

# **Ontology-Driven Information Retrieval: a Hermeneutical Approach to Information-Seeking Behavior**

**Fabiano D. Beppler**

*Knowledge Engineering and Management (EGC) UFSC. Trindade, 88040-900,*

*Florianópolis, SC – Brazil*

*Stela Institute. Rua Prof. Ayrton R. de Oliveira, 32 Florianópolis, SC, Brazil*

*Phone: +55 48 3239-2500 Fax: +55 48 3239-2505*

E-mail: [fbeppler@stela.org.br](mailto:fbeppler@stela.org.br)

**Frederico T. Fonseca \***

*College of Information Sciences and Technology*

*Pennsylvania State University, University Park, PA 16802, USA*

*Phone: (814) 865-6460 Fax: (814) 865-6426*

Email: [redfonseca@ist.psu.edu](mailto:redfonseca@ist.psu.edu)

**Roberto C. S. Pacheco**

*Knowledge Engineering and Management (EGC) UFSC. Trindade, 88040-900,*

*Florianópolis, SC – Brazil*

*Stela Institute. Rua Prof. Ayrton R. de Oliveira, 32 Florianópolis, SC, Brazil*

*Phone: +55 48 3239-2500 Fax: +55 48 3239-2505*

E-mail: [pacheco@stela.org.br](mailto:pacheco@stela.org.br)

## **Abstract**

Information Retrieval (IR) research has made a great progress in the last few years. However, extracting valuable information from IR systems is still very time consuming. Common problems affecting users, besides uncertainty and anxiety, are deficient query definition, information overload, and poor system interactivity. In our approach we addressed this uncertainty and anxiety which are part of the search process. If the user does not know exactly what s/he wants, how can a framework help in the information seeking process? We resorted to philosophy and the hermeneutic circle of Heidegger and Gadamer to provide the principles of such a framework. In our implementation of the hermeneutic circle in an IR system users develop their ideas while browsing the information and the concepts that represent the information. We chose ontologies to implement this hermeneutic approach. Ontologies improve IR systems regarding its retrieval and presentation of information, which make the task of finding information more effective, efficient, and interactive. We called our framework Hermeneus, which in Greek means the interpreter or translator. Hermeneus works as an intermediary that facilitates the user to move from the initial state of information need to the goal state of resolution. Our system intends to be the bridge between the user's question and the answer to be found while s/he navigates in the ontology concepts and the instances of these concepts in a back and forth way.

**Keywords:** Ontology-driven information retrieval, ontologies, information seeking, hermeneutics

## **1 Introduction**

The field of Information Retrieval (IR) has made a great progress in the last few years (Lin & Demner-Fushman, 2006; Witten, Moffat, & Bell, 1999). Nevertheless, users attempting to extract valuable information from IR systems still spend a lot of time trying to fulfill their information needs (Allan, Carterette, & Lewis, 2005; Lai & Soh, 2004). Common problems affecting users are deficient query definition, information overload, and poor system interactivity. Human behavioral studies show that the information seekers are lazy usually tending to create queries with the average length of just 2.2 terms (Baeza-Yates & Ribeiro-Neto, 1999) and rarely adopting boolean expressions in their criteria (Spink, Wolfram, Jansen, & Saracevic, 2001). In addition, a large percentage of users do not go beyond the initial results (Silverstein, Henzinger, Marais, & Moricz, 1999), i.e., they have low tolerance to go deep through what was retrieved.

Information seeking requires active intervention and interaction by the user with retrieved documents and with the representations of documents and their relationships (Belkin, 1993). According to Järvelin and Ingwersen (2004), “interactive information retrieval itself functions as a special case of information seeking.” Therefore, in order for IR systems to support information-seeking behavior, they should be highly interactive so that the searches can be easier, faster, and more intuitive (Ingwersen & Järvelin, 2005b; Järvelin & Ingwersen, 2004). They should be easier because users want to obtain results with as little effort as possible; they should be faster because users are impatient and require responsiveness, mainly on the Internet (Huberman & Lukose, 1997); and they should be intuitive, because users do not want to spend time in learning how to perform complex tasks in order to get better results (Hearst et al., 2002).

Uncertainty and anxiety are also an integral part of the search process (Kuhlthau, 1991, 1993). Another problem in retrieving the desired information is in how users describe their needs. The inability to express precisely what information is needed induces users to engage in an indeterminate sequence of interactions with the IR system (Capra III & Pérez-Quñones, 2005; Kuhlthau, 1993; Marchionini, 1989; Sonnenwald, Wildemuth, & Harmon, 2001). Belkin and Croft (1992) describe such a “problematic situation” as “an anomalous state of knowledge, which prompts the person to engage in active information-seeking behavior.” Marchionini (1989) says that information-seeking process “includes recognizing and interpreting the information problem, establishing a plan of search, conducting the search, evaluating the results, and if necessary, iterating through the process again.” According to Ingwersen and Järvelin (2005a), the cognitive model of users may be “dynamic but not self-contained.” The user’s state of knowledge changes by virtue of engagement with retrieved information which by its turn reflects in some change in the anomalous state of knowledge (Belkin, 1993).

But what people do when they are looking for information? How do people understand? How do people change their minds? How do people know that they found what they were looking for? What compels them to look for more? In trying to find an answer for these questions in an information seeking context, we resorted to philosophy. More specifically, we used Heidegger and Stambaugh (1996) and Gadamer’s (1989) hermeneutic circle as the main principle behind the IR system we describe in this paper. The hermeneutic circle is the concept created by Heidegger to explain how humans understand. Before approaching any subject or question, we need to have an idea (the fore-structure) of what the thing is beforehand. In the hermeneutic circle we go back and forth between the thing itself and its concept till we finally grasp its meaning.

The information seeker always has a question in mind even when he or she does not know exactly what it is. As Heidegger puts it “How is one to give a name to what he is still searching for? To assign the naming word is, after all, what constitutes finding” (1982 p.20). Our system was designed to help the user go through this process changing along with the user’s ideas and concepts. So, instead of investing in query composition, we invested in developing a structure that lets the user build his/her own query. Furthermore, this query building is not static but dynamic following the hermeneutic circle principles. In our approach the user can see partial results, can go back and forth between concepts in an ontology and instances of these concepts, and can change modes of visualization based either on concepts or instances of the concepts.

The architectural option we chose to implement this hermeneutic approach to information retrieval was to use ontologies. Ontologies improve IR systems regarding its retrieval and presentation of information. Ontologies can enhance the search process making the task of finding information more effective, efficient, and interactive. After the initial proposal by McGuinness (1999), the use of ontologies in information retrieval has not been developed as it should have been. In our work we follow some of the ideas of Guarino’s ontology-driven information systems (1998). Ontologies, according to Guarino (1998), refer to an engineering artifact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words. As an engineering artifact, ontologies may be used to generate or validate information systems components (Fonseca, 2007).

We called our framework Hermeneus following Kuhlthau (1991; 1993) who considered that the IR system is one intermediary that should facilitate the user to move from the initial state of information need to the goal state of resolution. Hermeneus in Greek means the interpreter or

translator (Kingsley, 1993). Our system is intended to be the bridge between the user's question and the answer that s/he will find in the back and forth between the concepts and the instances.

Next, we describe the hermeneutic circle and how users can get insights during their search process using our approach (section 2). In section 3, we show how ontologies can be used to implement the hermeneutic circle in an information retrieval system. The prototype that shows the feasibility of our proposal is presented in section 4. We discuss our approach describing the main functionalities and presenting some issues related to interactive IR systems based on ontologies in section 5. In section 6, we describe similar works that were important in the development of our approach. Finally, we present our conclusions and an outlook to future work.

## **2 Philosophical foundations: the metaphor of the hermeneutic circle applied to the information seeking process**

The original meaning and scope of hermeneutics as the art of interpretation of texts was greatly extended by the work of Heidegger and Gadamer. Heidegger says “the expression ‘hermeneutic’ derives from the Greek verb *hermeneuein*. That verb is related to the noun of *hermeneus*, which is referable to the name of the god Hermes by a playful thinking that is more compelling than the rigor of science. Hermes is the divine messenger. He brings the message of destiny; *hermeneuein* is that exposition which brings tidings because it can listen to a message” (Heidegger, 1982 p.29).

We think hermeneutic play is the context in which the user of an IR system will come to new insights about the information s/he is browsing. To be effective, designers of IR systems must structure a context that allows users to engage in the spontaneity of interpretive play. In our framework for information retrieval we took into account the four main components of

hermeneutics and used each one of them to support the components in our framework, as can be seen in Table 1.

<b>Components of hermeneutics</b>	<b>Components in the system</b>
Fore-structure	Computational ontology
Thing in itself	Instances
Play	User interaction with instances and computational ontology
Application	The user's question

Table 1. Components of hermeneutics related to components of the system

First, there is the *fore-structure* (which Gadamer calls prejudice) which is the initial idea that the user has about the subject. This is the first instrument that a user has to start browsing a data set. In our framework fore-structure corresponds to computational ontologies. Second, there is the *thing-in-itself* or things themselves. This corresponds to the objective world, the facts we have; the data we have gathered. In our framework, these are the instances of the concepts in the ontology. Finally, there are *play* and *application*. Here we consider that the user interaction with the instances and with the ontology is what represents *play*. Users go back and forth between the two (instances and concepts) always having an *application* in mind.

Heidegger says that the process of understanding involves a continual back and forth dialogue between the dimensions of analysis and synthesis aimed at apprehending the whole, and the dimensions that aim at apprehending the parts. We can never avoid the fact that one comes to the project of understanding with assumptions about the whole, and/or the parts. But one can, and must, reflect on those presuppositions, while keeping the object of inquiry, “the thing itself” clearly in view. Heidegger says that our “first, constant, and last task is not to let fore-having, fore-sight, and fore-conception be given to it by chance ideas and popular conceptions, but to guarantee the scientific theme by developing these in terms of the things themselves” (1996 p.143). The process of understanding something new is “a continuous dialectical tacking

between the most local of local detail and the most global of global structure in such a way as to bring both into view simultaneously” (Geertz, 1979) cited in (Bernstein, 1983).

Heidegger says that “every interpretation which is to contribute some understanding must already have understood what is to be interpreted” (1996 p.142). Heidegger understands that this proposition may be seen as a vicious circle. “But if interpretation always already has to operate within what is understood and nurture itself from this, how should it then produce scientific results without going in a circle, especially when the presupposed understanding still operates in the common knowledge of human being and world?” (Heidegger & Stambaugh, 1996 p.143) But Heidegger firmly denies that this is a vicious circle. He says that “the fulfillment of the fundamental conditions of possible interpretation rather lies in not mistaking interpretation beforehand with regard to the essential conditions of its being done” (Heidegger & Stambaugh, 1996 p.143). Furthermore, “this circle of understanding is not a circle in which any random kind of knowledge operates” but it is instead the expression of our most fundamental and intuitive knowledge. So following Heidegger’s suggestion of “what is decisive is not to get out of the circle, but to get in it in the right way;” we built Hermeneus as one of the ways of getting into the hermeneutic circle and enabling the information seeker to achieve what s/he was looking for.

To Heidegger’s description of the hermeneutic circle, Gadamer (1989) has added two crucial elements: application and play. Instead of conceiving the process of understanding in terms of analysis and synthesis alone, he has shown the importance of the dimension of application to that process. Understanding takes place in a context in which the concepts involved enter into some sort of practice. This requires a judgment about the implications for alternatives possible in local situations. Our whole framework tries to give users these alternatives mentioned by Gadamer.

Gadamer has also pointed to the inevitably playful, to and fro, nature of the hermeneutic process. In play, the participant is no longer trapped in a subjectivism in which the play is a kind of predicate over against the participating subject. Instead, the player loses him or herself in the play, becoming, as it were, the predicate of the play. The player's movements and decisions are constrained to conform to the regenerative, to and fro patterns of the play.

One of the foundations of Gadamer's work in hermeneutics is the connection between the theoretical and the practical. Hermeneutic understanding is not a selfless, purposeless activity. For Gadamer, interpretation, understanding, and application are all linked together in a unity. These are not independent events in time but they happen simultaneously, with one influencing and actually enabling the other. Gadamer starts saying that "interpretation is not an occasional, post facto supplement to understanding; rather, understanding is always interpretation, and hence interpretation is an explicit form of understanding" (Gadamer, 1989). Understanding and interpretation are applied to a text in Gadamer's example, and applied here to a specific subject domain which is the target of the user's initial question. The task of the interpreter is to adapt the text to a concrete situation to which the text is speaking. The same is valid here in the process of understanding a subject through the process of going back and forth between the concepts and instances. Understanding is a dynamic process and meaning is always being created and recreated. The objective of our framework is to give the user the medium to achieve this kind of understanding. The metaphor for the main user activity of going back and forth between instances and concepts in trying to shape his/her own questions is the hermeneutic circle.

In the next section we start making the shift from the philosophical foundations of our framework to its computational aspects. The main element in the translation of the hermeneutic

circle to an IR interface is the ontology. In our approach ontology corresponds to the fore-structures and has the role of anchoring the whole information seeking process.

### **3 Ontology-Driven Information Retrieval**

Information search process is the user's constructive activity of finding meaning from information in order to extend his or her state of knowledge on a particular problem or topic (Kuhlthau, 1991). This activity is consistently interactive because users are constantly expanding their state of knowledge along with the search process. Based on such issues, it is clear that IR systems should provide mechanisms to facilitate users' interaction. McGuinness (1999) states that ontology with standard IR tools can be used to create enhanced search capabilities. She emphasizes that "the challenges are to provide usable navigation tools, user expectation setting, and most importantly, to provide a friendly and usable search function." Ontologies can also be used to provide more accurate answers to queries because with ontologies knowledge is formally represented and users are able to compose semantic queries (Decker, Erdmann, Fensel, & Studer, 1999; Lei, Uren, & Motta, 2006; Müller, Kenny, & Sternberg, 2004). Furthermore, as ontologies include axioms to express relationships between concepts, it is possible to derive information which has been specified only implicitly (Decker et al., 1999).

We took into account the above issues and developed a framework called Hermeneus that uses ontologies to enhance users' interactivity with IR systems. The main component of our approach is the ontology. The domain specialist has the responsibility to define the meaning of terms that represent a domain and codify this vocabulary in an ontology. The ontology then is the foundation for creating a knowledge base composed of instances. Each instance represents a materialized view of a class defined by the ontology (e.g., given a professor's class, a specific professor is an instance of such a class). Therefore, our approach takes advantage of the ontology

itself so as to avoid domain specialist concerns with implementation details related to the IR system design (Fonseca, 2007; Guarino, 1998).

Ontology is a technology that can give support for the development of more intelligent systems (Sowa, 2002). The *intelligent* concept, in our approach, means to use the ontology to assist users to compose more valuable queries, improve the accuracy of retrieved information, and extract additional information that is not explicitly stated. All such features are in accord with information-seeking research that states the importance of interactive tools to support users during the search process. In an information-seeking process, according to Belkin (1993), Kuhlthau (1993), Marchionini (1989), and Ingwersen and Järvelin (2005a), active intervention and interaction are required by the user. In our solution we implemented a functionality that is in accordance with García and Sicilia (2003) theory, which states that “IR applications should explore new query formulation paradigms that enable interacting with users to a higher extent than current query-formulation interfaces.”

In our approach, the user is able to navigate between the instances and concepts of the ontology in a back and forth fashion. This way, the results set has a more active role during the retrieval and browsing process due to the use of the semantic value included in each instance. The retrieved information is not presented as simple textual documents but as ontology instances. As ontologies can be composed of more than one class, there can be relationships among instances. For instance, the “*The double role of ontologies in information science research*” instance, defined by the *paper* class, is connected by the relationship “*hasAuthor*” to the “*Frederico Fonseca*” instance, defined by the *author* class.

Ontologies can become a means of communication between users and the system and help overcome the bottlenecks in information access, which is primarily based on keyword searches

(Shah, Finin, & Joshi, 2002). In our case, each concept defined by the ontology can be used to formulate semantic queries. Ontology thus can guide users to compose queries and select appropriate terms that are likely to lead to nonempty answer sets. As users are able to explicitly define terms for specific concepts, they are composing more accurate queries automatically. It means that the ontology is also used to disambiguate query terms. Complex queries, using more than one concept, for instance, can also be easily built due to the facility of choosing concepts and typing the corresponding terms. For instance, selecting the “*author*” concept and typing the term “*fonseca*”, then selecting the “*paper*” concept and typing the term “*information systems*”. Figure 1 describes in EBNF the format of queries applied to a search.

```

search ::= query | semantic_query
query ::= term { term }
semantic_query ::= concept ":" term { term } ";" { concept ":" term { term } ";" }
concept ::= <classes described in the ontology>
term ::= letter { letter }- | digit { digit }-
letter ::= "a" | "b" | ... | "z"
digit ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"

```

Figure 1. EBNF describing a query in a search

In the *ontology navigator module*, where the ontology is graphically available, users can interact with ontology concepts and visualize how they are related to each other. In addition, users can select a *central concept* in the ontology to designate which kinds of instances are being retrieved (e.g., if *central concept* is *paper*, the main retrieved instances are papers). Besides information about the *central concept*'s instances, instances of directly related concepts to the *central concept* are also presented. So, for example, in an ontology with three concepts, *paper*, *author*, and *date*, where there is a relation between the *paper* and the *date* concepts but not between the *author* and the *date* concepts, and the *central concept* is *date*, only instances about *date* and *paper* are presented.

The use of ontologies in IR systems gives also an additional possibility which is the capability to use inference techniques to derive additional knowledge. This is possible because

the ontology is described in a formal language that includes axioms for specifying relationships between concepts. Hermeneus was conceived to allow the domain specialist to take advantage of such inference techniques and configure specific rules to be used during the search process. Therefore, besides the ontology's graphical visualization and the list of retrieved instances, users can also visualize and interact with additional information for a given retrieved instance. For example, selecting a retrieved instance about a specific paper, the *additional information module* could show, for instance, similar papers, authors with papers in the same area or papers related to the authors' institutions.

The interface takes advantage of the ontology and the distinct kinds of retrieved information, focusing on a layout that induces easy interaction. Figure 2 shows, on the left side, that the ontology is presented graphically in the *ontology navigator module*. In the middle, the *retrieved instances module* is responsible for presenting the instances that are retrieved according to the users' queries and the chosen *central concept*. Moreover, as instances hold semantic description, each retrieved instance can be used to compose semantic queries automatically. On the Figure 2's right side, the *additional information* is shown. This module depends on the ontology and on the knowledge base. The inference rules use concepts and relationships in the ontology and apply them to the knowledge base. The results are displayed and the user can interact with them and create new or modify existing queries.

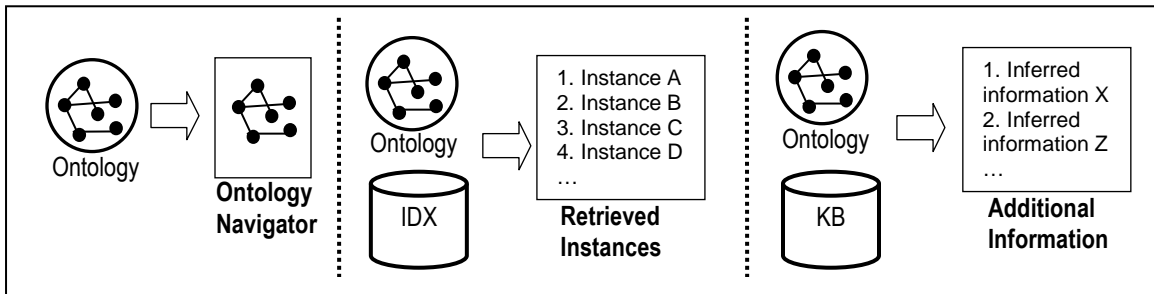


Figure 2. Modules available for user's interaction

Furthermore, because the ontology drives the whole process, all the three modules (*ontology navigator*, *retrieve instances*, and *additional information*) are strongly integrated. Therefore, actions in a module may affect the others. For example, if the user selects a piece of semantic content from a retrieved instance in the *retrieved instances module* to refine a query, the *ontology navigator module* highlights automatically the concept associated to such content. In addition, the user can select a new retrieved instance in the *retrieved instance module* so as to visualize additional information; thus, the *additional information module* interprets the selected instance and extracts new content based on the ontology and the instances directly related.

#### 4 Prototype description

In order to validate our approach, it was implemented a prototype. We created a database using 622 papers from the Journal of the American Society for Information Science and Technology (JASIST). Each instance contains information about the papers including title, authors, institutions, issues, and venue. Figure 3 shows an instance example that can be retrieved from the website of the Journal of the American Society for Information Science and Technology.

<p><b>AU:</b> Frederico T. Fonseca, James E. Martin  <b>TI:</b> Toward an alternative notion of information systems ontologies: Information engineering as a hermeneutic enterprise  <b>SO:</b> Journal of the American Society for Information Science and Technology  <b>VL:</b> 56  <b>NO:</b> 1  <b>PG:</b> 46-57  <b>YR:</b> 2005  <b>CP:</b> Copyright © 2004 Wiley Periodicals, Inc.  <b>ON:</b> 1532-2890  <b>PN:</b> 1532-2882  <b>AD:</b> School of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16802; Psychology Department, Pennsylvania State University, University Park, PA 16802  <b>DOI:</b> 10.1002/asi.20099  <b>US:</b> <a href="http://dx.doi.org/10.1002/asi.20099">http://dx.doi.org/10.1002/asi.20099</a>  <b>AB:</b> In this paper we discuss the construction of information systems ontologies. We summarize and discuss Barry Smith's review (2003a) of the field in the paper...</p>
---

Figure 3. Example of paper's citation

Based on the content of such citations, it was conceived an ontology composed of five classes with relations among them. The classes and their respective properties represent the concepts of *paper*, *author*, *institution*, *journal*, and *keyword*. Table 2 shows the ontology classes and their respective properties. Figure 4 presents the ontology with its concepts and their respective relationships. We created this ontology with the purpose of showing the feasibility of our approach. Any discussion about ontology structure and definition is beyond the scope of this paper.

<b>Class</b>	<b>Properties</b>
Paper	Identifier, Title, and Abstract
Author	Name
Institution	Description
Journal	Volume, IssueNumber, and Year
Keyword	KeywordName

Table 2. Ontology classes and their respective properties

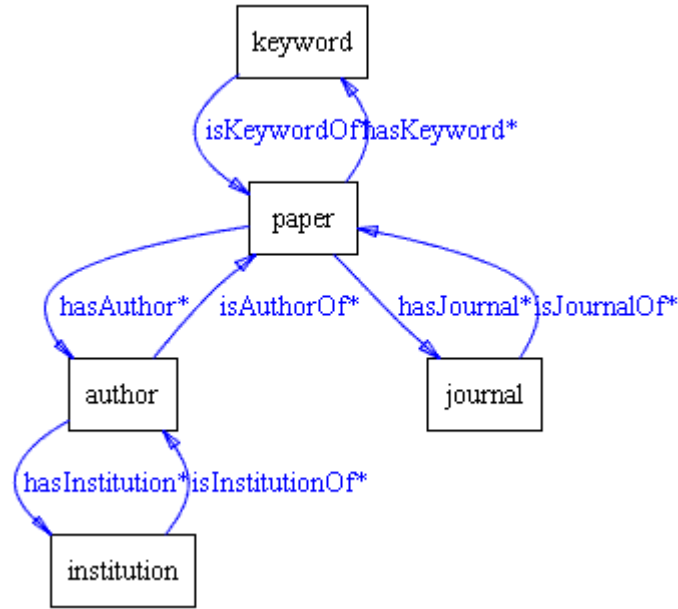


Figure 4. Ontology: classes and their relations

The knowledge base then was built by using the instances of the ontology. The properties of *paper*, *author*, *journal*, and *institution* classes were filled with respective fields extracted from the journal’s website. The property of *keyword* class, however, was filled with information (i.e., entities) extracted from the *title* and *abstract* fields using the Information Extraction technique called Named Entity Recognition – NER (Mikheev, Moens, & Glover, 1999).

It is important to emphasize that concepts are defined as classes and the respective properties (Noy & McGuinness, 2001) – properties, also known as slots, of each concept describes its features and its attributes (e.g., *author* concept is defined by the *author* class and the *name* property). Moreover, there can be concepts that comprise subclasses and in some cases more than one class – in this case the relations between classes are also important. In our ontology, each concept is described by a specific class. Therefore, an instance of a concept refers to the value of respective class’ properties (e.g., the content of *paper* instances refers to *title*, *abstract*, and *identifier* properties, described in *paper*’s class).

The IR system's interface is composed of two screens. The first screen, presented in Figure 5, shows a text field and the *ontology navigator module*. Users can type terms directly in the text field and interact with the ontology. When interacting with the ontology, users can define the *central concept*, which indicates which kind of information should be retrieved. When a user selects the *author* concept as *central concept*, for example, the instances about authors and their respective papers and institutions are shown, because according to our ontology the concepts *paper* and *institution* are direct related to the *author* concept (see Figure 4).

In addition, when clicking in a concept, in the *ontology navigator module*, the word that represents such a concept is inserted in the text field following by the colon symbol (e.g., clicking on the *author* concept, the word “*author:*” is inserted in the text field). Users can then type terms for specific concepts (i.e., composing semantic queries). Additionally, the combination of basic and semantic queries is allowed. In a basic query, which is composed of terms without context, all instances that contain the typed terms are retrieved (e.g., search for “*artificial intelligence*” means that all instances with these two terms are retrieved regardless where such terms are described); whilst in a semantic query, only instances that contain the concepts with their respective terms are retrieved (e.g., the semantic query “*author:fonseca*” retrieves only instances that contain the term “*fonseca*” specifically described in the *author* concept).

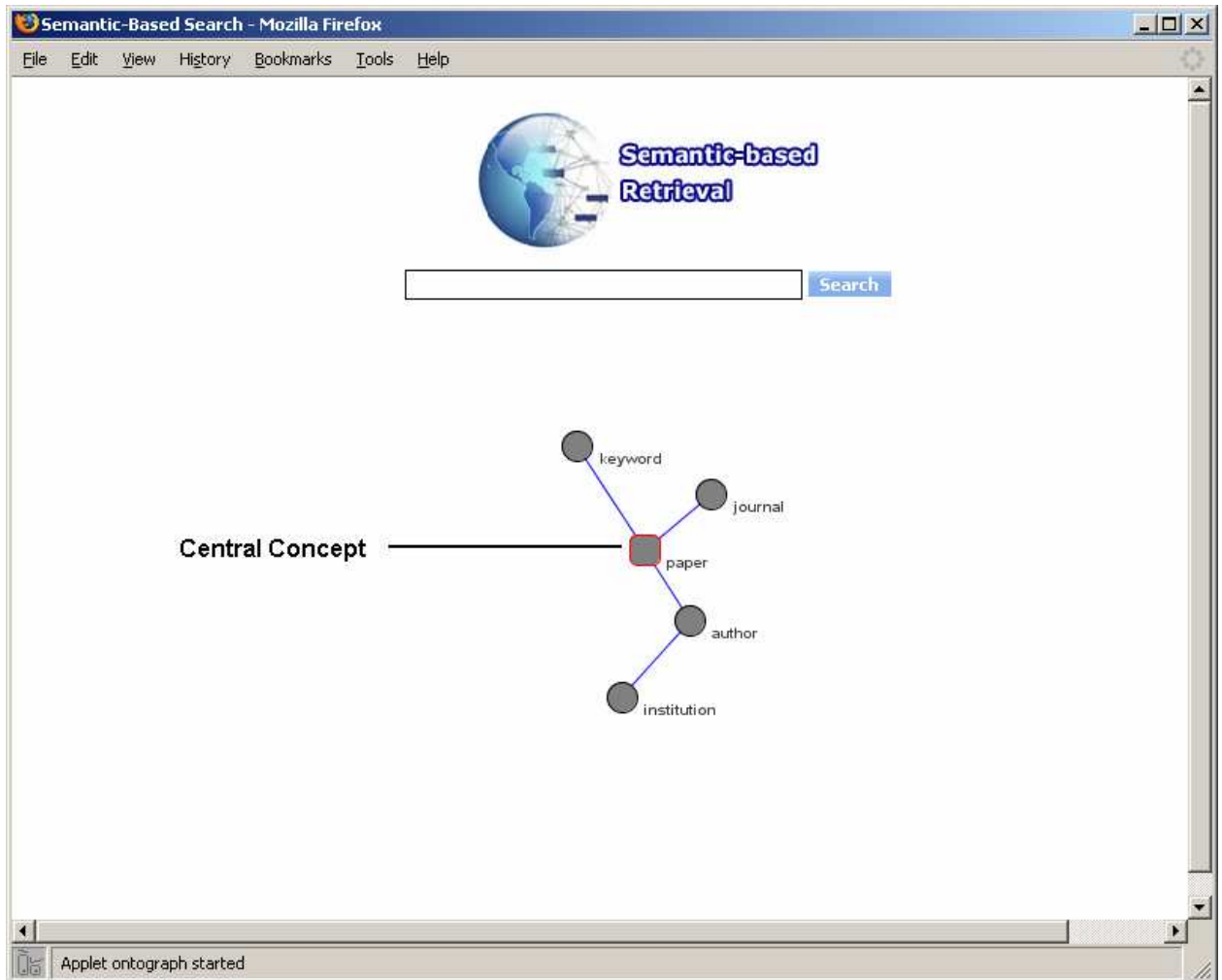


Figure 5. Prototype's initial screen

After the first search, the result window is shown (Figure 6). This window is divided basically in four sections: the first one makes a text field available to receive the query; the second section shows the *ontology navigator*, with the ontology presented graphically; the third section presents the list of retrieved instances; and the fourth section presents additional information about an instance. In Figure 6, as a result from our query (*ontology author:fonseca*), two *paper* instances were retrieved. In addition, instances of directly related concepts (*author*, *journal*, and *keyword*) are also presented. In the *ontology navigator module*, the *central concept* has a different shape (i.e., square) and the concept (i.e., *author*) used in the query is marked in

yellow. If the user clicks on the first retrieved instance additional information is displayed in the fourth section. Then the information available in *additional information module* will show that the author “*Frederico Fonseca*” has as main keywords “*ontologies*”, “*information science*”, and “*engineering*”, and cooperates with “*Gilberto Câmara*” and “*James E. Martin*”. Moreover, it presents the author’s institution, “*Pennsylvania State University*”, describing that such university has 18 papers published by 36 authors. It will also be shown that the institution’s main keywords are “*web search*”, “*web search engine*”, “*information retrieval*”, and “*ontologies*”.

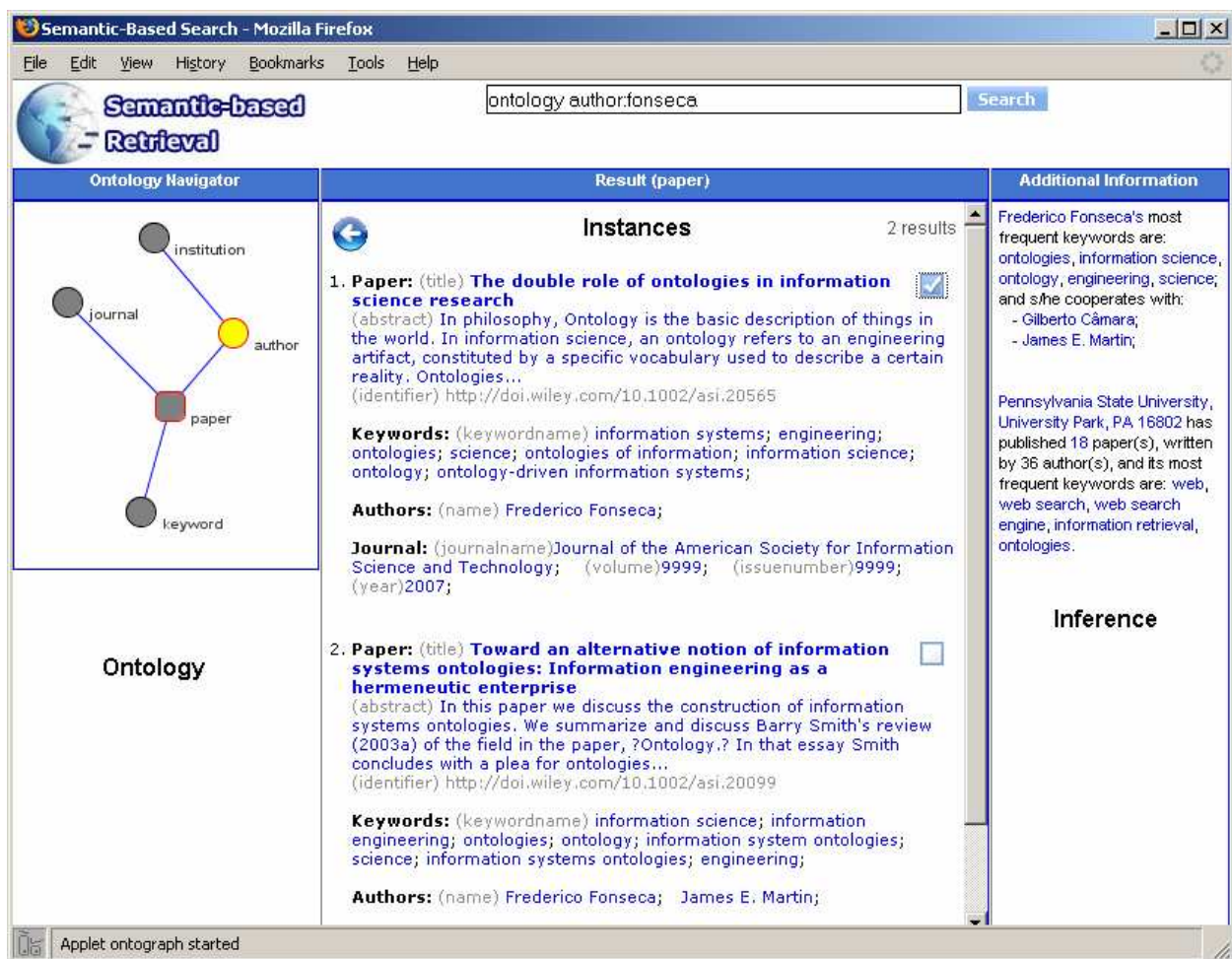


Figure 6. Prototype's result screen

Users can interact with all retrieved content available in the *retrieved instances* and *additional information* modules. In the *retrieved instances* module, when users select an instance

a new search is performed considering such content, like a refinement process. In the *additional information* module, this function may work differently, because the domain specialist is able to define specific behaviors for extracted information. For example, in our prototype, if we click in the amount of papers (i.e., eighteen), available in the *additional information module*, related to *Frederico Fonseca's* paper, we can automatically visualize eighteen papers in the *retrieved instance module*, published by authors linked to the *Pennsylvania State University* (i.e., the affiliation of author *Frederico Fonseca*).

## **5 Discussion**

Our approach uses an ontology to assist users throughout the information seeking process. In the following sections we discuss how our system can help users during the process and how the tools available provide support for users to develop and extend their state of knowledge about a specific information need.

### **5.1 Hermeneus as the translator of the user's information needs into semantic queries**

One of the main focuses of our system is to help users understand and develop their own information needs. Although it may seem awkward that users do not know exactly what they are looking for at first, the literature deals with this uncertainty called knowledge gap or state of incompleteness. It is this gap that drives users to use intermediaries, such as IR systems, to supply their information needs. Users, however, have to develop their own information needs first so as to be able to figure out what they are looking for. In this process of trying to understand their own needs, the users' cognitive state may contain doubts, problems in interpretation and, according to Ingwersen and Järvelin (2005a), reach a state of uncertainty. Belkin (1993) states that users normally have a lack of knowledge, or an anomalous state of knowledge, about the subject they are looking for. It is our opinion that both the process of

understanding the lack of knowledge and the searching for information are cognitive tasks narrowly related to each other during the whole cycle of information acquisition. It reinforces the fact that IR systems should allow users to interact with tools that can affect their state of uncertainty in order to stimulate a transformation in their state of knowledge (Hert, 1997; Ingwersen & Järvelin, 2005a; Kuhlthau, 1993; Marchionini, 1989). Hermeneus was designed with the purpose of giving the users the necessary support for them to develop and understand their own information needs. The hermeneutic circle was the principle behind the system.

When interacting with an IR system the most common way to express information needs is composing queries. Nevertheless, formulating precise queries is not an easy task. Precise queries can make the difference between good and bad results from a search (Chen, Shankaranarayanan, & She, 1998). We tried to address these issues allowing users to compose semantic queries. Since users are able to interact with the ontology, they can identify which concepts are available and use them to formulate queries. A semantic query basically is composed of the concept's name and the desired search terms. In our prototype, users can interact with the ontology through mouse-clicks; when they click on a concept it is automatically added to the current query's content so as to assist the composition of semantic queries. For example, a search for the term "*data modeling*" (i.e., the query contains just the "*data modeling*" terms) combined with a click on the concept *author* available in the ontology, adds the description of such a concept to the query with the colon symbol (i.e., the query now consists of "*data modeling author:*" content) – so the user may just insert the term specifically related to the *author* concept without needing to type the concept's name (e.g., "*data modeling author:Fonseca*"). In developing our solution we followed the principle that through the interaction with the ontology, users get insights about their needs and express themselves with more useful queries, and thus obtain results more

precisely. Additionally, because users can define explicitly the context of terms, they can formulate less ambiguous queries (García & Sicilia, 2003; Hyvönen et al., 2005).

As described by Kuhlthau (1991) and Hert (1997), the user's state of knowledge is dynamic rather than static, changing as s/he proceeds during the searching process. In our approach, since we retrieve instances and present them in a highly contextualized environment, and each piece of retrieved information has a semantic link with the concepts in the ontology, users can improve their queries by adding semantic content from relevant instances automatically. For example, in a very common situation where a user types just a few terms (e.g., "*semantic*") to express his/her need, the retrieved instances can help him/her to find a more precise description and use such description dynamically (e.g., in our prototype when searching for the term "*semantic*", users can visualize and interact with the following terms related to the *keyword* concept: *latent semantic indexing*, *semantic network*, *semantic information*, *semantic-based information retrieval*, *semantic-based web retrieval systems*, *semantic web*, and *semantic search*). With this kind of exploratory interaction users are shielded from details of query formulation besides being able to construct more useful search statements (Capra III & Pérez-Quñones, 2005). For example, searching for the term "*web*" in our prototype, we received 169 retrieved instances; but, clicking on "*semantic web*" content, presented in one of the retrieved instances, related to the *keyword concept*, and adding this new information to our query (i.e., our query now is composed of "*web*" term and "*keyword: semantic web;*" semantic content), gave as a result only 11 instances. With this interactivity, users can easily experiment with retrieved content to formulate new queries as part of the process of understanding and developing their information needs.

The ontology can also be used to define which kind of retrieved instances should be shown through the selection of a *central concept*. This functionality enables users to have different

perspectives of retrieved instances for the same query. Ontologies can be used to derive further views of knowledge, particularly in the exploration of new navigation purposes (Staab, Studer, Schnurr, & Sure, 2001). Since the process of finding information is dynamic and constantly changing (Ingwersen & Järvelin, 2005a; Kuhlthau, 1993; Marchionini, 1989; Vakkari, 2003), visualizing retrieved information with different perspectives may help users get insights about their needs. In our prototype, for example, when searching for “*information science*”, there were 58 retrieved instances grouped by the *paper* concept (i.e., the *central concept* is *paper*). After changing the visualization of the results set to be grouped by the *author* concept (i.e., the *central concept* was changed to *author*) there were then 73 instances of authors. In this case, the user was able to visualize the results set with authors and their respective information (i.e., instances of papers and institutions for each retrieved author). Such functionality is in accordance with Vakkari (2003) theory, who quotes that “modeling and developing systems features, like features of interfaces that support users in articulating their information needs by using terms from potentially relevant documents grouped according to the concepts in the users' queries reflecting the task, would enhance retrieval performance.”

## **5.2 Extracting additional information**

During the process of searching for information users tend to focus on details that can help them find what they are looking for. In traditional IR systems, for instance, users receive a result list with just a small part of retrieved documents – normally it is not enough to make the document's meaning understood. In our approach, the *additional information module* enables users visualize information that is not explicitly stated in the result list of retrieved instances. Additional information helps users have a better understanding about the domain and facilitates them to learn more about a specific instance (Golbreich, 2004; Murdock, Pinheiro da Silva,

Ferrucci, Welty, & McGuinness, 2005). Therefore, inferring new information for each retrieved instance enables users to get insights about their information needs.

Inference, according to Decker et al. (1999), is the logical process of derivation of new data from a collection of data. We use this technique combined with rules created by domain specialists to extract information that was only implicitly expressed before. In our prototype, for example, when the retrieved instances are grouped by the concept *paper* (i.e., the *central concept* is *paper*) and users pick out a specific instance, the *additional information module* presents inferred information about authors and their institutions. In this case, as shown in Table 3, it is presented the authors' most frequent keywords, considering all papers stored in the knowledge base, the institution to which the authors are affiliated with, and the authors' collaborators. In addition, users can see the most frequent keywords for each institution and the papers that mention such institution (Table 3). The information available in the *additional information module* can also be used for additional exploratory activity, because users can interact with such information and refine queries dynamically. This interactive process has the same principle available in the *retrieved instances module*.

Retrieved instance (paper)	Additional Information
<p>Title: The double role of ontologies in information science research</p> <p>Author: Frederico Fonseca</p> <p>Keywords: information systems; ontology; ontologies; engineering; science; information science; ontology-driven information science;</p>	<p>Frederico Fonseca is connected to Pennsylvania State University, University Park, PA 16802, his most frequent keywords are: ontologies, ontology, information science, engineering, science; and s/he cooperates with:</p> <ul style="list-style-type: none"> <li>- Gilberto Câmara;</li> <li>- James E. Martin;</li> </ul> <p>Pennsylvania State University, University Park, PA 16802 has published 18 paper(s), written by 36 author(s), and its most frequent keywords are: web, web search, web search engine, information retrieval, ontologies.</p>

Table 3. A sample of the additional information column

### 5.3 Ontology graphical interaction

The aim of a graphical tool is to help human beings gain insight into data. Visualizing an ontology means to view its concepts and respective relations (Spence, 2000). The added value of an ontology’s graphical presentation lies in its expressivity because concepts and their relationships (the vocabulary of the domain) become easier to detect (Fluit, Sabou, & Harmelen, 2002). Our approach takes advantage of this and enables users’ interaction with the ontology during the whole process of searching for information. The interface of the *ontology navigator module* has been designed to stimulate the user to explore the information shown. It can assist users in understanding their knowledge gap and then specify their information needs more precisely.

Using Hermeneus, users can visually identify which concepts were used in a search. This functionality marks all concepts belonging to a semantic query with a distinct color. For example, searching for “*author: Pacheco*,” the concept *author* has a different color (yellow) in the ontology. Users can also view what information is specifically related to each concept in accordance with the query. Considering the prior example, the *author* box in the ontology shows

“*Pacheco*” as its content. These kinds of graphical features help users to contextualize their needs at a time that their understanding of the problem is still developing.

The graphical tool also enables users to use the ontology to have different perspectives of the results set. This functionality has already been described before, but it is important to emphasize that such changes in the results set’s perspective is done through mouse-clicks on the concepts available graphically. The type of information the results set is showing is related to the *central concept*. The concept used as the *central concept* is drawn with a different shape (square) in order to facilitate visual identification of which type of information is being presented.

## **6 Related work**

Although there are some works applying ontology in IR systems, none of them addresses all the points we do. For instance, in MuseumFinland (Hyvönen et al., 2005; Hyvönen, Saarela, & Viljanen, 2004) is presented a semantic view-based search engine that uses ontology to define categories to be inserted in a multi-facet view. The facets are built based on taxonomies and are exposed to the end-users so as to provide a broad view of the repository content. However, multi-facet view requires a hierarchic structure in order to organize information in its facets or categories. This approach works well when the domain knowledge can be modeled as a taxonomy. In most cases, though, an ontology is much more than a taxonomy of concepts. The best ontology representations have particular constraints and interrelations among concepts (Guarino, 1997). Furthermore, multi-facet view does not allow for combination of distinct concepts belonging to different levels in the hierarchy in the same query. Our approach does not offer restrictions with the type of ontology and is flexible enough to allow users configure more complex queries using any concept in the ontology.

McGuinness (1999) developed the initial idea of ontology driven information retrieval. Her work presented an ontology-enhanced online search for medical documents. Users can combine a simple keyword search with the choice of one of the content areas displayed in the same window. The content areas are presented in a tree structure as a taxonomy where users are able to navigate and choose one of them. Again, this approach takes advantage of a hierarchical structure and limits the composition of more complex queries relating concepts from different parts of the hierarchy. It is not possible to link one keyword to a specific concept (i.e., the typed keywords are applied to all selected categories). Furthermore, the results screen shows only the retrieved records and does not present the ontology associated with the information used in the search. Thus it is harder for the user to contextualize the results in the domain knowledge. In our approach, users are allowed to use the ontology as a guide to compose queries and define explicitly for each concept its specific terms (e.g., *author: fONSECA paper: ontology*). In addition, in the results window, users can still interact with the ontology and see which concepts were used in the current query, besides being able to interact with the retrieved instances to narrow down the scope of their queries more dynamically.

*Textpresso* (Müller et al., 2004) is an ontology-based information retrieval and extraction system. The system has an ontology defined as categories and subcategories used to support the searches on biology literature. The retrieval system is composed of two user interfaces in order to support users with different levels of knowledge. It is possible to create queries using a limited combination of categories, subcategories, and keywords. Nevertheless, users cannot visualize the entire list of categories and subcategories organized in an easy way to navigate and choose one or more of them when configuring the query. Additionally, categories and subcategories are normally related to taxonomies which do not represent the majority of ontology structures. Our

approach is not restricted to specific types of ontologies and allows a clear visualization of the concepts and relations described in the ontology structure. In our system, basic and advanced queries can be built using the same tools, because users can configure semantic queries with ontology concepts, and they can define as many keywords as needed for as many concepts as available in the ontology. Finally, our approach is not restricted to a specific domain thus being able to be applied to any domain.

*SemSearch*, described in Lei et al. (2006), proposes a semantic search that intends to hide the complexity of user queries in order to overcome the problem of knowledge overhead. Using a Google-based interface, according to the authors, allow the explicit specification of query subject and the combination of keywords. Indeed, we also consider that a simple place (i.e., text field) to type in the queries keeps the system easy to use. Nevertheless, to use semantic queries in their proposal, users need to know the subject of their necessity and use specific words to define such subject along with keywords. In our work we make the ontology graphically available to users as a guide to compose semantic queries, so they do not need to guess which concepts are defined in the knowledge domain. The ontology, in our case, is also used to enable users to have different perspectives for the same search. Moreover, the results set has an active role in the searching process in opposition to *SemSearch*, which presents just passive information.

## **7 Conclusions and future work**

In our approach we addressed the uncertainty and anxiety which are part of the search process. If the user does not know exactly what s/he wants, how can a framework for information seeking help in this process? We looked into philosophy of Heidegger and Gadamer to find the principles that guided our solution. We tried to implement the hermeneutic circle as the support

the information seeking process. In the hermeneutic circle users develop their ideas as they browse the information and the concepts that represent the information.

The solution we presented here was the framework called Hermeneus. We choose the name Hermeneus, which in Greek means the interpreter or translator (Kingsley, 1993), following Kuhlthau's (1991; 1993) concept that an IR system is an intermediary between the user's need for information and the achievement of the resolution of that need. Information seeking theory helped us understand that the user requires an environment where s/he can intervene and interact more actively. Finding information is an exploratory activity where users apply their knowledge and intuition combined with strategies and tools to find the desired information (Capra III & Pérez-Quñones, 2005). Our framework enabled users to engage in a sequence of interactions that helped them understand what they were looking for.

Hermeneus is based on ontologies in order to enhance the interactivity with the user. We used a dynamic process for query building so that users could browse the ontology concepts and the instances of these concepts in a back and forth way. Doing so users could also change modes of visualization based on each concept described in the ontology. Users could also interact with retrieved instances so as to use their contents to begin new searches automatically. Furthermore, additional information could be extracted for each retrieved instance. Users were able to visualize and interact with knowledge which before was just implicit stated in the ontology and in the knowledge base.

Hermeneus addressed Järvelin's (2004) and Ingwersen's (2005b) requisites of a search activity that is easier, faster, and intuitive. In terms of easiness, our approach allowed users to access the domain knowledge through keywords search and to interact with the ontology graphically – therefore they could compose queries using terms according to their initial state of

knowledge and get insights when navigating in the ontology concepts. Regarding fastness, because users' state of knowledge can be influenced by the retrieved information, our approach enabled users to use query refinement as an assistant to compose semantic queries dynamically and to extract additional information for each retrieved instance. Such operations are in accordance with Ingwersen and Järvelin (2005b) theory about information seeking who states that "relevance is a cognitive, situational, and dynamic phenomenon." Finally, concerning intuitiveness, users could use concepts described in the ontology to compose semantic queries, use the retrieved content to redefine their searches, and use the ontology to visualize the results set in different perspectives. These functionalities, according to Card et al. (1999) and Albertoni et al. (2005), augment user's cognition and thus reach the goal state of resolution more quickly.

As future work, we are working on a new way to present the retrieved information to take advantage of semantic content, described in the instances and the ontology. We think that a more graphical way to show the results set can enrich even more the interactivity between the users and the system. We also understand that users' traceability could be stored so as to help other users when they are searching for similar information. In this case, the system could "understand" new users' behavior *on-the-fly* enhancing the assistance throughout the seeking process and giving new clues about their searches. We also envisage the creation of a mechanism that could be used by users to define their own ontologies dynamically and thus configure the IR system according to their notion of reality for a specific domain. User, in this case, will be able to create a personalized IR system and share it with others if s/he wants.

## 8 References

- Albertoni, R., Bertoni, A., & Martino, M.D. (2005). Information Search: The Challenge of Integrating Information Visualization and Semantic Web. Sixteenth Int. Workshop on Database and Expert Systems Applications, 529-533.
- Allan, J., Carterette, B., & Lewis, J. (2005). When will Information Retrieval be "Good Enough?" Proceedings of the 28th annual ACM SIGIR, 433-440.
- Baeza-Yates, R.A., & Ribeiro-Neto, B.A. (1999). Modern Information Retrieval: ACM Press, Addison-Wesley.
- Belkin, N.J. (1993). In Interaction with Texts: Information Retrieval as Information-Seeking Behavior (pp. 55-66). Paper presented at the Information Retrieval '93: Von der Modellierung zu Anwendung.
- Belkin, N.J., & Croft, W.B. (1992). Information Filtering and Information Retrieval: Two Sides of the Same Coin? Communications of the ACM, 35(12), 29-38.
- Bernstein, R.J. (1983). Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis. Philadelphia, PA, USA: University of Pennsylvania Press.
- Capra III, R.G., & Pérez-Quiñones, M.A. (2005). Using Web Search Engines to Find and Refind Information. IEEE Computer Society, 38(10), 36-42.
- Card, S.K., Mackinlay, J.D., & Schneiderman, B. (1999). Readings in Information Visualization: Using Vision to Think: Morgan Kaufmann.
- Chen, H., Shankaranarayanan, G., & She, L. (1998). A Machine Learning Approach to Inductive Query by Examples: An Experiment Using Relevance Feedback, ID3, Genetic Algorithms, and Simulated Annealing. Journal of the American Society for Information Science, 49(8), 693-705.

- Decker, S., Erdmann, M., Fensel, D., & Studer, R. (1999). In *Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information* (pp. 351-369). Paper presented at the DS-8. Kluwer Academic Publisher.
- Fluit, C., Sabou, M., & Harmelen, F.v. (2002). *Ontology-Based Information Visualization*. In *Visualising the Semantic Web*: Springer Verlag.
- Fonseca, F. (2007). The Double Role of Ontologies in Information Science. *Journal of the American Society for Information Science and Technology*, 58(6), 786-793.
- Gadamer, H.-G. (1989). *Truth and Method* (Original publication 1960, as *Wahrheit und Methode*) (Second edition ed.): Crossroad.
- García, E., & Sicilia, M.-Á. (2003). User Interface Tactics in Ontology-Based Information Seeking. *PsychNology Journal*, 1(3), 242-255.
- Geertz, C. (1979). From the Native's Point of View: On the Nature of Anthropological Understanding. In P. Rabinow & W.M. Sullivan (Eds.), *Interpretive Social Science: A Reader* (pp. 225-241). Berkeley: University of California Press.
- Golbreich, C. (2004). In *Combining Rule and Ontology Reasoners for the Semantic Web* (pp. 155-169). Paper presented at the RuleML, Hiroshima, Japan. Springer-Verlag.
- Guarino, N. (1997). Understanding and Building, Using Ontologies. *Int. Journal of Human-Computer Interaction*, 46(2), 293-310.
- Guarino, N. (1998). *Formal Ontology and Information Systems*. *Formal Ontology in Information Systems*, Netherlands: IOS Press, 3-15.
- Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K.-P. (2002). Finding the Flow in Web Site Search. *Communications of the ACM*, The consumer side of search, 45(9).

- Heidegger, M. (1982). *On the Way to Language*. San Francisco: Harper & Row.
- Heidegger, M., & Stambaugh, J. (1996). *Being and Time: A Translation of Sein und Zeit*. Albany, NY: State University of New York Press.
- Hert, C.A. (1997). *Understanding Information Retrieval Interactions: Theoretical and Practical Implications*. London, England: Ablex Publishing Corporation.
- Huberman, B.A., & Lukose, R.M. (1997). Social Dilemmas and Internet Congestion. *Science*. American Association for the Advancement of Science, 277(5325), 535-537.
- Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., et al. (2005). MuseumFinland: Finnish Museums on the Semantic Web. *Journal of Web Semantics*, 3(2), 25.
- Hyvönen, E., Saarela, S., & Viljanen, K. (2004). In *Application of Ontology Techniques to View-Based Semantic Search and Browsing*. Paper presented at the *The Semantic Web: Research and Applications (First European Semantic Web Symposium)*.
- Ingwersen, P., & Järvelin, K. (2005a). The Sense of Information: Understanding the Cognitive Conditional Information Concept in Relation to Information Acquisition. *Lecture Notes in Computer Science*, Springer, 3507, 7-19.
- Ingwersen, P., & Järvelin, K. (2005b). *The Turn: Integration of Information Seeking and Retrieval in Context*. Netherlands: Springer.
- Järvelin, K., & Ingwersen, P. (2004). Information Seeking Research Needs Extension Towards Tasks and Technology. *Information Research: An International Electronic Journal*, 10(1).
- Kingsley, P. (1993). Poimandres: The Etymology of the Name and the Origins of the Hermetica. *Journal of the Warburg and Courtauld Institutes*, 56, 1-24.

- Kuhlthau, C.C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, 42(5), 361-371.
- Kuhlthau, C.C. (1993). *Seeking Meaning: A Process Approach to Library and Information Services*. USA: Ablex Publishing Corporation.
- Lai, J., & Soh, B. (2004). In Similarity Score for Information Filtering Thresholds (Vol. 1, pp. 216-221). Paper presented at the ISCIT 2004. IEEE International Symposium on Communications and Information Technology.
- Lei, Y., Uren, V., & Motta, E. (2006). In SemSearch: A Search Engine for the Semantic Web. Paper presented at the 3rd European Semantic Web Conference (ESWC), Montenegro.
- Lin, J., & Demner-Fushman, D. (2006). The Role of Knowledge in Conceptual Retrieval: A Study in the Domain of Clinical Medicine. *Proceedings of the 29th ACM SIGIR*, 99-106.
- Marchionini, G. (1989). Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia. *Journal of the American Society for Information Science*, 40(1), 54-66.
- McGuinness, D.L. (1999). In Ontology-Enhanced Search for Primary Care Medical Literature. Paper presented at the Int. Medical Informatics Association, Phoenix, Arizona.
- Mikheev, A., Moens, M., & Glover, C. (1999). In Named Entity Recognition without Gazetteers. Paper presented at the 9th Int. Conf. of the European Chapter of the Association for Computational Linguistics.
- Müller, H.-M., Kenny, E.E., & Sternberg, P.W. (2004). Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature. *PLoS Biol*, 2(11).
- Murdock, J.W., Pinheiro da Silva, P., Ferrucci, D., Welty, C., & McGuinness, D.L. (2005). In Encoding Extraction as Inferences (pp. 92-97). Paper presented at the AAAI Spring Symposium on Metacognition on Computation. AAAI Press, Stanford University.

- Noy, N.F., & McGuinness, D.L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Knowledge Systems Laboratory.
- Shah, U., Finin, T., & Joshi, A. (2002). In *Information Retrieval On The Semantic Web* (pp. 461-468). Paper presented at the 11th international Conference on Information and Knowledge Management, Virginia, USA.
- Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a Very Large Web Search Engine Query Log. *ACM SIGIR Forum*, 33(1), 6-12.
- Sonnenwald, D.H., Wildemuth, B.M., & Harmon, G.L. (2001). A Research Method to Investigate Information Seeking Using the Concept of Information Horizons: An Example from a Study of Lower Socio-Economic Students' Information Seeking Behaviour. *The New Review of Information Behaviour Research*, 2(November), 65-86.
- Sowa, J.F. (2002). Architectures for Intelligent Systems. *Systems Journal, Artificial Intelligence*, 41(2), 331.
- Spence, R. (2000). *Information Visualization (1ST edition ed.)*: Addison Wesley.
- Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). Searching the Web: The Public and their Queries. *Journal of the American Society for Information Science*, 53(2), 226-234.
- Staab, S., Studer, R., Schnurr, H.-P., & Sure, Y. (2001). Knowledge Processes and Ontologies. *Intelligent Systems, IEEE*, 16(1), 26-34.
- Vakkari, P. (2003). Task-Based Information Search. *Annual Review of Information Science and Technology*, 37, 413-464.
- Witten, I.H., Moffat, A., & Bell, T.C. (1999). *Managing Gigabytes: Compressing and Indexing Documents and Images (Second Edition ed.)*. San Francisco, CA: Morgan Kaufmann.