does not mean it is confidential. They will more than likely use the word “confidential.”
- **Proprietary Rights Required?** Are there proprietary rights to the information desired? Just because something is confidential or requires payment does not mean it is proprietary. They will more than likely use the word “proprietary.”
- **Cost:** A monetary value can be inputted here if you receive one. If they mention a range (i.e. \$15-20), pick the higher number (\$20).
- **Worth:** This is getting at the idea of how much is the information worth given its cost? So for instance some clients will say something was “just too expensive.” The underlying meaning is “Not Worth It – too expensive.” Or sometimes they will be more explicit and say that it was too expensive and they didn’t need it bad enough to pay the price. Or sometimes they will say that the information is very costly but they really need it and so they wish they had the money for it but they don’t. This would be “Worth It – but expensive.” If they talk in terms of the search task taking a really long time or was too laborious and the information just wasn’t worth it, then you can consider “expensive” in terms of time/labor consumption. So worth is to be measured in terms of expense (either monetary or time/labor).
- **Centrality:** Could the client find the information they needed in one place or did (or would) they have to gather bits and pieces of information across multiple locations in order to get the exact meaning/notion/answer they needed?
- **How well did the information match the client’s needs?** How relevant was the information? How satisfied with it was the client? Was it on topic?

### Description of Information Sought???

You might be wondering why I am not providing a place to enter the actual information sought, something like “demographic information about people aged 65 or older who receive Medicare.” Theoretically, a whole other ontology would need to be built to describe the actual information sought. My ontology is focused on the process of seeking and how people perceive this process and its constituent components. The actual information sought is outside the scope of the project.

### Location of Information

Check all that were specifically mentioned. Don’t assume. For instance, just because a client says “website” does not mean that you should click “Server” or “Internet.” Only check what they actually said or that which can be corroborated by other features of the expression. If they said it *might* be in the library, go ahead and check “Library.” Again, it’s a matter of *their* perception. If you run across a mention of NCLive, you can assume it is a Digital Library on the Internet and it is a Database. You can check all 3 of these. You can also assume any information sought on NCLive is electronic and required Logging In and Registration/subscription. If they say “journal article,” you can assume Journal. If they say “newspaper article,” you can assume Newspaper. If they say LexisNexis, you can assume Database and Digital Library.

### Location Characteristics

- **Amount of Information Contained:** Was the amount of information they encountered overwhelming? In other words, did they experience information overload? For instance, if a client talks about surfing the web and there being sooo much information, then check “Internet” for the Location and “Overwhelming” for this form object. Or was the information contained sparse and therefore too limited?
- **Interface Usability:** How easy was the Location to use or navigate? If they say there were too many drill-downs or too many options on the site, then assume “Difficult” at the least. If they go on and on about this or use a more vehement term, you might be able to assume “Challenging.” If you have any question about “Difficult” or “Challenging,” just put “Difficult.”
- **Pertinence:** Were the location(s) they searched or knew about the *right* locations, in their opinions. If they talk about *needing* to find the *right* location (including people to talk to), assume they did not and select “Incorrect location.”
- **Organized:** Was the information in the Location organized or disorganized? One expression talks about how the client wishes he were better at filing information. This assumes his information is disorganized. He uses his own body of information as a resource, so this counts as a Location that is disorganized.
- **Active?** Was the Location (usually a Website, Server, etc.) up or down? Active or inactive? For example, “the server was down” would require a “No” for this form object and a check of “Server” under Location of Information.
- **Address:** Was the client able to successfully find the address (usually URL) of the Location a second/third/etc. time or did it change? In other words, did they get a “Page Not Found” error or did the site move, etc.?



### Means Of Access

- **Borrowing** – specifically refers to libraries (also interlibrary loan).
- **Logging In** – specifically refers to virtual Locations which require a username and password (registration/subscription).
- **Navigation** – specifically refers to clicking hyperlinks or buttons to move from screen/page to screen/page in virtual Locations.
- **Used Search Engine** – specifically refers to use of a search engine (entering search terms) in a virtual Location.
- **Communication** – specifically refers to people; but people can be in a Library.
- **Uploading/downloading files** – specifically refers to transferring data/files.

### Surrogates

The only surrogate that might give you pause is “Metadata.” I am specifically referring to XML or RDF or something of that nature. Yes, a subject heading and a citation, etc. are metadata; but please limit “Metadata” to just that thing that geeks do to mark up documents, usually web documents. Synonyms for “Subject Heading” include “topic” or “subject.” It doesn’t have to mean an actual Subject Heading. More, it means searching by subject/topic.

### Any Access Requirements Encountered

What, if any, restrictions/barriers were placed on access to a Location/Information? Registration/subscription refer to having to give personally identifying information and being assigned a username and password. Registration/subscription may be free or may require payment. So you might check more than one access requirement.

### Client’s Knowledge and Learning that Might Prepare Him/Her for the Search Task

The knowledge aspect of this is: did the client feel that there was something s/he needed to know in order to succeed at the search task? Did they need to know

- search terms (the idea that sometimes we don’t have the correct vocabulary to formulate a query because we are new to the subject area),
- registration information (did the client forget his/her username/password or not have it on him/her),
- existence of an appropriate location to search (the idea that sometimes we just don’t know where to go first for information, we don’t know what resources there are),
- or the actual address (URL, for example) of a potential location?

If they mention something outside of these, please put it in the text field. I believe the training/learning questions are fairly self-evident.

### Characteristics of the Search Task

- **Required Skill Level:** Does the client feel that s/he needed more/new skills to accomplish the task? If the client makes no mention of skills, don’t assume “No more or new skills required.” Just leave this untouched (as you would for other form objects that are not referenced in the problem expression).

- **Difficulty Level:** How difficult/hard/challenging does the client feel the task is?
- **Level of Familiarity:** Has the client performed the task before? If so, is the task simply familiar or is it now routine (done frequently)?
- **Search Plan Devised:** You may not run across a mention of this. Did the client think strategically about the search prior to beginning it or think that this was needed to improve the task?
- **Level Of Involvement:** What is the client's perception about how tedious or simple the task is?
- **Number Of Steps:** You may not run across a mention of this; but if you do, the actual number would be inputted here.
- **Duration:** This is also rarely mentioned. Did the client specify an actual time frame or range of times? If s/he specified a range, please input the larger number. For example, for "15-20 minutes," please put 20 minutes.
- **Timelines:** Does the client perceive the time the task took or will take to be fast or slow?
- **Urgency Level:** How important is the task?
- **Origin of Need:** Did the need for the information originate with the client ("Personally Motivated"), with a supervisor or other person positioned above the client ("Motivated from Above"), from someone at the same level as the client, such as a colleague ("Motivated Laterally"), or from someone under the client, such as his/her subordinate ("Motivated from Below").
- **Purpose:** If the client expresses a reason why they need the information, input it here. This will be infrequently encountered.

### Search Result

If you feel like inputting text here, please do. Examples that I could see entered would be: "information not found," "information located but was not helpful," etc.

### Coding Examples:

I have made up a fictitious problem expression and will show you how I would code it.

"I really hate it when I search the web and, first of all, I can't find the right website. So I get kind of confused. So then I'll ask someone I work with and they'll suggest a website. Like one time, I found this website that would give me information about dogs that have red fur and curly tails, right? But they never said what kinds of dogs these were. I was trying to figure out if all Pomeranians have red fur. My boss wanted to know. So I searched NCLive and found a journal article about red, curly-tailed dogs but I couldn't get access to the full-text. Apparently, our subscription didn't include everything that was available. That was annoying."

Search Task Title: Looking for red, curly-tailed dog names

Exists? Yes <i>(The client found something, so they did not question the existence of the information.)</i>	Usable As Is? leave untouched <i>(Because the client is not specifically speaking about any use of the information)</i>	Cost: leave untouched  Worth: leave untouched  Centrality: leave untouched <i>(Although they looked in 2 different locations, this does not mean the information they needed was in more than one place.)</i>
Available? No <i>(Ultimately, they could not access the full-text of the journal article.)</i>	Available Format: Electronic  Confidentiality: No	
Complete? No <i>(They only got the abstract and the earlier website didn't give them all the information they needed.)</i>	Proprietary: No <i>(Remember this usually requires a specific mention of "proprietary.")</i>	How well did the information match the client's needs? Somewhat relevant <i>(I would put Somewhat relevant because of the first piece of information the client found – it was partially helpful. The journal article we don't know about because the client wasn't able to get it.)</i>
Current? leave untouched		

#### Location of Information

I would check

- another person *(because they asked another person where to find the information, this information is part of the overall search task and can be seen as a supporting component of the total information sought)*
- website,
- database *(because I've said explicitly that assuming this of "NCLive" is okay)*
- Digital Library *(because I've said explicitly that assuming this of "NCLive" is okay)*
- Internet *(because the client specifically said "web" and because I've said explicitly that assuming this of "NCLive" is okay)*
- Journal *(because the client specifically says "journal article")*

So even though the client mentioned different locations for different types of information s/he found, I will check *all* locations mentioned not just the one for the information the client seemed to feel might be the most helpful.

### Location Characteristics

Amount of Information: leave untouched

Active? leave untouched

*(I suppose you could select Active and that would be a reasonable assumption; but leaving it blank because of it not being explicitly mentioned is okay too.)*

Interface Usability: leave untouched

Pertinence: Incorrect location *(because they made mention of not finding the “right website;” and because the information they found on the one website was not quite right)*

Address: leave untouched

### Means Of Access

- I would check Navigation because they mentioned using a website and navigation is *always* a part of using a website.
- I would also check Logging In, in part because I know the NCLive requires some type of identification process and I have said that it is okay to assume this, but more because they confirm this by saying “subscription.” So if in doubt about something (like if you didn’t know that NCLive required logging in) don’t include it, don’t guess. Look to the definitions of questions above as to what assumptions you can make. Registration/subscription is defined as needing a username and password which necessarily implies Logging In.

### Surrogates

- Abstract

### Any Access Requirements Encountered

- Registration/subscription *(It was required to even get into NCLive and the quality of it was limited making the nature of the subscription even more of a restriction. I could make an assumption that the article could have been purchased; but since the client did not specifically speak about that and because s/he was focusing more on the idea of incompleteness and access requirements as the sources of frustration, I would not check either of the requirements that talk about making a purchase.)*

### Client’s Knowledge and Learning

Did the client feel that there was some knowledge or data s/he needed to perform the search task? Yes

Leave all checkboxes blank except: If a location exists for the desired information *(since the client said that finding the right website – implying that there exists a website that would serve his/her needs – was a frustration).*

Other knowledge needed: leave untouched

Did the client take any trainings to assist him/her with performing the Search Task? leave untouched *(I don’t know if they did or didn’t. It was never mentioned.)*

If yes, did the client feel that the training(s) was helpful? leave untouched

### Characteristics of the Search Task

Required Skill Level: leave untouched	Level Of Involvement: leave untouched	Urgency Level: leave untouched
Difficulty Level: leave untouched ( <i>“difficulty” was never mentioned; “annoyance” was but that’s not the same</i> )	Number of Steps: leave untouched	Origin of Need: Motivated from Above
	Duration: leave untouched	Purpose: leave untouched
Level of Familiarity: leave untouched	Timeliness: leave untouched	

Search Plan Devised?  
leave untouched

### Search Result:

“appears to have never been able to answer question due to limited subscription and therefore limited information received – so couldn’t ever evaluate the relevance of a particular document”

### Cultural Factors

- Work Colleague – cooperative
- Work Supervisor – leave cooperative drop down untouched

**APPENDIX E: CODER QUESTIONNAIRE**

1. How difficult was the tool to use?
  - Not difficult at all
  - Somewhat difficult
  - Very difficult
  - Extremely difficult
2. Did you ever run across an expression where you wanted to model some aspect of it but felt that there was nothing on the tool/form that made that possible?
  - Yes
  - No
3. If yes in question 2, roughly, how many times do you think this occurred?
  - 1-2 times
  - 3-4 times
  - 5-6 times
  - 7-8 times
  - 9-10 times
  - more than 10 times
4. Any other comments you might have for me that I can report on?

## APPENDIX F: HOW CODINGS WERE ANALYZED

For any given problem expression, a coder could select any number of form items with which to interact. For example, the problem expression: “There’s specific proprietary data we’d love to get our hands on that we basically have to negotiate to get” was coded as such:

Researcher:

Proprietary = Yes

Access Requirement = Proprietary Agreement

Coder 1:

Proprietary = Yes

Coder 2:

Exists = Yes

Available = No

Available Format = Electronic

Proprietary = Yes

Access Requirement = Departmental Expenditure

Access Requirement = Personal Expenditure

Access Requirement = Registration

All form items were tallied on a binary scale, but also with consideration of form items that were left untouched since untouched form items demonstrate judgments as well. Form items were either tallied as Y for Yes, N for No, or B for blank (left untouched). The above problem expression was registered in an Excel spreadsheet as follows:

Excel Field Label	Judgments
Exists	YBB
Available	NBB
Electronic	YBB
Proprietary	YYY
Departmental Expenditure	YBB
Personal Expenditure	YBB
Registration	YBB
Proprietary Agreement	YBB

All other fields = BBB

Judgments for form items were recorded in an Excel spreadsheet according to the following rules. **R** is the number of rows that were used in the Excel spreadsheet to represent the form item.

<b>R</b>	<b>FORM ITEM</b>	<b>HOW JUDGMENTS WERE RECORDED</b>
<b>Characteristics of Information Sought (or Found)</b>		
1	Exists? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Available? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Complete? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Current? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Usable As Is? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Available Format: Electronic <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Available Format: Print <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Confidentiality Issues Involved? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Proprietary Rights Required? <i>yes/no radio button</i>	Yes radio button = Y No radio button = N Untouched = B
1	Cost: <i>text field</i>	If the number entered was the same = Y If the number entered was different = N Untouched = B <i>Example: If two coders said the cost was \$5 and one said it was \$10, then the judgment would be YYN.</i>



4	<p>Worth: <i>drop-down menu with the following options:</i></p> <ul style="list-style-type: none"> <li>▪ <i>Worth It - but expensive</i></li> <li>▪ <i>Worth: Worth It – inexpensive</i></li> <li>▪ <i>Worth: Not Worth It - too expensive</i></li> <li>▪ <i>Worth: Not Worth It - yet inexpensive</i></li> </ul>	<p>Drop-down menu item selected = Y Untouched = B <i>There were 4 rows in the Excel spreadsheet to represent Worth: one row per drop-down option. For example, if one coder selected “Worth It – but expensive,” another selected “Worth It – inexpensive,” and the third left it untouched, the cells would appear as follows:</i></p> <p><i>YBB = Worth It - but expensive</i>  <i>YBB = Worth: Worth It – inexpensive</i>  <i>BBB = Worth: Not Worth It - too expensive</i>  <i>BBB = Worth: Not Worth It - yet inexpensive</i></p>
1	<p>Centrality <i>drop-down menu with the following options:</i></p> <ul style="list-style-type: none"> <li>▪ <i>Located in one place</i></li> <li>▪ <i>Scattered</i></li> </ul>	<p>“Scattered” selected = Y “Located in one place” selected = N Untouched = B <i>Example: If one coder selected “Located in one place” and another selected “Scattered” and the third left it blank, then this cell would receive a YNB.</i></p>
3	<p>How well did the information match the client’s needs? <i>drop-down menu with the following options:</i></p> <ul style="list-style-type: none"> <li>▪ <i>Not Relevant</i></li> <li>▪ <i>Somewhat Relevant</i></li> <li>▪ <i>Highly Relevant</i></li> </ul>	<p>Drop-down menu item selected = Y Untouched = B <i>There were 3 rows in the spreadsheet to represent this form item: one row per drop-down option. For example, if one coder selected “Not Relevant” and the other two left it untouched, the cells would appear as follows:</i></p> <p><i>YBB = Not Relevant</i>  <i>BBB = Somewhat Relevant</i>  <i>BBB = Highly Relevant</i></p>
<b>Location of Information</b>		
1	Computer <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Server <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Library <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Person <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Book <i>checkbox</i>	Checkbox checked = Y Untouched = B

1	Journal <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Newspaper <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Video Tape <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Website <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Database <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Digital Library <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Archive <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Unspecified article <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Internet <i>checkbox</i>	Checkbox checked = Y Untouched = B
<b>Location Characteristics</b>		
3	Amount of Information Contained: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ <i>Sparse or Limited</i></li> <li>▪ <i>Satisfactory</i></li> <li>▪ <i>Overwhelming</i></li> </ul>	Drop-down menu item selected = Y Untouched = B <i>There were 3 rows in the Excel spreadsheet to represent this form item: one row per drop-down option.</i>
1	Interface Usability: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ <i>Easy to Use</i></li> <li>▪ <i>Difficult to Use</i></li> </ul>	“Difficult to use” selected = Y “Easy to use” selected = N Untouched = B
1	Pertinence: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ <i>Correct location to find desired information</i></li> <li>▪ <i>Incorrect location to find desired information</i></li> </ul>	“Correct location” selected = N “Incorrect location” selected = Y Untouched = B
1	Active? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Address: <i>radio button with the two options:</i> <ul style="list-style-type: none"> <li>▪ <i>Changing</i></li> <li>▪ <i>Static</i></li> </ul>	“Changing” selected = Y “Static” selected = No Untouched = B

1	Organized? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
<b>Means of Access</b>		
1	Borrowing <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Logging In <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Navigation <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Used Search Engine <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Communication <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Uploading/downloading files <i>checkbox</i>	Checkbox checked = Y Untouched = B
<b>Surrogates Used or Involved in the Search Task</b>		
1	Abstract <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Citation <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Subject Heading <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Metadata <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Video Clip <i>checkbox</i>	Checkbox checked = Y Untouched = B
<b>Any Access Requirements Encountered</b>		
1	Client's department would have to or did purchase information. <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Client would have to or did personally purchase information. <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Library membership <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Registration/subscription <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	A confidentiality waiver must be signed. <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	A proprietary agreement must be negotiated. <i>checkbox</i>	Checkbox checked = Y Untouched = B

1	Permeation of a firewall required. <i>checkbox</i>	Checkbox checked = Y Untouched = B
<b>Client's Knowledge and Learning that Might Prepare Him/Her for the Search Task</b>		
1	Did the client feel that there was some knowledge or data s/he needed to perform the search task? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	Search Terms <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Registration Information <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	If a location exists for the desired information <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	The location's address <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Other knowledge needed: <i>text field</i>	If coders entered something into this text field that was similar in nature = Y Untouched = B
1	Did the client take any trainings to assist him/her with performing the Search Task? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
1	If yes, did the client feel that the training(s) was helpful? <i>yes/no radio button</i>	Yes = Y No = N Untouched = B
<b>Characteristics of the Search Task</b>		
1	Required Skill Level: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ <i>No more/new skills required</i></li> <li>▪ <i>More/new skills required</i></li> </ul>	"No more/new skills " selected = N "More/new skills" selected = Y
1	Difficulty Level: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ <i>Easy</i></li> <li>▪ <i>Difficult</i></li> </ul>	"Easy" selected = N "Difficult" selected = Y Untouched = B
3	Level of Familiarity: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ <i>New task</i></li> <li>▪ <i>Familiar task</i></li> <li>▪ <i>Routine task</i></li> </ul>	Drop-down menu item selected = Y Untouched = B <i>There were 3 rows in the Excel spreadsheet to represent this form item: one row per drop-down option.</i>

3	Search Plan Devised? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ No plan</li> <li>▪ Partial plan</li> <li>▪ Full plan</li> </ul>	Drop-down menu item selected = Y Untouched = B <i>There were 3 rows in the Excel spreadsheet to represent this form item: one row per drop-down option.</i>
1	Level of Involvement: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Simple</li> <li>▪ Tedious (No = Simple)</li> </ul>	“Simple” selected = N “Tedious” selected = Y Untouched = B
1	Number of Steps: <i>text field</i>	If the number entered was the same = Y If the number entered was different = N Untouched = B
1	Duration <i>text field with drop-down menu with the following options</i> <ul style="list-style-type: none"> <li>▪ Minute(s)</li> <li>▪ Hour(s)</li> <li>▪ Day(s)</li> <li>▪ Week(s)</li> <li>▪ Month(s)</li> </ul>	If the number entered in the text field and the drop-down option selected were the same = Y If the number entered in the text field and the drop-down option selected were different = N Untouched = B
1	Timeliness: <i>drop-down menu with the following options</i> <ul style="list-style-type: none"> <li>▪ Very Slow</li> <li>▪ Slow</li> <li>▪ Quick</li> </ul>	“Very Slow” or “Slow” selected = Y “Quick” selected = N Untouched = B
3	Urgency Level: <i>drop-down menu with the following options</i> <ul style="list-style-type: none"> <li>▪ Low</li> <li>▪ Medium</li> <li>▪ High</li> </ul>	Drop-down menu item selected = Y Untouched = B <i>There were 3 rows in the Excel spreadsheet to represent this form item: one row per drop-down option.</i>
4	Origin of Need: <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Personally motivated</li> <li>▪ Motivated laterally</li> <li>▪ Motivated from above</li> <li>▪ Motivated from below</li> </ul>	Drop-down menu item selected = Y Untouched = B <i>There were 4 rows in the Excel spreadsheet to represent this form item: one row per drop-down option.</i>
<b>Cultural Factors</b>		
1	Work Colleague <i>checkbox</i>	Checkbox checked = Y Untouched = B

1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Work Subordinate checkbox	Checkbox checked = Y Untouched = B
1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Work Supervisor checkbox	Checkbox checked = Y Untouched = B
1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Other Professional checkbox	Checkbox checked = Y Untouched = B
1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Work Department checkbox	Checkbox checked = Y Untouched = B
1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Overarching Work Organization checkbox	Checkbox checked = Y Untouched = B

1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Outside Organization <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
1	Professional Community <i>checkbox</i>	Checkbox checked = Y Untouched = B
1	Uncooperative? <i>drop-down menu with the following options:</i> <ul style="list-style-type: none"> <li>▪ Cooperative</li> <li>▪ Uncooperative</li> </ul>	Cooperative = N Uncooperative = Y Untouched = B <i>Each type of human entity listed had a corresponding drop-down menu to determine the human entity's level of cooperation.</i>
102	Total Rows	

## APPENDIX G: CALCULATION OF THE KAPPA

Using Joseph Fleiss' (1971) formula for calculating the percent of agreement with more than two raters and adapting it to suit non-mutually exclusive categories (per the assistance of Dr. Michael Schlessinger):

### Probability of Agreement (**P**)

$$P = 1/n(n-1)N_1 \times \text{sum of squares} - 1/n-1$$

$n$  = the number of raters for any given problem expression

$n(n-1)$  = the number of ordered pairs of raters if you have  $n$  raters

$$n(n-1) = 3(3-1) = 6$$

$1/n-1$  = simplification of part of Fleiss' formula explained below

$$1/n-1 = 1/3-1 = 1/2 \text{ or } .5$$

$N_1$  = the number of categories (rows in the Excel spreadsheet / form items) multiplied by the number of subjects (problem expressions) = the total number of cells

$$\text{So, } P = 1/6N_1 \times \sum(Y^2 + N^2 + B^2) - 1/2$$

$$\text{Sum of squares} = \sum(Y^2 + N^2 + B^2) = S^s + S^s + S^s + \text{etc.}$$

$$\text{For one cell, } S^s = Y^2 + N^2 + B^2.$$

( $Y$  = Yes judgments,  $N$  = No judgments,  $B$  = left untouched/blank. See Appendix F.

The  $S^s$  notation is used in the table below to reduce column width.)

In a single cell, for instance,  $Y(Y-1)$  is the number of Yes pairs. Thus the number of agreeing pairs in a cell is:

$$\begin{aligned} &= Y(Y-1) + N(N-1) + B(B-1) \\ &= Y^2 + N^2 + B^2 - (Y + N + B) \\ &= Y^2 + N^2 + B^2 - 3 \end{aligned}$$

All together, there are  $n \times (n-1) \times N_1 = 6N_1$  pairs of judgments. So the proportion of agreeing judgments in all  $N_1$  cells is:

$$\begin{aligned} &= 1/6N_1 \times \sum(Y^2 + N^2 + B^2 - 3) \\ &= \sum(Y^2 + N^2 + B^2)/6N_1 - 3N_1/6N_1 \\ &= 1/6 N_1 \times \sum(Y^2 + N^2 + B^2 - 3) - 1/2 \\ &= \sum(Y^2 + N^2 + B^2)/6 \times N_1 - .5 \text{ or } (S^s + S^s + S^s + \text{etc.} / 6 \times N_1) - .5 \end{aligned}$$

For example: a judgment of YYB in one cell in the Excel spreadsheet would be calculated:

$$S^s = 2^2 + 0^2 + 1^2 = 4 + 0 + 1 = 5$$



Then all cell values are added together as referenced by  $\Sigma(Y^2 + N^2 + B^2)$ .

Total Instances of:	Y	N	B	$S^s$	Instances x $S^s$
95      YYY	3	0	0	9	855
18      NNN	0	3	0	9	162
3      YNN	1	2	0	5	15
324    YBB	1	0	2	5	1620
78      NBB	0	1	2	5	390
11      YNB	1	1	1	3	33
2      YYN	2	1	0	5	10
134    YYB	2	0	1	5	670
23      NNB	0	2	1	5	115
4412   BBB	0	0	3	9	39708
Sum Total					43578

$$S^s + S^s + S^s + \text{etc.} = 43578$$

In Appendix F, the number of rows = 102. The number of problem expressions = 50.

$$N_1 = 102 \times 50 = 5100$$

$$\begin{aligned} P &= (43578 / 6 \times 5100) - .5 \\ &= (43578 / 30600) - .5 \\ &= 1.4241176 - .5 \\ &= .9241176 \end{aligned}$$

### Probability of Expected Agreement (**Pe**)

$$P_e = P_Y^2 + P_N^2 + P_B^2$$

$P_Y$  = the number of times Y appears in all judgments /  $3N$

$$P_Y = 895 / 3N_1 = 895 \times 15300 = .0584967$$

$P_N$  = the number of times N appears in all judgments /  $3N$

$$P_N = 197 / 3N_1 = 197 \times 15300 = .0128758$$

$P_B$  = the number of times B appears in all judgments /  $3N$

$$P_B = 14208 / 3N_1 = 14208 \times 15300 = .9286274$$

$$\begin{aligned} P_e &= .0584967^2 + .0128758^2 + .9286274^2 \\ &= .0034218 + .0001657 + .8623489 \\ &= .8659364 \end{aligned}$$

**Fleiss' Kappa (K)**

$$K = P - P_e / 1 - P_e$$

$$\begin{aligned} K &= .9241176 - .8659364 / 1 - .8659364 \\ &= .0581812 / .1340636 \\ &= .433982 \\ &= .43 \text{ or } 43\% \end{aligned}$$