Designing Ontology-Based Interactive Information Retrieval Interfaces

Elena García[1] and Miguel-Ángel Sicilia[2]

[1] Computer Science Department, Alcalá University,
Ctra. Barcelona km. 33600 – 28871 Alcalá de Henares (Madrid), Spain
elena.garciab@uah.es

[2] Computer Science Department, Carlos III University,
Av. Universidad, 30 – 28911 Leganés (Madrid), Spain
msicilia@inf.uc3m.es

Abstract. The so-called Semantic Web advocates the future availability of machine-understandable metadata, describing Web resources by means of ontologies expressed in description logics. This would eventually entail changes in Information Retrieval (IR) indexing and matching algorithms, but also in the user interface design of IR tools. This second aspect can be informed by existing Interactive Information Retrieval (IIR) research, but it requires also further investigations about the interaction of users with terminological structures and iterative, browsing-oriented query construction paradigms. In this paper, preliminary experiences and reflections regarding ontology-based query formulation interface design are described.

1 Introduction

Information Retrieval (IR) can be defined as a discipline concerned with the formulation and resolution of queries issued by users against a database of information items (possibly heterogeneous in format and structure). According to [3], the effective retrieval of relevant information is affected both by the *user task* and the *logical view* of the information items. The interaction of the user with the IR system usually comprises both retrieval and browsing activities, oriented toward fulfilling an information need. The logical representation of an item is a representation of its contents, in many cases consisting on a set of keywords extracted by humans or by means of automated mechanisms. Both elements have been extensively studied in the last decades from two complementary perspectives: a system approach and an interaction approach. The system approach is mainly concerned with the process of matching a query against the database of items, in an attempt to produce optimal rankings. The interaction approach — often referred to as Interactive Information Retrieval (IIR) — puts emphasis in human factors, stressing the iterative nature of information searching (a review can be found in [20]), and extending the scope of IR to the needs, motivations and strategies of users in their interaction with IR systems.

Classical IR algorithms are founded on the assumption that information items (or documents) are modeled logically by keywords pertaining to some natural language. But the vision of a Semantic Web [7] radically changes this perspective. The Semantic Web essentially advocates "crossing the chasm" from unstructured keyword-based models to richer logic-based annotations that would eventually provide a basis for reasoning. This entails that the logical model of a document becomes a set of logical assertions about its contents (and perhaps also about its physical structure, its relationships with other documents and other information). In addition, the form of the queries becomes a logic expression with an arbitrary level of complexity in its structure. Both consequences, when taken together, lead to a reconsideration of existing IR user interfaces, beyond the mere typing of search terms. The problem can be divided again in a system-oriented part and an interaction-oriented aspect. The former is mainly a matter of logical satisfaction and concerns annotation approaches, and eventually problems of logical approximation in the sense given in [17] and elaborated, for example, in [25]. The latter concerns the interaction strategies of the user with one or several linked terminological structures, and the interpretation of her actions as query-formulating criteria, comprising activities at various levels of granularity, that, according to [5], can be roughly categorized as moves, tactics, stratagems and strategies. In this paper, we mainly focus on this second interaction aspect, in an attempt to provide a point of departure for the design of a new generation of IR interfaces based on formal ontologies (it should be noted that formal ontologies are different to lexical thesauri [27] in that they are free of some problems of ambiguity that occur in natural language) that provide an effective and efficient interaction relying on logics-based techniques like those described in [18]. Moreover, the results and design guidelines provided in this paper are based on previous research on a concrete ontology-driven IR system described in [22] and [13], and they are also informed by previous research on IIR in a broad sense. It should be noted that the criteria sketched in this paper do not exhaust the wide range of IR possibilities opened by Semantic Web technologies, and further research is needed both to accumulate experimental evidence (whenever semantic metadata becomes mainstream available) and to develop a methodological framework to construct and evaluate such systems. In consequence, the ideas discussed here are mainly directed toward stimulating further research.

The rest of this paper is structured as follows. In Section 2, the context and requirements of the problem is delimited, providing a set of assumptions about the evolution and the eventual deployment of the Semantic Web as a global infrastructure, just as the Web is today. From that initial context, a number of query formulation issues (and their related search strategies) are discussed in Section 3. Section 4 summarizes and discusses preliminary findings. Finally, conclusions and future research directions are provided in Section 5.

2 Some Assumptions regarding Interactive Ontology-Based Search

Since the Semantic Web is an evolving and growing reality, any attempt to investigate one of its applications should first state the assumptions in which it's grounded. In consequence, we first provide a reasonable set of assumptions that conform the point of departure of our ontology-based IR research. In addition, a number of design requirements are established as tentative goals elaborated from existing research.

The first assumption states simply that the Semantic Web relies in description logics (DL) as its foundational technical substrate. Currently, this appears to be largely uncontroversial given the increase of research results in this direction[3], and provided that the OWL language, endorsed by the W3C consortium as a Semantic Web standard [12], includes a description logic sub-language.

Assumption 1 *The $\mathcal{ALC}$ description logic (or some of its extensions) will be used as the language(s) both for metadata annotation and to describe their associated ontologies.*

We mention $\mathcal{ALC}$ as a minimum, since it's a basic, significant representative of DLs. Of course, more expressive DLs are currently used (and can be expected to be used) in Semantic Web languages and prototypes.

Assumption 2 *Shared ontologies expressed in DL by means of standard formats will be available to IR systems.*

Although it may take a long time to have available a set of significant consensual ontologies covering most domains of everyday's information seeking, it must still be considered a prerequisite for Semantic Web applications, specially for general-purpose Web search engines. The emergence and availability of large conceptualizations like OpenCyc[4] represent an important step in that direction. Both assumptions (1) and (2) are in accordance to existing analysis regarding the technological basis of the Semantic Web, as recently stated by Horrocks et al. [14], since formal models are obviously better suited to machine understandability, and common conceptualizations are required to enable interoperability.

Assumption 3 *Web resources annotated through DL assertions regarding shared ontologies will be available as the item base and logical item model for IR systems.*

Assumption 3 entails that a ontology-based metadata record will be provided for *every* searchable Web resource. Of course, and given the growing size of the Web, this provision may be expected to become a reality in a gradual way, giving room to hybrid models of IR. Nonetheless, here we'll approach the problem from an ideal perspective in which metadata exists for the whole database of

[3] http://dl.kr.org/
[4] http://www.opencyc.org/

resources. Moreover, annotation is still an open problem from the viewpoint of the quality and consistency of metadata records, and currently information extraction techniques appear to be a good compromise semi-automated solution to annotation [10]. In any case, we'll assume here that annotations are properly constructed, just as conventional IR engines rely in their keyword-based logical models.

Assumption 4 *The user population consists of people who do not specialize in search and are who have not knowledge about ontologies or knowledge models.*

Assumption 4 states that the query formulation and resolution mechanisms should be devised to reach the vast majority of the profiles that currently can be found in the Web. This precludes designs based on specialized query languages, and also user interfaces that include technical elements that are part of the solution domain. An example of such design is the `Ontobroker` query interface as described in [11], in which the technical ontology concepts of object, class, attribute and value are used in 'combo-boxes' to form logical query expressions.

Some specific design requirements must now be added to the just discussed assumptions to conform the space of design possibilities we're looking for.

Design Requirement 1 *Information retrieval interfaces must provide support to iterative query refinement and, as a supplementary function, also to serendipitous discovery.*

One possible approach to design an ontology-based IR interface is that of simply adopting the existing layout of current search engines, in which the user types some words of phrases, and then browses the list of results. This is the approach taken by querying systems like `OntoQuery` [2], that still relies on natural language processing for query resolution. But here we focus on IR interface designs in which the query formulation process is interactive and more user-controlled. These kind of alternative, more sophisticated UI designs have been advocated by Bates [5], resulting in a classification of user involvement levels and associated search tactics, stratagems and strategies. The topic of discovery by chance has been raised in recent studies [26] as a way to enhance the overall usefulness of information seeking behaviors. Despite the available evidence about its value, it still remains to be investigated the concrete design tactics and the limits of this feature. In consequence, we'll restrict ourselves here to provide some hints about potential ways in which ontology-driven search may trigger serendipity encounters.

Design Requirement 2 *The design and affordances of the interface must be independent of the domain.*

Requirement 2 entails that we are concerned with general purpose designs. Some existing ontology-based search tools provide interfaces specific to a given domain. For example, `Melisa` [1] provides a Windows-Icon-Menu-Pointer (WIMP) interface that gives the user the possibility to indicate 'evidence quality' and 'integration of the evidence' that are specific to medical or related domains. It should be noted that we don't claim that domain-specific elements should be absent from the interface, we only commit to plug-ability of ontologies in the same runtime of the IR system, so that for all the domain-specific query formulation information must be encoded associated to the ontology, in a language understandable for the IR system.

Design Requirement 3 *The user interface must be Web-browser based.*

Web browsers are the dominant interaction paradigm in the current Web, and their particular, uniform hypertext navigation mechanisms will likely stay largely unaffected while the HTTP protocol remain stable. Consequently, Web usability and information architecture [21] guidelines are still relevant to Semantic Web-enabled interfaces. Additionally, our present research is restricted to human information seeking. Further studies should address the possibility of building information-seeking software agents and also collaborative information seeking strategies.

3 Query Formulation and Search Strategy Design

In surface, the browsing of the concept network determined by ontologies appears as analogous to navigation on existing link catalogs like *Yahoo!* [15]. But although the clickstreams of both designs may appear as similar, the cognitive intention of user actions would eventually be fairly different, and the result of the user task is not the same. In the latter case, information items are found by simply browsing nested subject categories lacking a formal taxonomic criterion, while in the former, relationships between terms and also subsumed categories can be used to specify an underlying query that may be resolved following diverse heuristic approaches. Thus, the ontology-driven interface designer is faced with both usability and technical issues that open a number of currently unexplored alternatives.

In this section, we describe the rationale for the main design problems we faced in the design of the prototype ontology-based IR system described in [22] and [13], that will be referred to as `OntoIR` from here on. Technically, the system works by querying a `Rdf`-based ontological model comprised by three related sub-ontologies and implemented on the `Tomcat` Java-based Web server[5] using the `Jena` `Rdf`-processing libraries[6]. The three sub-ontologies describe respectively the *domain*, the Web *resources* annotated with concepts in that domain, and the *bibliographic sources* in which the domain concepts and relations are described. Any ontology serialized in the DAML+OIL language can be plugged into the tool, thus satisfying Design Requirement #2. Evaluation and user comments about `OntoIR` system work have resulted in a number of identified improvement areas,

[5] http://jakarta.apache.org/tomcat/
[6] http://www.hpl.hp.com/semweb/jena.htm

so that other design alternatives that were discarded or previously neglected are also suggested here, with the intention of motivating further research in the area.

The first problem encountered in the design of an ontology-based IR interface is simply where to start, that is, which elements of the ontology are provided as a guide for the user to begin the specification of the search.

Problem 1 *What are the elements that must be provided as a point of departure for the user task?*

In DL ontologies we have three basic candidates: concepts, relations between concepts, and their instances. If a search process begins from scratch, the problem of deciding which elements to show is limited by the practical design rule of limiting the length of pages [16] (thus avoiding scrolling). The number of concepts or terms in a given ontology may in most cases be in the range between a dozen and several hundred terms, while the number of objects is typically larger — a factor of the number of concepts. Since relationships may difficult the comprehension of the initial user move, it appears reasonable to begin with ontology concepts. Nonetheless, the use of relationships may be subject to future studies.

`OntoIR` is based in a top-down approach to query refinement, organized around concepts. According to it, the user first selects a domain from which to start the query, and then, the interface provides him/her with a number of what we call *entry points*, along with its descriptions. Entry points are a number of terms that are marked (with a form of meta-metadata) as query initiators, or that are selected from the profile of previous queries of the users, in the case of having an adaptive approach. The results and the terms that are provided by the system in a given search process are subject to become initiators for new searches, as discussed later. Thus, the information seeking problem is broken up in two (possibly iterative) steps:

– **Domain selection**. First, the user selects a topic (i.e. a domain or ontology) from which to start the search. It should be noted that this does not entail that the entire search will be limited to that ontology, since relationship traversal may lead to a query using terms from other domains, or combining terms from different ontologies. Since this step is not strictly related to query formulation, we'll not go into details about it. Given that in the future, the number of available ontologies may be relatively large (in accordance with Assumption #2), some form of hierarchical browsing — or even a classic IR process — should be provided for this selection step.
– **Query formulation and resolution**. The entry points of the selected ontology are provided as a point of departure. Therefore, it's advisable that the number of entry points be small enough to fit into the screen (or at least to minimize scrolling). The current UI design is shown in Figure 1.

If the system is provided with user modeling and adaptive capabilities, some form of *search-by-example* may be devised, taking previous queries or results of the user, or even from similar users if some form of *collaborative filtering* approach (see for example [19]) were included in the search interface.

Once in the query formulation step, a second range of alternatives arise.

Problem 2 *What are the user moves that contribute to query formulation and how do they contribute?*

Problem #2 refers to the interpretation of user interactions. The overall problem may be stated as how to translate a variable number k of discrete, sequential user moves (belonging to a set M) or interactions into a DL-based form. We have reduced the problem to the simpler one of collecting a set of terms $\mathcal{C}$ from user interactions as expressed in (1).

$$t : M^{(k)} \longmapsto \mathcal{C} \tag{1}$$

The rationale for such simplification is that studies on current search practices like [24] have shown that queries are short (most of them including less than ten terms) and also simple in structure. Nonetheless, it should be noted that the transformation expressed in (1) loses the ordering of the selection of terms and their selection context, and precludes selecting relations, so that further research should generalize this model.

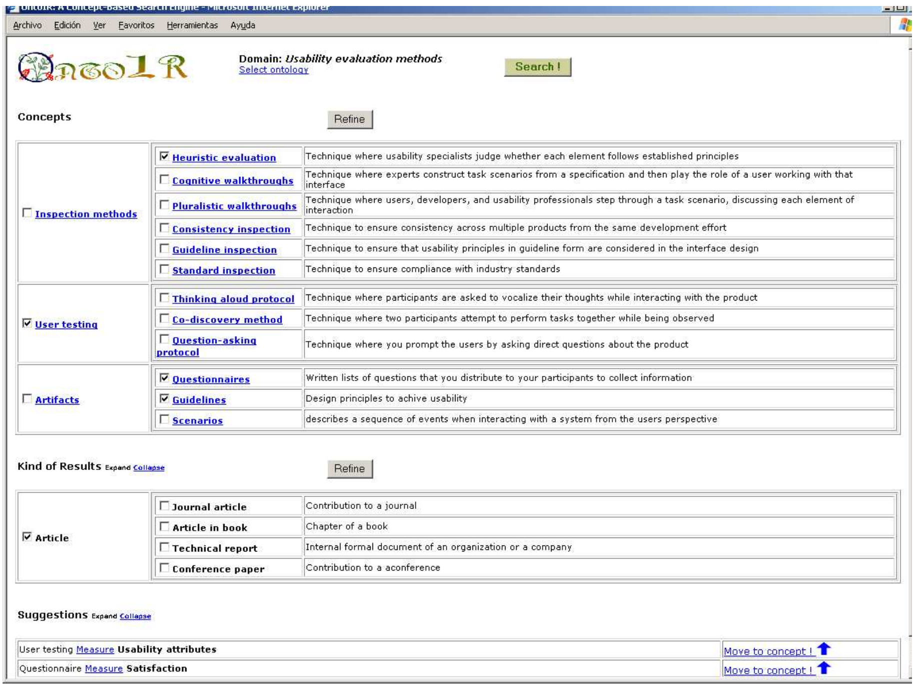

Fig. 1. Overall appearance of the `OntoIR` tool

The current version of `OntoIR` allows for the following kinds of user moves that build the query in a top-down fashion:

– The “Concepts” table allows for the selection (via *checkboxes*) of terms. The ‘refine’ functionality allows the user to explore the next level of the hierarchy for the selected terms, till the leaves of the generalization-specialization network is reached. This means going from more general categories to more specific ones, enabling wider or narrower searches. One major alternative that was discarded in `OntoIR` was that of allowing the user to move from terms to their generalizations (i.e. bottom-up moves).
– The “Kind of Results” table allows for restricting the type(s) of Web resources to be retrieved, i.e. the form of the document. This is not limited to the format, but to the type of document. The $(KA)^2$ ontology [6] provides an example taxonomy of scholarly documents that has been adopted in `OntoIR`, but richer categorizations, perhaps coming from library science, may be expected in the future. The selection process follows the same iterative top-down approach used for the domain concepts.
– The “Suggestions” area provides a way to enrich the query formulation process with related concepts. In this area, concepts bearing relationships with elements in the ongoing term collection ($c' \in \mathcal{C}'$) are provided (i.e. domain concepts $d \in D$ such that $R(c', d)$ or $R(d, c')$ being R any domain relation). The ‘move to concepts’ functionality makes appear the concept involved in the given relationship in the “Concepts” area, that is, the related concept is added to the ongoing $\mathcal{C}'$. The terms shown in this area are ramdomly selected from the possible ones.

The second and third areas in the list above are “collapsible” and “expandable” to save screen space if required. In synthesis, query formulation proceeds top down, suggesting related terms, and providing an optional document type filter. The terms in the “Concepts” area are links that trigger a separate browser window showing the bibliographic source(s) from which the concept definition was borrowed. This is a useful user function to prevent concept misunderstandings.

Once the elements that conform the query have been collected, an interpretation step is required, stated in Problem #3.

Problem 3 *What is the approach taken to match the query with the contents of the item database?*

Now the problem may be stated as how to translate $\mathcal{C}$ into a DL-based expression denoting a number of Web resources. The overall form of this translation step may be denoted as $r : \mathcal{C} \rightarrow \Psi$, where Ψ is a concept expression denoting a number of Web resources (direct or indirect instances of a top class `Resource`). It should be noted that Ψ should not be substituted by a logically equivalent expression $E \equiv \Psi$, since the terms selected by the user carry a semantic content that may eventually determine the results of the query and also contribute to future analysis of search behaviors.

Other alternatives to mapping r may combine concept expressions with example instances and/or relation names, allowing for more flexible retrieval approaches. Expression (2) describes the query approach of `OntoIR` for this mapping.

$$r_{OntoIR} : \mathcal{C} \longmapsto \overbrace{\mathcal{R}_{\mathcal{D}} \sqcap (T_1 \sqcup T_2 \ldots \sqcup T_l)}^{\Psi} \quad (2)$$

Expression (2) describes matching items as a concept expression $\mathcal{R}_{\mathcal{D}}$ restricted to the types of documents $T_1, \ldots, T_l$ selected in the "Kind of Results" area (all types are allowed if no T_j was selected). $\mathcal{R}_{\mathcal{D}}$ is a combination of three sets of Web resources related with the sets of instances C, E and S. Expression (3) defines S in terms of the concepts $D_1, D_2, \ldots, D_h$ selected in the "Concepts" interface area, as the concept covering all the instances of the domain concepts in $\mathcal{C}$, and C is defined as a subset of S in which an arbitrary domain relationship R exists to other element of S. In the same expression, E is defined as a set of concepts external to S that are related to one or more concepts in S.

$$C \doteq S \sqcap \exists R.S \;\; and \; E \doteq \neg S \sqcap \exists R.S \;\; given \; that \;\; S \doteq D_1 \sqcup D_2 \ldots \sqcup D_h \quad (3)$$

Given that instances of `Resource` (i.e. Web information elements) are linked to elements in domain ontologies through relations (with concrete positive or negative meanings like `describes`, `criticizes` or other kind of relation semantics) subsumed by a top relation $ABOUT$, $\mathcal{R}_{\mathcal{D}}$ can be determined by expression (4).

$$\mathcal{R}_{\mathcal{D}} \doteq Resource \sqcap (\exists ABOUT.C \sqcup \exists ABOUT.E \sqcup \exists ABOUT.S) \quad (4)$$

Where *Resource* is the top class of all kind of Web resources in the ontology. From expression (4), relevance criteria for resources can be implemented regarding to each of the three sets C, E and S. `OntoIR` currently considers a simple preference relation $\exists ABOUT.C \succ S \sqcap \neg\exists ABOUT.C \succ \exists ABOUT.E$, but more elaborated approaches could be devised from the same setting. For example, the number of connections inside C may be considered as an indicator of relevance, and even flexible approaches similar to quantified statements in [8] may be approached.

An example of a page of query results is showed in Figure 2. These results come from the query formulated in Figure 1 on a *usability evaluation methods and techniques* domain. The query has been built to retrieve all kind of on-line articles about questionnaires and guidelines in user testing and heuristic evaluation methods. It should be advised that all these concepts are the result of a previous refinement of several entry points and some of them could be refined in turn (e.g. questionnaires). In the basis of a sample of annotated resources, `OntoIR` shows results according to the priority rule defined above. The (internal) set S of concept instances comprises `severity rating`, `QUIS 5.0`, `QUIS 6.0`, `SUS` and `OSF/Motif guidelines`, among others. The first instance represents an opinion questionnaire, the following three are satisfaction questionnaires and the last one is a set of guidelines. As the `Standard Inspection` technique is not selected and it represents an inspection method that uses, for example, `OSF/Motif` `guidelines`, the generic `Standard Inspection` instance is included in set E. The set C contains instances like `severity rating`, `QUIS 5.0`, `QUIS 5.0` or the generic instance `heuristic evaluation`, since `QUIS 6.0 is a version of QUIS 5.0` and `heuristic evaluation uses severity rating` questionnaires. For each retrieved document the following information is provided: (a) The kind of resource, (b) a brief extract of the content of the document, that allows the user to evaluate its suitability for his/her search, (c) some relevant citation information about the document, and (d) the relations it maintains with other instances of the selected terms. These relations may be to initiate new searchs, since users can elaborate a new query with the terms involved in the relations via the "search using related concepts" functionality. In addition, the related instances are showed as links to access their description.

Obviously, interpretation requires further alternative explorations on realistic ontologies to have an idea of the appropriateness of such schemes. In addition, several alternative interpretations could be implemented, allowing the user of the system to decide which is better for the task at hand.

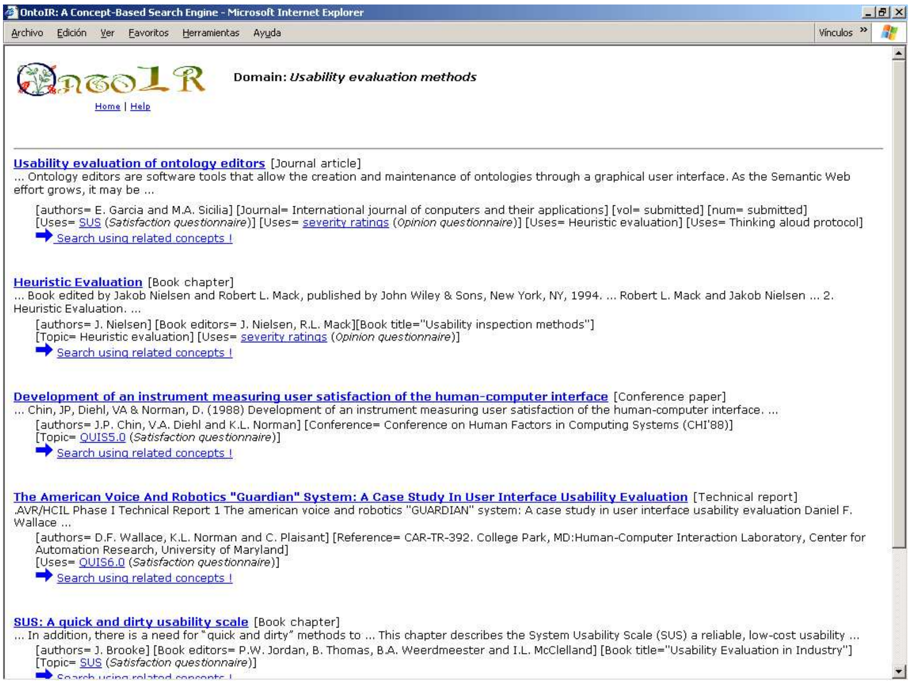

Fig. 2. A search results page in the `OntoIR` tool

The last problem that must be addressed is how to foster iterations and even casual encounters, as prescribed in Design Requirement #1.

Problem 4 *What mechanisms are provided to initiate a new search from a previous one?*

As we have advanced in the example, `OntoIR` provides a number of features that work as initiators of new search processes:

- Relationships suggested in the query formulation interface work as links leading to a separate search interface including the terms in the relation and also the entry points of their respective ontologies.
- Each of the search results is provided with the concepts related with it (along with the name of the relationship), and the "search using related concepts" functionality initiates a new search including the concept involved.

These affordanceses can be complemented by basic personalization features, including the history of past searches of the user, sorted by the number of times the user repeated it, and also a mode of search in which the user begins selection from the most selected concepts used by him/her in past searches. This is consistent with the considerable amount of repeated queries reported in [24].

In addition, some specific ontological axioms has been recently incorporated to `OntoIR` to experiment on casual searching behavior. Concretely, the `disjointWith` axiom of `OWL` has been used to provide query initiators based on concepts that are the opposite of those included in the ongoing query.

4 Some Preliminary Results and Reflections on Evaluation Techniques

According to the terminology proposed by Bates, our prototype system supports two levels of activity (falling into the "area of recommended development" proposed in [5]):

- At the "stratagem" level, the "Kind of Results" area has proved to be a separate mechanism for users to "Filter by Type" the ongoing search, in most cases carried out once at a time and before selecting domain terms.
- At the "tactic" level, the SUB (to move downward to a more specific term), RELATE (to move sideways to a coordinate term) and CONTRARY (to search for the term logically opposite) tactics [5] are provided. In addition, the SELECT tactic, i.e. breaking complex search queries into subproblems is somehow supported in the "Domain Selection" step, and the process of refinement — that only affects to selected terms — may be considered sometimes as a CUT tactic (to choose the option that eliminates the largest part of the search domain).

Although many search activities remain unsupported in our current prototype, its querying model is consistent to a large extent with evolving searching models like "berrypicking" [4], and the concatenation of search processes also facilitates term relevance feedback [23], as preliminary evaluation has pointed out.

While we wait for the emergence of TREC[7]-like evaluation suites for ontology-based IR systems, IIR models of evaluation like [9] can be tailored to the specifics of the design space described in this paper. `OntoIR` has experienced two evaluation processes. The first one, reported in [13] leaded to the separation of the overall query formulation interface in three areas, and also to a reformulation of the r_{OntoIR} mapping. The second one is more recent and was approached as a user testing process with ten users using the thinking aloud protocol. The users were first introduced to the features of the tool, and then they were given an example query and several tasks consisting on concrete searches involving from three to six refinement moves. Two groups of users were formed: profile A included programming students (daily users of the Web), while group B was formed by non-technical people with casual experiences with conventional search engines. Groups A and B were given six concrete search tasks regarding the domain of sailing, group A was provided also with four tasks regarding an ontology of programming languages. Both ontologies (and their associated resource bases) were built specifically for evaluation purposes, containing about a hundred annotated resources. Failure rates (i.e. unfinished tasks) were of less than one in average, and the average time per search task was of about three minutes. No significant differences were found between groups A and B, with the exception of the results of an informal brief Likert-scale (one to five) questionnaire administered at the end of the sessions regarding ease of use, learnability and perception of usefulness (compared to conventional search engines). A difference of more than one point in the scale pointed out to a worse perception of ease of formulating queries in group B. Observations also pointed out to the appropriateness of including ways to move up in the hierarchy, and also to provide more clear differentiations of the "Refine" and "Search" moves in the interface. Other minor possible enhancements are providing a text describing the overall domain while formulating the query and also a way to indicate that no more sub-hierarchy levels are available for a given term.

5 Conclusions

The realization of the Semantic Web vision calls for a revised conception of classical IR user tasks and logical models. Since ontologies are considered a key service in the prospective Semantic Web [12], the design of ontology-based IR interfaces has become an important research issue in this area. In this paper, our first experiences in designing and evaluating such systems has been described, highlighting the main assumptions, design requirements and problems that are faced when approaching the problem from a human factors perspective. A prototype has been described that enables a number of search tactics in the context of top-down iterative query refinement. Further research is required to explore

[7] TREC is the acronym for Text REetrieval Conferences: http://trec.nist.gov/pubs.html

other design alternatives and also to develop standards and best practices regarding the evaluation of ontology-based IIR systems.

References

1. Abasolo,J.M., Gómez, M.: MELISA: An ontology-based agent for information retrieval in medicine. In: Proceedings of the First International Workshop on the Semantic Web (SemWeb2000), Lisbon, Portugal (2000) 73–82
2. Andreasen, T., Fischer-Nilsson, J., Erdman-Thomsen, H.: Ontology-based Querying. In: Larsem, H.L. et al. (eds.) Flexible Query Answering Systems, Flexible Query Answering Systems, Recent Advances, Physica-Verlag, Springer (2000) 15–26
3. Baeza-Yates, R., Ribiero-Nieto, B.: Modern Information Retrieval. ACM Press, Addison-Wesley NY (1999)
4. Bates, M.J.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review **13** (1989): 407–424
5. Bates, M.J.: Where Should the Person Stop and the Information Search Interface Start?. Information Processing & Management **26** (1990): 575–591
6. Benjamins, R., Fensel, D., Decker, S.: KA2: Building Ontologies for the Internet: A Midterm Report. International Journal of Human Computer Studies, 51(3) (1999) 687–713
7. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American, **284(5)** (2001) 34–43
8. Bordogna, G., Pasi, G.: Flexible querying of structured documents. In: Larsem, H.L. et al. (eds.): Flexible Query Answering Systems, Flexible Query Answering Systems, Recent Advances, Physica-Verlag, Springer (2000) 350–361
9. Borlund, P.: The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. Information Research, 8(3), (2003) paper no. 152
10. Ciravegna, F., Dingli, A., Petrelli, D., Wilks, Y.: User-System Cooperation in Document Annotation based on Information Extraction. In: Gómez-Perez, A., Benjamins, R. (eds.): Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), Lecture Notes in Artificial Intelligence Vol. 2473. Springer-Verlag, Berlin Heidelberg New York (2002) 122–137
11. Fensel, D., Angele, J., Decker,S., Erdmann, M., Schnurr,H.P., Studer,R., Witt, A.: On2broker: Lessons Learned from Applying AI to the Web. Research report no. 383, Institue AIFB, Karlsruhe University (1998)
12. Dieter Fensel: Language Standardization for the Semantic Web: The Long Way from OIL to OWL. Proceedings of the 4th International Workshop on Distributed Communities on the Web, Lecture Notes in Computer Science Vol. 2468. Springer-Verlag, Berlin Heidelberg New York (2002): 215–227
13. García, E., Sicilia, M.A., Díaz, P., Aedo, I.: An Interactive Ontology-Based Query Formulation Approach for Exploratory Styles of Interaction. In: Jacko, J.A., Stephanidis, C. (eds.): Human Computer Interaction. Theory and Practice. Lawrence Erlbaum (2003)
14. Horrocks, I., Patel-Schneider, P.F., van Harmelen, F.: From SHIQ and RDF to OWL: The making of a web ontology language. Journal of Web Semantics, 2003 (to appear).
15. Labrou, Y., Finin, T.: Yahoo! as an Ontology: Using Yahoo! Categories to Describe Documents. In: Proceedings of the Eighth International Conference on Information Knowledge Management (1999) 180–187

16. Lynch, P.J., Horton, S.: Web Style Guide: Basic Design Principles for creating Web Sites (1999) Yale University Press
17. McCarthy, J.: Approximate Objects and Approximate Theories. In: Cohn, A.G., Giunchiglia, F., Selman, B. (eds.): Proceedings of the 7th International Conference on Principles of Knowledge Representation and Reasoning (2000) 519–526
18. Papazoglou, M.P., Porpoer, H.A., Yang, J.: Landscaping the Information Space of Large Multi-Database Networks. Data & Knowledge Engineering, **36**(3) (2001) 251–281.
19. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, Riedl J.: GroupLens: An open architecture for collaborative filtering of netnews. In: Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, Chapel Hill, NC: ACM (1994) 175–186
20. Robins, D.: Interactive Information Retrieval: Context and Basic Notions. Informing Science Journal, **3**(2) (2000) 57–62
21. Rosenfeld, L.,Morville, P.: Information Architecture for the World Wide Web. O'Reilly (1998)
22. Sicilia, M.A., García, E., Aedo, I., Díaz, P.: A literature-based approach to annotation and browsing of Web resources. Information Research Journal **8**(2) (2003)
23. Spink, A., Saracevic, T.: Human-computer interaction in information retrieval: nature and manifestations of feedback. Interacting with Computers, 10(3) (1998) 249–267
24. Spink, A., Wolfram, D., Jansen, B. J., Saracevic, T.: Searching the Web: The public and their queries. Journal of the American Society for Information Science and Technology, 52 (3) (2001) 226–234
25. Stuckenschmidt, H., van Harmelen, F.: Approximating Terminological Queries. In: Andreasen, T., Motro, A., Christiansen, H., Legind-Larsen, H. (eds.): Flexible Query Answering Systems. Lecture Notes in Computer Science, Vol. .2522. Springer-Verlag, Berlin Heidelberg New York (2002) 329-343
26. Toms, E.: Serendipitous Information Retrieval. In: Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries, Zurich, Switzerland European Research Consortium for Informatics and Mathematics (2000)
27. Voorhees, E. M.: Using WordNet for text retrieval. In: Fellbaum, C. (ed.):WordNet: An Electronic Lexical Database, MIT Press (1998) 285–303