Ontology selection: ontology evaluation on the real Semantic Web

Ontology Selection: Ontology Evaluation on the Real Semantic Web

Marta Sabou V. Lopez, Enrico Motta Victoria Uren
Knowledge Media Institute & Centre for Research in Computing
The Open University
Milton Keynes, United Kingdom
{R.M.Sabou, V.Lopez, E.Motta, V.Uren}@open.ac.uk

ABSTRACT

The increasing number of ontologies on the Web and the appearance of large scale ontology repositories has brought the topic of ontology selection in the focus of the semantic web research agenda. Our view is that ontology evaluation is core to ontology selection and that, because ontology selection is performed in an open Web environment, it brings new challenges to ontology evaluation.

Unfortunately, current research regards ontology selection and evaluation as two separate topics. Our goal in this paper is to explore how these two tasks relate. In particular, we are interested to get a better understanding of the ontology selection task and filter out the challenges that it brings to ontology evaluation. We discuss requirements posed by the open Web environment on ontology selection, we overview existing work on selection and point out future directions. Our major conclusion is that, even if selection methods still need further development, they have already brought novel approaches to ontology evaluation.

1. INTRODUCTION

In recent years, a conscious effort has been made by the Semantic Web community to migrate and apply its semantic techniques in open, distributed and heterogeneous Web environments. In fact, this tendency is suggestively captured in the thematic slogan of this year's Semantic Web Track[1] of the WWW conference: *"Where is the Web in the Semantic Web?"* Indications that the Semantic Web is evolving towards a *real* Semantic Web are already there. Not only has the number of ontologies dramatically increased, but also the way that these ontologies are published and used has changed. Ontologies and semantic data are published on the open Web, crawled by semantic search engines (e.g., Swoogle [6]) and reused by third parties for other purposes than originally foreseen (e.g., Flink [18] derives social networks from automatically crawled FOAF profiles). These changes motivate research on ceratin research topics, such as emergent semantics or semantic data integration. In particular, the increasing number of ontologies has led to the development of large scale ontology repositories and motivated a need for mechanisms that allow selecting the right ontology for a given task and context.

The *real* Semantic Web has implications on the future of ontology evaluation as well. Much of the work in this field was performed in the context of ontology learning leading to the exploration of different evaluation perspectives (e.g., application performance, similarity with a Gold Standard) and levels (e.g., lexical, conceptual). However, the scope of the evaluation was limited to asses the quality of one ontology at a time. Unlike ontology learning, ontology selection is performed in a significantly different setting which is characteristic for the real Semantic Web (large scale, dynamic, heterogeneous). Because ontology evaluation is core to the task of ontology selection, understanding the requirements for and the current status of ontology selection could provide a significant insight in the future challenges and opportunities for ontology evaluation. However, there have been no considerations about how these two tasks are related and on how to explore this relationship for the benefit of both fields. In this paper we aim to fill in this gap by seeking answers to the following questions:

1. *How does selection relate to evaluation?*
2. *Which are the requirements for selection that have an impact on the future of evaluation?*
3. *How are these requirements addressed by evaluation and selection approaches?*
4. *Which are the future directions for evaluation?*

Our analysis is structured as follows. We first discuss the relevant characteristics of the state of the art ontology evaluation (Section 2). To answer the first two questions we perform a requirements analysis for the selection task by (a) providing a definition that isolates its core evaluation function from other factors and (b) by exploring the requirements that are imposed by Web libraries and by semantic Web applications. In Section 4 we provide an overview of current work on ontology selection. In Section 5 we contrast the established requirements with current approaches in order to identify shortcomings and possible future directions. This discussion addresses our last two questions. We provide our final conclusions in Section 6.

2. ONTOLOGY EVALUATION

An important body of work exists in the context of ontology evaluation (see two recent surveys for an overview [2], [12]). In this section we provide an overview of the major characteristics of this field that are relevant for our analysis.

[1] `http://www2006.org/tracks/semweb.php`

WWW2006, May 22–26, 2006, Edinburgh, UK.
.

Ontology evaluation approaches are unevenly distributed in two major categories. On one hand, a few principled approaches exist that define a set of well-studied, high level ontology criteria to be manually assessed (e.g., OntoClean [11], Ontometric [16]). On the other hand, the use of ontology evaluation in the context of ontology learning has lead to the development of *automatic* approaches that cover different evaluation perspectives and levels. Evaluation levels refer to the aspects of the ontology that are evaluated (e.g., labels, conceptual structure). Perspectives are defined by what is considered to be a good "quality" ontology.

In an **application specific ontology evaluation** the quality of an ontology is directly proportional to the performance of an application that uses it. While several papers report on successfully using ontologies in various tasks such as text clustering (e.g., [13]), initial considerations on task-based ontology evaluation are only reported in [21]. Two problematic issues surface for such evaluations related to (a) the difficulty of assessing the quality of the supported task (e.g., search) and (b) creating a "clean " experimental environment where no other factors but the ontology influences the performance of the application.

In a **Gold Standard based ontology evaluation** the quality of the ontology is expressed by its similarity to a manually built *Gold Standard ontology* (e.g., [5]). One of the difficulties encountered by this approach is that comparing two ontologies is rather difficult. According to [17], one of the few works on measuring the similarity between ontologies, one can compare ontologies at two different *levels*: lexical and conceptual. Lexical comparison assesses the similarity between the lexicons (set of labels denoting concepts) of the two ontologies. At the conceptual level the taxonomic structures and the relations in the ontologies are compared. In fact, to our knowledge, the algorithm in [17] is the only one to offer a solution to the conceptual comparison problem.

In a **corpus coverage** scenario the quality of the ontology is represented by its appropriateness to cover the topic of a corpus. A rather lexical approach is taken to solve this task in EvaLexon ([22]) where well-established recall/precision type metrics are used to evaluate how well ontology triples was extracted from a corpus. A more conceptual level evaluation is attempted in [3]. In this case a set of important terms are determined in the corpus, then this set of terms is extended by adding two levels of hypernyms from WordNet. A probabilistic approach is used to compare ontology labels with the (extended set) of query terms.

Note that all these approaches have been developed for the scenario of evaluating a single ontology at a time. As a result, they often incorporate manual stages and are not too concerned about performance. Also, many of the automatic methods only address the lexical layer of ontologies and only [17] and [3] attempt a conceptual evaluation.

3. REQUIREMENTS FOR ONTOLOGY SELECTION

In this section we take a closer look to the task of ontology selection. We first provide a definition of this task in which we identify how it relies to ontology evaluation and which are the major factors that have the potential to bring up novel requirements (Section 3.1). In particular, we discuss the influence of ontology libraries and innovative semantic Web applications in Sections 3.2 and 3.3. We finalize this section with a discussion in Section 3.4.

3.1 Definition

We define *ontology selection* as the process that allows identifying one or more ontologies or ontology modules that satisfy certain criteria. The actual process of checking whether an ontology satisfies ceratin criteria is, in essence, an ontology evaluation task. For example, when one needs to select an ontology that has the best coverage for a given corpus, a prerequisite of the selection lies in evaluating all considered ontologies on this criterium. Therefore, ontology evaluation is core to the ontology selection task. To better understand the relation between ontology selection and ontology evaluation, we attempt to define the major elements of a selection process. This definition will be used through the paper as a basis for analysis and discussion. We distinguish the following elements that characterize the ontology selection process:

The information need. The aim of the selection process is to identify an ontology structure that satisfies a certain information need. The information need can be expressed differently. For example, it could be expressed as a set of keywords, a logical query, it could be represented by a corpus or by an ontology. Obviously, the way the information need is expressed is influenced by the requirements of the application that will use the results of the selection (an application defines a usage scenario for the selection task).

The selection criteria. The core task of the ontology selection algorithm is to **evaluate** a set of ontologies in order to identify the ones that fulfill the selection criteria. These criteria can be related to topic coverage, ontology structure or ontology popularity.

The Ontology library. Ontology selection is performed on top of a collection of ontologies, i.e., an ontology library. As we will discuss in the following sections, there are several issues related to ontology libraries that pose extra challenges on selection. These issues have been accentuated in the last years, as the Semantic Web has continuously evolved.

The output. The ontology selection process could have different outputs. For example, selected ontologies can be presented as a ranked list of ontologies. In other cases selection might return possible combinations of ontologies that jointly satisfy a certain information need. Often, consumers are only interested in a part of the ontology - so the relevant module should be presented from the perspective of the users and using the right level of granularity. In [19], N. Noy points out that objective evaluations do not often support the ontology users to their best and that particular care should be taken to help naive users find ontologies and evaluate their suitability for the user's tasks. From the perspective of ontology selection this translates in providing a friendly output for ontology selection facilities. For example, a summarization of the selected ontologies.

This definition of the ontology selection task illustrates our perspective that evaluation is core to selection. At the same time it is clear that this evaluation is influenced by the characteristics of the underlying libraries and the requirements of the scenarios that use selection (these influence the way the information need is expressed, the selection criteria to be used and the way output should be provided). In what follows we investigate how ontology libraries and semantic web applications are evolving and what constraints does this evolution towards a Web environment pose on the core evaluation part of the selection mechanism.

3.2 Ontology Libraries

Ontology libraries are crucial to facilitating ontology reuse. Several ontology library systems have been developed during the last years. However, a comparison between the libraries described by an earlier survey [7] and those developed in the last two years (e.g., Swoogle [6], OntoSelect [4], OntoKhoj [20]) shows an important change in the role and characteristics of these systems. The above mentioned survey perceives ontology libraries as core to an infrastructure that offers a wide range of management (editing, reasoning, versioning) and standardization facilities (alignment to upper level ontologies, use of standard languages). In contrast to this, more recent libraries have a narrower scope and they concentrate on automatically crawling the Web for ontologies, storing these ontologies and providing ways to access the crawled knowledge. This novel libraries inherit and accentuate a couple of library characteristics that pose extra challenges on selection:

Size. One of the aspects that have more dramatically changed with the evolution of the Semantic Web is the size of ontology libraries. Dynamically populated ontology repositories are significantly bigger than traditional libraries. For example, Swoogle [6] contains over 10,000 semantic documents, while OntoSelect [4] contains about 800 ontologies. Not surprisingly, due to their smaller size, few of the traditional ontology libraries provide automatic mechanisms for retrieving ontologies (the only exceptions being the Ontolingua server [10] and WebOnto [8]). Most of them rely on simple browsing interfaces. In contrast, automatic ontology selection is mandatory in the large Web based libraries.

Heterogeneity in domain, quality, level of detail. As Web based libraries acquire their ontologies by dynamic crawling rather than by static registration, the resulting ontology collection is characterized by heterogeneity at different levels. The collected ontologies cover a wider range of topics than traditional libraries. There are also large variations in the level of quality of ontologies collected from the Web. They range from rigorously checked, highly formal foundational ontologies to simple taxonomies. The level of details that are provided by different ontologies on the same domain can differ a lot. This poses extra challenges to select the ontology that has the right level of granularity/detail.

Level of control. The tendency of traditional libraries was to provide a controlled environment for ontology storage. Ontologies were manually uploaded and categorized in some sort of index structure. In contrast, Web libraries automatically crawl and store ontologies.

Modularity. Several traditional ontology libraries consider ontology modularisation as a prerequisite for reuse and encourage users to provide modularised ontologies. Current Web libraries do not offer any support for modularisation. Therefore, ontology selection mechanisms will have to deal with this aspect as well.

Summarizing the discussion above, we conclude a set of requirements for ontology selection in the context of newly developed ontology libraries. First, the *size* of these libraries requires automatic ontology selection methods. Because, as discussed in the following section, several Semantic Web applications rely on using ontology selection at run time, selection algorithms will need to provide a high performance. Second, ontology selection algorithms will have to consider issues related to ontology *modularity.* Finally, given the heterogeneity of the ontology libraries (regarding *domain, quality, granularity*) ontology selection methods will need to consider this issues and compare ontologies from the above mentioned perspectives.

3.3 Scenarios for Ontology Selection

Another source of novel requirements for ontology selection is represented by the practical scenarios in which it is employed. Such scenarios are defined by tools which use ontology selection. Ontology based tools are evolving from tools that were relying on a single, often fixed ontology to tools that open up to harvest the rich ontological knowledge available on the Web - the ontology becomes the variable rather than the hardcoded/stable part of the system.

In particular, we are working on two semantic Web tools that are evolving from using a single, rich and manually crafted ontology to exploring and combining ontologies available on the Web. These tools rely on automatic ontology selection and pose different requirements on the way information need is expressed, on the selection criteria to be used as well as the format in which results are requested. Awareness of such practical scenarios is essential to understand and develop ontology selection.

3.3.1 Ontology based Question Answering

AquaLog [15] is an ontology based question answering system. The novelty of this system with respect to traditional question answering systems is that it relies on the knowledge encoded in an underlying ontology to disambiguate the meaning of the questions and to provide the answers. To shortly give an impression about how the system operates, consider that it is aware of an ontology about academic life[2] which has been populated to describe KMi related knowledge[3]. Also, suppose that the following question is asked[4]:

Which projects are related to researchers working with ontologies?

In a first stage the system interprets the natural language question, and, using domain knowledge to resolve modifier attachments, translates it in triple-like data structures. For this compound question, two triples are identified:

(projects, related to, researchers)
(researchers, working, ontologies)

In the next step these triples are compared to the underlying ontology centered knowledge base using a set of string

[2] `http://kmi.open.ac.uk/projects/akt/ref-onto/`.
[3] This populated ontology can be browsed through a semantic portal at `http://semanticweb.kmi.open.ac.uk`
[4] For the interested, the AquaLog demo is available online at: `http://kmi.open.ac.uk/technologies/aqualog/`

comparison methods and WordNet. For example, the term *projects* is identified to refer to the ontology concept *Project* and *ontologies* is assumed equivalent to the *ontologies* instance of the *Research-Area* concept. The relations of the triples are also mapped to the ontology. For example, in the second triple, there is only one known relation in the ontology between a *Researcher* and a *Research-area*, namely *has-research-interest*. This relation is assumed as the relevant one for the question. However, when disambiguating the relation that is referred to by *related to*, the system cannot find any syntactically similar relation between a *Project* and a *Researcher* (or between all more generic and more specific classes of the two concepts). Nevertheless, there are four, alternative relations between these two concepts: *has-contact-person, has-project-member, has-project-leader, uses-resource*. The user is asked to chose the relation that is closest to his interest. Once a choice is made, the question is entirely mapped to the underlying ontological structure and the corresponding instances can be retrieved.

While the current version of AquaLog is portable from one domain to the other (the system being agnostic to the domain of the underlying ontology), the scope of the system is limited by the amount of knowledge encoded in the ontology. One way to overcome this limitation is to extend AquaLog in the direction of open question answering, i.e., allowing the system to benefit from and combine knowledge from the wide range of ontologies that exist on the Web. This new implementation of AquaLog is called PowerAqua [14]. One of the challenges in PowerAqua is the selection of the right ontology for a given query from a Web based library.

There are several requirements for ontology selection in the context of AquaLog. First, the information need supplied to the selection algorithm is expressed in terms of one or more triples (where both concepts and relations between them are important). Second, the selection mechanism needs to identify ontologies that contain these triples - i.e., the selection criteria requires coverage of such knowledge. Note that for retrieving the maximum number of results the system should check for ontologies that also contain synonyms of the search terms. If no ontology exists that mentions all triples (or their semantic equivalents) then the system should return ontologies that entirely cover at least one triple. This leads us to the third requirement, that the output should consist not only of specific ontologies but also of combinations of ontologies that jointly satisfy an information need. Finally, it would be beneficial to only retrieve the relevant modules rather than the entire ontologies.

3.3.2 Semantic Browsing

The goal of semantic browsing is to exploit the richness of semantic information in order to facilitate Web browsing. The Magpie [9] semantic web browser provides new mechanisms for browsing and making sense of information on the semantic Web. This tool makes use of the semantic annotation associated with a Web page to help the user get a quicker and better understanding of the information on that Web page. Magpie is portable from one domain to another as it allows the user to chose the appropriate ontology from a list of ontologies that are known to the tool. However, similarly to AquaLog, the current version relies on a single ontology active at any moment in time. This limits the scope of the sense making support to the content of the current ontology.

Our current research focuses on extending Magpie towards an open browsing. This means that the tool should be able to bring to the user the appropriate semantic information relevant for his information need from *any* ontology on the Web. This extension relies on a component that can select, at run time, the appropriate ontologies for the given browsing context.

In the case of Magpie, the information need for the ontology selection is more complex than for AquaLog as it is defined by the current browsing context. This includes the content of the currently accessed Web pages. Optionally, one can also consider the browsing history of the user as well as its user profile. For example, it should be taken into consideration that users are often unwilling to use radically different ontologies. Therefore an extra requirement would be that the result is similar to the currently used ontology (this would extend the specification of the information need with that ontology).

While in the case of AquaLog it is clear which concepts and relations should be covered by the extracted ontologies, in the case of Magpie this information needs to be filtered out from the textual data supplied. Unfortunately, this is a cumbersome task because Web pages can contain terms relevant for different topics. For example, the following short news story[5] is both about trips to exotic locations and talks. The question here is how to decide on the focus of the text, or at least how to identify the different topics mentioned in the story.

"For April and May 2005, adventurer Lorenzo Gariano was part of a ten-man collaborative expedition between 7summits.com and the 7summits club from Russia, led by Alex Abramov and Harry Kikstra, to the North Face of Everest. This evening he will present a talk on his experiences, together with some of the fantastic photos he took."

There are several different selection criteria that could be combined to select ontologies in the context of Magpie. Obviously, content coverage is an important criteria here, as it is in the context of AquaLog. We also need to keep in mind that the selected ontologies will serve as support for browsing so other characteristics become important. For example, these ontologies should have the right level of granularity/generality. In this case, more than in the case of AquaLog, it is important to only return the relevant modules of selected ontologies (large ontologies are difficult to present). Due to the "unfocused" nature of our queries, the selection algorithm is more likely to come up with collections of ontologies that jointly cover the indicated information need as an output.

3.4 Requirements Summary

In this section we synthesize the requirements brought forward both by libraries and new ontology based tools (see Table 1 for an overview). The *large size* of ontology libraries requires selection methods to be *automatic* and to deal with several ontologies at the same time. This also implies that selection methods should have a *good performance*. The same requirement is enforced by our ontology based tools which would ideally employ selection at run time.

A further requirement, imposed both by libraries and scenarios, is dealing with the heterogeneous ontology collec-

[5] `http://stadium.open.ac.uk/stadia/preview.php?s=29&whichevent=657`

tions.

Knowledge reuse is closely related to ontology modularization. Indeed, current applications would require selection mechanisms to return a relevant ontology module rather than an ontology. However, unlike their ancestors, recent ontology libraries do not provide any modularization information about ontologies, thus requiring selection mechanisms to deal with this aspect.

Another requirement relates to combining more ontologies in the answer of the selection task. This is a requirement for both our applications.

Finally, note that our applications require the retrieval of ontologies that semantically match their queries (i.e., the response and the query match at the meaning level, rather than just lexically). There is also a clear need for considering relations as well not just concepts when selecting an ontology.

4. CURRENT APPROACHES TO ONTOLOGY SELECTION

Several approaches have already been proposed for the problem of selecting (ranking) ontologies. A distinguishing feature between these approaches is the selection criteria that they rely on (i.e., the kind of evaluation that is performed). Based on this feature, we identified three categories of approaches that select ontologies according to their popularity (Section 4.1), the richness of semantic data that is provided (Section 4.2) and topic coverage (Section 4.3).

4.1 Popularity

Approaches from this category select the "most popular" (i.e., well established) ontologies from an ontology collection. They rely on the assumption that ontologies that are referenced (i.e., imported, extended, instantiated) by many ontologies are the most popular (a higher weight is given to ontologies that themselves are referenced by other popular ontologies). These approaches rely on metrics that take into account solely the links between different ontologies. In fact, these approaches use the same principle as current Web search engines (the importance of a Web page is proportional to the number of pages that reference it) and they often use a modified version of the PageRank algorithm. To our knowledge there are three approaches that consider ontology popularity.

OntoKhoj [20] is an ontology portal that crawls, classifies, ranks and searches ontologies. For ranking they use the OntoRank algorithm which is in spirit similar to the PageRank algorithm but instead of relying on HTML links it considers the semantic links between ontologies. Semantic links are denoted by instantiation and subsumption.

Swoogle [6] is a search engine that crawls and indexes online semantic Web documents. Swoogle allows querying its large base of semantic data and provides also some metrics for ranking ontologies. They rely on a similar principle as OntoRank and use a PageRank-like algorithm on semantic relations between ontologies (i.e., using terms of an ontology to define new terms, populating ontology terms, importing ontologies).

OntoSelect [4] is one of the first comprehensive ontology libraries that offers a complex ontology selection algorithm relying, among others, on selecting the most "well established" ontologies. The authors name this as the "connectedness" criteria since they look at how well an ontology is connected to other ontologies in order to determine its popularity. Unlike Swoogle and OntoKhoj, they use a less complex metric which only considers ontology imports as denoting semantic links between ontologies.

4.2 Richness of Knowledge

Another way to rank ontologies is to estimate the richness of knowledge that they express. When approximating this aspect, most approaches investigate the structure of the ontology.

The **ActiveRank** [1] algorithm is the only selection algorithm that has been developed independently from an ontology library. ActiveRank combines a set of ontology structure based metrics when ranking ontologies. To determine the richness of the conceptualization offered by the ontology they use the *Density Measure(DEM)* metric. This measure indicates how well a given concept is defined in the ontology by summing up the number of its subclasses, superclasses, siblings, instances and relations.

ActiveRank introduces two other measures that rely on the ontology structure and aim to evaluate the quality of the ontological knowledge. First, the *Centrality Measure(CEM)* metric relies on the observation that concepts which are in the "middle" of the ontology are the most representative and have the right level of generality. CEM is computed by taking into account the longest path from the root through the branch that contains a concept C to its node and the path from the root to the concept C. Second, the *Semantic Similarity Measure (SSM)* measures how close the concepts that correspond to the query are placed in the ontology by relying on the links between these concepts. The assumption is that an ontology that contains all queried concepts close enough to be treated as a module is better than an ontology in which these concepts are spread in different parts of the hierarchy.

In **OntoSelect** a similar metric, called *Structure*, is used. The value of the *Structure* measure is simply the number of properties relative to the number of classes in the ontology. The rationale behind this metric is that "more advanced ontologies have a large number of properties".

4.3 Topic Coverage

Finally, ontologies can be ranked based on the level to which they cover a certain topic. To determine this, most approaches consider the labels of ontology concepts and compare them to a set of query terms that represent the domain.

The *Class Match Measure (CMM)* of **ActiveRank** denotes how well an ontology covers a set of query terms. It is computed as the number of concepts in each ontology whose label either exactly or partially matches the query terms. Note that the matching is purely syntactic and no attention is paid to discovering synonyms or indeed to make sure that the concept is used in the same sense as intended by the query term.

The **OntoSelect** algorithm allows to specify the information need by supplying a whole corpus. The concepts that are the most relevant for a corpus are determined by statistical processing of the corpus. Then, coverage is measured by comparing the number of concept/property labels of the ontology with the query terms extracted from the corpus. This selection algorithm relies on the evaluation approach

Requirement	Imposed by		Addressed by	
	Libraries	Scenarios	Evaluation	Selection
Automation	y	-	[3], [17], [21], [22]	[1], [4], [6], [14], [20]
Performance	y	y	-	[1], [4], [6], [20]
Heterogeneity	y	y	-	[1]
Modularity	y	y	-	[1]
Combining Knowledge Sources	-	y	-	[14]
Relations	-	y	[17]	[4], [14]
Semantic Match	-	y	[3], [17]	[14]

Table 1: Requirements for ontology selection and the evaluation and selection approaches that address them.

proposed in [3].

In **OntoKhoj** they have considered word senses when ranking ontologies to cover a topic. In their algorithm they accommodate a manual sense disambiguation process, then, according to the sense chosen by the user, hypernyms and synonyms are selected from WordNet. The algorithm first tries to determine ontologies that contain the supplied keyword. If no matches are found, the algorithm queries for the synonyms of the term and then for its hypernyms. The algorithm was designed for a single word and it does not take into account relations.

Swoogle also offers a limited search facility that can be interpreted as topic coverage. Given a search keyword Swoogle can retrieve ontologies that contain a concept (or a relation) matching the given keyword. The matches are lexical and one can select between different levels of matches (exact, when the keyword matches exactly the concept label, prefix, when the keyword appears at the beginning of the concept label; suffix, when the keyword appears at the end of the concept label and fuzzy, when the keyword appears at any position in the concept label).

While still under development, the ontology selection algorithm which is part of the **PowerAqua** [14] question answering tool should be mentioned here. This algorithm aims to find the ontologies that cover a set of triples derived from a question. The minimum requirement is that any of the triples submitted as a query should be completely covered by an ontology. If elements of a triple are discovered in distinct ontologies than the triple is broken down in two more specific triples and the selection is reiterated. The output can contain more than one ontologies if different triples are covered by different ontologies. The selection itself is more semantic than existing approaches because it relies on WordNet senses, it checks for coverage of relations as well as concepts and considers the position of concepts within an ontology hierarchy to perform the selection.

4.4 Summary of Selection Approaches

To summarize Section 4, we provide a comparative overview of the ontology selection methods described above in Table 2. Our first obvious observation is that all the existing methods (except PowerAqua) rely on rather simple ways to specify an information need (a keyword, a set of keywords or a corpus from which a set of keywords are distilled) and use the same format to provide the output (i.e., a list of ranked ontologies). We also note that all approaches are designed for large-scale, automatically built ontology libraries.

Several interesting conclusions can be drawn with respect to the selection criteria used by these approaches. It is interesting to see that all methods offer some functionality for estimating topic coverage. This functionality is complemented with support for other selection criteria such as popularity or knowledge richness. In our discussion we have seen that there is a correlation between the selection criteria and the aspect of ontology that is evaluated. Namely, popularity based evaluation takes into considerations the links between ontologies, when knowledge richness is estimated then the structure of the ontologies is considered. Finally, topic based approaches consider the labels of the ontology elements when comparing them to the query terms. It is surprising that none of these approaches take advantage of ontological knowledge but rather treat ontologies as interconnected objects, graphs or bags of labels.

5. DISCUSSION

In this section we synthesize the material presented so far in order to answer the last two questions that we stated in the introduction. In particular, regarding question number three, we are interested in the way the requirements identified in Section 3 are addressed by evaluation and selection methods. These topics will be covered in Sections 5.1 and 5.2 respectively. To answer the fourth question, we look at several issues such as the shortcomings of current selection techniques (Section 5.3), the overlaps between selection and evaluation (Section 5.4) as well as the topics that should be addressed in the future (Section 5.5).

5.1 Requirements Met by Evaluation

As evident from Table 1, only a few of the identified requirements for the real Semantic Web are addressed by current evaluation approaches. Indeed, these approaches have been developed for dealing with one ontology at a time. As a result, not all of them are automatic (OntoMetric and OntoClean are purely manual). Further, there is little concern about performance or about topics such as dealing with heterogeneity, tackling modularization or combining more ontologies in the result set. The issue of a semantic match has been considered in current evaluation techniques and proved to be rather difficult. Illustrative in this aspect are the approaches described in [17] and [3] which have considered evaluation of ontologies at a more semantic level.

5.2 Requirements Met by Selection

Since selection algorithms have been developed to be employed on Web based ontology libraries, they partially satisfy the requirements that we have stated in Section 3.4. Indeed, all approaches provide automatic ontology selection. To maintain performance, the methods rely on rather

	OntoKhoj [20]	Swoogle [6]	OntoSelect [4]	ActiveRank [1]	*PowerAqua [14]
Information need	one keyword	one keyword	corpus	set of keywords	triples
Selection Criteria					
* Popularity	Yes	Yes	Yes	No	No
* Knowledge Richness	No	No	Yes	Yes	No
* Topic Coverage	Yes	Yes	Yes	Yes	Yes
Ontology Library	OntoKhoj	Swoogle	OntoSelect	Swoogle	Any
Output	ranked list of ontologies				combinations of ontologies

Table 2: Comparison of existing ontology selection approaches. * PowerAqua is still under development.

simplistic evaluation criteria that can be performed fast. ActiveRank, the most advanced selection algorithm, aims to tackle the heterogeneity and modularity requirements. Namely, the goal of its *Centrality Measure* metric is to determine an acceptable level of generality in the heterogeneous collection of ontologies. Then, the *Semantic Similarity Measure* is a first step to evaluate modularity issues - to approximately measure which ontology is more modular than the others for a given query set. Despite these achievements, several issues still need to be addressed in the future by ontology selection mechanisms, in particular related to more semantic evaluations, considering relations and combining knowledge sources.

5.3 Shortcomings of Selection

Following our analysis of the ontology selection methods we can identify two major shortcomings.

First, **the *meaning* of the concepts is ignored**. In particular, there is no considerations about synonyms (except OntoKhoj) or about the meaning of the terms as given by their position in the ontology (e.g., Politician/President, Executive/President or a president concept defined in a soccer/baseball ontology have different meanings). OntoKhoj is the only approach which has considered sense disambiguation (but only manually). We think that selection algorithms should be concerned with the actual meaning of the concepts rather than only with their lexical realization.

Second, **relations are ignored.** All approaches (except OntoSelect) focus only on concepts and ignore testing the existence of relations between given concepts. This is a major limitation given that relations provide valuable information to narrow down the search to the right ontologies. From the PowerAqua scenario it is clear that evaluating the coverage of relations between concepts is crucial. Relations are only moderately important in the Magpie scenario but, at the same time, finding the concepts with the right level of granularity is crucial.

5.4 Overlaps and Differences between Evaluation and Selection

When compared to the existing methods for ontology evaluation (we rely here on two recent surveys, [12] and [2]) we remark that few of the existing ontology evaluation techniques are employed by the selection algorithms described above. The only similarity is between the topic coverage algorithm of OntoSelect and the work described in [3]. On the other hand, the new context in which evaluation is performed has led to innovative methods for ontology evaluation. For example, the fact that a large number of ontologies is crawled from the Web gives an overview of all the semantic relations between ontologies and allows evaluating which one is the most popular. Also, the heterogeneity of ontological knowledge has prompted the work in [1] to tackle the evaluation of the right level of abstraction for a concept. The same work has provided the first steps towards estimating ontology modules based on structural features. Undoubtedly, these new approaches could provide inspiration for the ontology evaluation field.

5.5 Future Directions

Based on our analysis, we consider that ontology evaluation and selection have addressed complementary issues that can benefit both fields. First, ontology selection brings a set of new requirements for ontology evaluation that should be further analyzed and possibly addressed.

Second, selection and evaluation provide a complementary coverage of these requirements. Work on selection has made the first steps towards addressing requirements that are not considered by evaluation approaches (e.g., modularity, heterogeneity). In doing so, it has introduced new techniques for evaluation that could be adapted and refined for generic evaluation methods. On the other hand, evaluation methods are stronger on the side of performing semantic comparisons (though further development is needed in this area). This is definitely a shortcoming of selection techniques that should be addressed. It is, however, important to keep a balance between the complexity of the used evaluations and the performance of the algorithms.

6. CONCLUSIONS

In this paper we have explored the relation between ontology selection and ontology evaluation. Our hypothesis is that, by understanding ontology selection, its requirements and current state, we can provide useful insights into the future challenges and opportunities for the development of ontology evaluation. Our major conclusions are:

Ontology evaluation is core to ontology selection. This hypothesis has been captured in our definition of the ontology selection task. This definition identifies the function of evaluating how ontologies comply to a certain selection criteria as central to the selection task. Our overview of existing approaches has also confirmed that all selection methods perform a (or a combination of) evaluation task. Most frequently they evaluate the topic coverage of an ontology, as well as its popularity and the richness of the knowledge that is conceptualized.

The context of ontology selection poses extra requirements on evaluation. Ontology selection is performed in a Web context. It fulfills the needs of Web based applications and often operates on large scale, automatically crawled Web libraries of ontologies. This context brings new requirements for the ontology evaluation task such as the need for automation and a good performance. Also, topics such as heterogeneity, modularity, combining multiple ontologies and performing semantic matches should be considered.

Ontology selection and evaluation are complementary. Ontology selection and evaluation cover different requirements. While selection techniques have started to address issues such as modularity and heterogeneity, evaluation techniques are more advanced in covering semantic matches. While this complementary behavior is mutually beneficial for the development of both fields, all the mentioned requirements need further research.

More semantic evaluation needed. One of our critiques of the current approaches on ontology selection is that they do not regard ontologies as knowledge artifacts. Ontologies are treated as intelinked objects (when measuring popularity), as graphs (when using structural metrics) or as collections of strings (when estimating topic coverage). The meaning of the search terms and the concepts identified in potential ontologies is completely ignored. Our view is that selection techniques should be extended in the direction of a more semantic evaluation where the sense and context of search terms is considered.

While the analysis presented here is just a starting point for exploring the synergies between selection and evaluation, we are convinced that further work in this line should contribute to the establishment of the *real* Semantic Web.

7. REFERENCES

[1] H. Alani and C. Brewster. Ontology Ranking based on the Analysis of Concept Structures. In *Proceedings of the Third International Conference on Knowledge Capture(K-CAP 05)*, Banff, Canada, 2005. ACM.

[2] J. Brank, M. Grobelnik, and D. Mladenic. A survey of ontology evaluation techniques. In *In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)*, Ljubljana, Slovenia, 2005.

[3] C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks. Data-driven Ontology Evaluation. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, 2004.

[4] P. Buitelaar, T. Eigner, and T. Declerck. OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. In *Proceedings of the Demo Session at the International Semantic Web Conference*. Hiroshima, Japan, 2004.

[5] P. Cimiano, A. Hotho, and S. Staab. Clustering concept hierarchies from text. In *Proceedings of LREC*, 2004.

[6] L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, and P. Kolari. Finding and Ranking Knowledge on the Semantic Web. In Y. Gil, E. Motta, V.R. Benjamins, and M.A. Musen, editors, *Proceedings of the 4th International Semantic Web Conference*, volume 3729 of *LNCS*, pages 156 – 170, Galway, Ireland, November 6-10 2005. Springer-Verlag GmbH.

[7] Y. Ding and D. Fensel. Ontology Library Systems: The key to successful Ontology Reuse. In *Proceedings of SWWS'01, The first Semantic Web Working Symposium*, pages 93 – 112, Stanford University, California, USA, July 30 - August 1 2001.

[8] J. Domingue. Tadzebao and WebOnto: Discussing, Browsing, and Editing Ontologies on the Web. In *Proceedings of the 11th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Canada, 1998.

[9] M. Dzbor, J. Domingue, and E. Motta. Magpie - towards a semantic web browser. In *Proceedings of the Second International Semantic Web Conference*, Florida, US, October 2003.

[10] R. Fikes and A. Farquhar. Large-Scale Repositories of Highly Expressive Reusable Knowledge. *IEEE Intelligent Systems*, 14(2), 1999.

[11] N. Guarino and C.A. Welty. An Overview of OntoClean. In S. Staab and R. Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems. Springer-Verlag, 2004.

[12] J. Hartmann, Y. Sure, A. Giboin, D. Maynard, M. C. Suarez-Figueroa, and R. Cuel. Methods for ontology evaluation. Knowledge Web Deliverable D1.2.3, 2005.

[13] A. Hotho, S. Staab, and G. Stumme. Wordnet improves Text Document Clustering. In *Proceedings of the Semantic Web Workshop at SIGIR-2003, 26th Annual International ACM SIGIR Conference*, Toronto, Canada, July 28-August 1 2003.

[14] V. Lopez, E. Motta, and V. Uren. PowerAqua: Fishing the Semantic Web. In *Proceedings of the Third European Semantic Web Conference*, 2006.

[15] V. Lopez, M. Pasin, and E. Motta. AquaLog: An Ontology-portable Question Answering System for the Semantic Web. In *Proceedings of the European Semantic Web Conference*, 2005.

[16] A. Lozano-Tello and A. Gomez-Perez. ONTOMETRIC: A Method to Choose the Appropriate Ontology. *Journal of Database Management*, 15(2):1 – 18, 2004.

[17] A. Maedche and S. Staab. Measuring Similarity between Ontologies. In *Proceedings of European Knoeledge Ackquisition Workshop (EKAW)*, 2002.

[18] P. Mika. Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks. *Journal of Web Semantics*, 3(2), 2005.

[19] N.F.Noy. Evaluation by Ontology Consumers. *IEEE Intelligent Systems*, 19(4):74 – 81, July/August 2004.

[20] C. Patel, K. Supekar, Y. Lee, and E. K. Park. OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification. In *Proceeding of the Workshop On Web Information And Data Management*. ACM, 2003.

[21] R. Porzel and R. Malaka. A Task-based Approach for Ontology Evaluation. In P. Buitelaar, S. Handschuh, and B. Magnini, editors, *Proceedings of the ECAI Workshop on Ontology Learning and Population: Towards Evaluation of Text-based Methods in the Semantic Web and Knowledge Discovery Life Cycle*, Spain, 2004.

[22] P. Spyns and M.-L. Evaluating ontology triples generated automatically from texts. In *The Semantic Web: Research and Applications (ESWC05)*, volume 3532 of *LNCS*. Springer Verlag, 2005.