

CONTEXTUAL SEMANTIC SEARCH

Capturing and using the User's Context to Direct Semantic Search

Caio Stein D'Agostini, Renato Fileto, Mário Antônio Ribeiro Dantas and
Fernando Alvaro Ostuni Gauthier

Depto. de Informática e Estatística, Universidade Federal de Santa Catarina, Caixa Postal 476
88.040-900, Florianópolis - SC, Brazil

Keywords: Semantic web, context, personalization, user profile, semantic search.

Abstract: One orthodox perspective of the semantic Web depends on establishing a unified representation of ontology for an universe of discourse. However this expectation is unrealistic, because different people have different views of the world, making it impossible to devise a unified knowledge view that satisfies everyone in certain cases. One solution for this problem is to allow different users' views of consensual knowledge formalized in an ontology, and keep track of the mappings between these views and the underlying ontology. This work proposes to collect context information from the users interactions with a semantic search system, in order to gradually build individual users' views mapped to an ontology. This approach allows the user to pose queries based on keywords or his personal knowledge view. In addition, each personalized knowledge view captures the preferences of a single, specific user, enabling the system to provide better search results, based on its previous experience with that user.

1 INTRODUCTION

Semantic search approaches (Mangold, 2007) try to augment and improve searches on a set of resources that are initially unknown to the user, by using ontologies and semantic annotations of the resources. Even though there are several problems to be addressed when dealing with semantic technologies, such as the ontology creation and the semantic annotation process, we chose to center the scope of this paper on only one issue. The focus of our work is on how to capture user's contextual information and how to employ it, along with a given ontology and resources annotated according to its terms, to improve the precision, the coverage and the ranking of search results for individual users.

The ontologies used to describe resources in the semantic Web are potentially conflicting with particular users' views of an universe of discourse. These conflicts create a tension between individual users and formal ontologies. The former wants to see and interact with semantic Web systems according to their particular views, while the latter tries to describe a domain in a standardized way. In order to alleviate this tension, the systems' appearance and behavior can be customized according to particular users' views or, at least, the system must avoid forcing the user

to strictly comply to cumbersome ontologies, sometimes built by forcing consensus even when they are developed for small groups of people. Nevertheless, allowing different users' views characterizes just the stage of *Semantic Coexistence* in the Web (Naeve, 2005). In order to reach the stage of *Semantic Collaboration* (Naeve, 2005) it is also necessary to establish bridges between individual users' views and standardized ontological descriptions (Park and Cheyer, 2006).

This paper reports the progress of an ongoing study to create a personalized knowledge view for each user mapped to a consensual ontology. The goal is to provide individual tailored views of the knowledge base, so that he can benefit from the systems' use of ontologies and semantic annotations, without the need to cope with their full extent and details. These knowledge views are automatically created and updated by capturing contextual information during user interactions with the system, without additional user's effort. It alleviates the tension between individual users' views and the consensual ontology. In addition, as the individual views capture the user's context and preferences, they enable the estimation of user intentions in order to drive the disambiguation and semantic extension of keyword-based searches for resources annotated to a domain ontology.

This paper is organized as follows. Section 2 presents a definition of context relevant for the current work and some proposals on how to represent and use contexts. Section 3 presents our proposal to capture the context corresponding to the users' views and use it to improve the search process. Section 4 describes our representation of particular users' views and their mappings to an underlying ontology. Section 5 provides algorithms capture the users' views and to use it to drive the search process and potentially improve the search results by inferring the probable user intentions. Finally, section 6 presents closing comments and work left to be done or currently under development.

2 RELATED WORK

Context has a broad set of definitions according to how it is used. An appropriate definition for this work is the one from (Mani and Sundaram, 2007), which says that context is a dynamic subset of knowledge that affects the communication between entities. For information retrieval, context is a representation of a user's view of a domain. The context describes his preferences according to this particular view of a domain. It affects how the user's queries are interpreted and processed.

(Challam et al., 2007) and (Sieg et al., 2007) use ontological profiles to represent individual user contexts. In these profiles, the degree of interest of a user for a subject is represented by a weight in the ontology term referring to that subject. These weights are used to direct the searching algorithm (Sieg et al., 2007) and to rank the retrieved results (Challam et al., 2007). On both solutions the user's behavior is monitored. Information such as the websites visited by the user can be used to update his profile. However, this type of ontological profile only captures the interest for terms, failing to capture relevant relations between them.

(Michlmayr et al., 2007) uses publicly available bookmarks, from sites such as del.icio.us¹, to analyze how users correlate words. Since bookmarks are created at an specific time and stored for years, it is possible to create contextual profiles that reflects the user context at different time periods. A weighted graph is created by parsing all tags employed by the user to describe his bookmarks. Each tag is represented by a node in the graph, and every co-occurrence between two tags in the annotation of the same bookmark contributes to increase the weight of the edge between the

¹www.del.icio.us

corresponding nodes. The edges with higher weights are selected an user profile. This profile can be used to expand search expressions with other words that are probably also relevant to the user.

A similar graph structure is suggest by (Park and Cheyer, 2006). However, this work separates the users' views, represented as topic maps, from the ontological description of a domain. The topic maps are graphs where the nodes - called topics - represent subjects relevant to the users and the edges represent correlations between topics. The topics are named according to the users' views and are mapped to the terms with corresponding names in the ontology. However, (Park and Cheyer, 2006) fails to provide a detailed description of how to use these topic maps. Several questions are left unanswered, such as details of the topic maps structure (e.g., how homonym topics are stored and differentiated), their mapping to the ontology (e.g., how to know that different topics are synonyms), the criteria used to define browsing options in a topic map (i.e. which correlated topic is more important when there are several edges leaving a topic) and the creation and maintenance of topic maps.

The analysis of these works shows that there are several alternatives to represent contextual information. The solution proposed in (Michlmayr et al., 2007) individually represents the context from different users, but with no relation with a commonly accepted ontology. However, the ideias from (Michlmayr et al., 2007) might help to solve the issues left in (Park and Cheyer, 2006), by providing suitable mappings between users' views and ontologies.

3 CAPTURING THE USER CONTEXT

This paper proposes to explicitly represent particular users' views of an universe of discourse and establish mappings between each view and a consensual ontology. It helps to alleviate the tension between particular users' views and formal ontologies. In addition, it enables the alignment of the search results with the user's context.

Example 1 (Keyword-search for São Paulo). Consider a user who is travelling to Brazil by airplane. Suppose that he wants to learn about the airport he is going to arrive. He is going to São Paulo, which is a city, but this composite word is also the name of a state and a soccer team, among other entities. However, this foreign user does

not have all this knowledge. Assume that the user submits a search for the keyword *São Paulo* and waits for the results.

The first time the user poses a search, his context is empty. Since the system cannot infer his intentions, it checks the occurrences of *São Paulo* in the ontology. The system returns pointers to resources annotated with the keyword *São Paulo*, grouped according to different concepts associated to *São Paulo*, such as *city*, *state* and *soccer team*, as shown in Figure 1.

| Possible meanings for the keyword São Paulo | | |
|---|-----|-------------------------------------|
| <input checked="" type="checkbox"/> | | city(São Paulo) |
| Resources related to city(São Paulo) | | |
| | Doc | Annotated to: |
| <input type="checkbox"/> | B | city(São Paulo), city |
| <input checked="" type="checkbox"/> | C | city(São Paulo), Guarulhos |
| <input checked="" type="checkbox"/> | E | city(São Paulo), airport, Guarulhos |
| <input type="checkbox"/> | | state(São Paulo) |
| <input type="checkbox"/> | | soccer team(São Paulo) |

Figure 1: Results for a search for the keyword *São Paulo*.

The user selects only the results that, besides *city of São Paulo* *city(São Paulo)*, are also related to *airport* or *city(Guarulhos)*. The system then updates the weighted topic map representing the user's context. In the next time the user searches for *São Paulo*, the system employs the contextual information to elevate the rank of resources annotated to *city(São Paulo)*, *city(Guarulhos)* and *airport*.

Note that the association of a keyword with a specific concept and the respective instance from the ontology (e.g., *city(São Paulo)*) provides a more specific connotation for the keyword. The relevance of a connotation in the user context enables the disambiguation of future queries, and the relevance of the associations of the keyword with the annotation of the resources selected by the user enables means to drive semantic expansion of future queries. To make this context-driven search possible, there are some problems that need to be addressed: How to formally represent the user's context aligned to the ontological description of a universe of discourse; how to develop algorithms to keep the context information up-to-date with the continuously changing user's view; how to employ this information to improve search results. The proposed solutions for these problems are presented next.

4 REPRESENTING PERSONAL CONTEXTS ALIGNED TO AN ONTOLOGY

Our contextual semantic search system follows the three-layer architecture proposed by (D'Agostini et al., 2007). Figure 2 illustrates this architecture. Its layers are organized as follows:

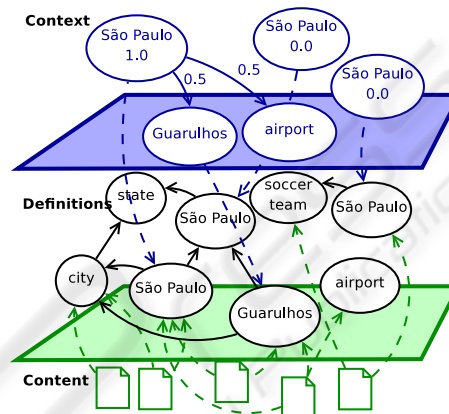


Figure 2: Mapping between the user's context and domain ontology.

- Users' Contextual Views (Weighted Topic Maps):** It maintains the users' knowledge views. These views are updated with information collected from previous user's interactions with the system. Each user has his individual view. The subjects of the views are indexed by their names in order to speed up the retrieval algorithms.
- Definitions (Ontology):** It maintains the ontology used to formally describe the universe of discourse and the non-intrusive annotations of resources based on that ontology. The terms of the ontology and the semantic annotation are also indexed to speed up their retrieval. The resources are considered to be already annotated.
- Content (Annotated Resources):** This layer refers to the stored content. It includes a repository of resources associated with ordered semantic annotations to support their retrieval. A resource can be a data set, a corpus (e.g., a document in a digital library) or a service (e.g., a resource in a computational grid), etc, depending on the type of application.

We represent each user's context by a weighted topic map. A topic can refer to any subject and has a name, properties and relations with other topics (Garshol, 2004). Topic maps have been chosen to

describe the users' contexts because they are human readable and organize knowledge in a similar way as people do (Novak and Gowin, 1984), while they can still provide a formal and machine processable knowledge representation. For the purpose of this work, a topic map is represented as a graph $TM(T,A)$, where T is the set of topics representing subjects of interest to the user and A is the set of associations which represent how the topics are correlated according to the user's view.

Each topic $t \in T$ is tagged with the word employed by the user to name the respective subject. Each topic corresponds to only one term (concept or instance of concept) formally described in the ontology. It is possible for two topics to have the same name, but they must refer to different terms with the same name.

In Example 1, the keyword São Paulo renders an ambiguous search, since São Paulo is the name of different subjects. To disambiguate searches, each topic $t \in T$ has a weight $w \in [0, 1]$. The sum of the weights from all topics with the same name equals 1. The topic weights are assigned based on how frequently the user referred to each topic (i.e., each connotation of the same word) in the past. Associations also have weights. Each association $a \in A$ has a weight $w \in [0, 1]$. The sum of the weights from all associations departing from the same topic also equals 1. The weights of the associations express the likelihood that a user searching for one subject is (also or instead) interested in another subject, as in Example 1, where the user searches for São Paulo, when he is in the airport located in the neighboring city of Guarulhos.

The process of creating new topics and associations in the topic map, as well as keeping them aligned to the underlying ontology is described next.

5 CONTEXT MANAGEMENT

The topic map represents a user's knowledge view. It is created and updated based on the user's feedback over the returned results for each search he poses. The system considers that the more frequently used topics are more relevant to the user than the less frequently used topics with the same name. This assumption allows the system to disambiguate the searched keyword. Since the searches are directed by the user's topic map, the topic map evolution and the search process are dependent on each other. The same context-search-context dependency happens at (Sieg et al., 2007), where it is called *the cold-start problem*. This dependency is broken by allowing searches to be performed directly on the ontology.

That type of search happens when there is no information available in the topic map to infer the user's particular intentions for a keyword, the search for that keyword is performed directly in the ontology. If the system finds different connotations for that keyword in the ontology, it presents them to the user, and asks him to choose results related to particular connotations. On the other hand, when the keyword is already in the topic map, the system can use the weights of topics referring to different connotations of the keyword to make a ranked list of results. The user's choices over the returned results are used, in both cases, to update his context. This process can be repeated for each search or until enough contextual information has been gathered.

5.1 Context Maintenance

The topic map evolves continuously at each search, by adjusting the topic and association weights according to the user choices over the search results. A topic weight is increased when the term it refers to is used to annotate a resource considered relevant by the user. The weights of the associations are updated according to the correlation between relevant topics. The associations correlating the topics corresponding to the terms used to annotate the selected results and the topic corresponding to the relevant meaning of the search keyword have their weights increased. The higher the number of times a term is used to annotate the chosen resources, the higher the weight of its corresponding association with the keyword topic. If the association does not exist, a new one is created. This process is described in Algorithm 1.

Algorithm 1 (Context Maintenance Algorithm).

1. The user checks the resources he considers relevant
2. For each topic checked by the user{
3. Increment the weight from the topic matching;}
4. For each resource checked by the user{
5. Increment the weights from the associations between the topic corresponding to the keyword and the topics corresponding to the terms used to annotate the checked resources;}
6. Normalize the weights from topics and associations to the interval $[0,1]$;

The weights of topics and associations which are not considered relevant in a search can be decayed, so that feedbacks from recent searches have more influence in the weights than older ones. This procedure can be used to reflect the temporal aspects of the context.

Consider the results from the search in Example 1, in Figure 1. The user selected the resources C, E, which are annotated to $city(\text{São Paulo})$. But so is resource B, which was not checked by the user. Also, both resources C and E are annotated

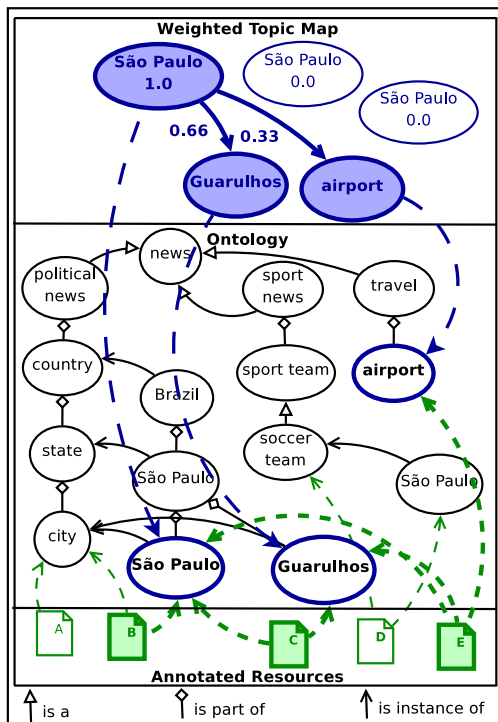


Figure 3: Possible relevant result based on the user's topic map.

to Guarulhos and E is also annotated to the term airport, to which none of the rejected resources are. With this information the system can infer that, when the user searched for São Paulo, he was interested in the city, not in the state or soccer team. And since the Guarulhos and airport annotation are the ones which differs resources C and E from the rejected ones, it can be inferred that, by searching for city(São Paulo), the user was actually searching for Guarulhos or airport. Figure 3 shows the corresponding created topic map and its mapping to the ontology. The correlation between São Paulo (the keyword searched for) and Guarulhos is stored by applying a weight 1.0 to a topic corresponding to city(São Paulo) and creating an association between that topic and the topic Guarulhos. This association has assigned a weight of 0.66, since it correspond to 2 of the 3 annotations characteristic to the selected resources (2 x Guarulhos and 1 x airport). The topic airport receives a lower weight, since airport is used less times than Guarulhos to annotate the results checked by the user.

5.2 Contextual Semantic Search

During the search process, summarized in Algorithm 2, the weighted topic maps are used to direct the

search in the ontology. This directioning is done by disambiguating keywords and expanding the search to include other subjects the user relates to the subject he is searching for, allowing the search to consider the user's particular view of how different elements from a domain relate. Depending on the keywords searched, three different situations might occur: the searched keyword has a correspondence in the topic map, meaning that there is context information available in the topic map to direct the search process; the keyword has no correspondence in the topic map, but there are terms in the ontology which are referred by the keyword (the 'cold-start' problem); the keyword has no correspondence either on the topic map or at the ontology and no results are returned. For all these situations the search process described here considers that the keyword received as parameter does not need any type of processing. This means that steps such as identifying if a search expression is either a composite word or two independent words have already been performed.

Algorithm 2 (Search Algorithm).

```



1. The user inserts a search keyword;
2. For each keyword{
3.   Look in the topic map for matching the keyword;
4.   If system finds topic(s){
5.     Disambiguate the topics using their weights;
6.     Expand the search using the disambiguated
       topic's associations;
7.     Identify the terms in the ontology corresponding
       to the topics;
8.   }else{
9.     Look in the ontology for terms matching the
       keyword;
10.  If system does not find term(s){
11.    Ends search;
12.  }else{
13.    Identify those terms; } }
14. Recover and list to the user the resources annotated
    by the terms identified terms;

```

For the searches driven by the context information, the first step is, if necessary, to disambiguate the keyword. The disambiguation is done by selecting the topics with the higher weight in the topic map, since they are likely the ones the user is interested at. After selecting a topic, it is necessary to verify how the user correlates that topic with the others. For this verification, the associations departing from the selected topic are ordered according to their weights, from higher to lower weights. The search is then expanded, considering also the terms in the ontology related to the topics on the other end of those associations. This process is exemplified in Example 2 and illustrated in Figure 3.

Example 2. Take the results from Example 1. Consider that the user performed a new search for São Paulo. Through the weight of the topics named São Paulo the system can infer that the user means city(São Paulo). By also consid-

ering the weights from the associations originating at the topic *city(São Paulo)*, it can also be inferred that he is most likely trying to find information about Guarulhos and airport and not about the city of São Paulo. This last inference is only possible because of the association established between *city(São Paulo)* and Guarulhos or airport, which is not represented in the ontology, only in the user's topic map.

| Resources related to city(São Paulo) | | |
|--------------------------------------|---|-------------------------------------|
| Doc | Annotated to: | |
| <input type="checkbox"/> |  | city(São Paulo), airport, Guarulhos |
| <input type="checkbox"/> |  | city(São Paulo), Guarulhos |



| Other results for the keyword São Paulo | | |
|---|---|-----------------------|
| Doc | Annotated to: | |
| <input type="checkbox"/> |  | city(São Paulo), city |
| <input type="checkbox"/> |  | ... |

Figure 4: Results from the search from Example 2.

6 CONCLUSIONS

The use of personalized knowledge views mapped to an ontology reduce the tension between usability and consensual knowledge and reach the Semantic Collaboration stage of the Web. This paper presents a proposal to (i) gradually build the user's view and keep it updated, by gathering contextual information from his interactions with a semantic search system; (ii) represent the user's context as a weighted topic map, whose topics are connected to specific connotations described in an ontology; (iii) use these weighted topic maps to infer probable user intentions, based on his previous choices, in order to make a ranked list of search results tailored for the specific user.

At the current stage, experiments are being planned and the data necessary for them are being collected. Thus, there are no experimental results available yet. Nevertheless, many of the assumptions made so far can be verified by experiments realized in related works, such as (Challam et al., 2007), (Sieg et al., 2007) and (Michlmayr et al., 2007).

Currently, the topic map representing a particular user's view only allows concepts and instances described in the underlying ontology as topics. One challenge is to find a trade-off between the level of independence given to the users to express their knowledge views and the limitations and cost of the facilities to establish semantic bridges between these

views. Thus, allowing topics with no corresponding term in the ontology and using these topics to evolve the ontology is a theme for future work. The visual presentation, browsing and edition of the user's topic map are also considered for future developments.

ACKNOWLEDGEMENTS

This work is supported by CNPq (grant 48139212007-6), CAPES and FEESC.

REFERENCES

- Challam, V., Gauch, S., and Chandramouli, A. (2007). Contextual search using ontology-based user profiles. *Conference RIAO2007*.
- D'Agostini, C., Fileto, R., Dantas, M. A. R., and Gauthier, F. O. (2007). Inferring user's intentions through context. *Sessão de Posters do Simpósio Brasileiro de Bancos de Dados (SBBDD)*, pages 7–10.
- Garshol, L. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, 30(4):378.
- Mangold, C. (2007). A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontologies*, 2(1):23–34.
- Mani, A. and Sundaram, H. (2007). Modeling user context with applications to media retrieval. *Multimedia Systems*, 12(4):339–353.
- Michlmayr, E., Cayzer, S., and Shabajee, P. (2007). Tech report: Hpl-2007-72: Adaptive user profiles for enterprise information access. Technical report.
- Naeve, A. (2005). The Human Semantic Web-Shifting from Knowledge Push to Knowledge Pull. *International Journal of Semantic Web and Information Systems*, 1(3):1–30.
- Novak, J. and Gowin, D. (1984). *Learning How to Learn*. Cambridge University Press.
- Park, J. and Cheyer, A. (2006). Just For Me: Topic Maps and Ontologies. *Lecture Notes in Computer Science*, 3873:145–159.
- Sieg, A., Mobasher, B., and Burke, R. (2007). Ontological user profiles for personalized web search. *Proceedings of the 5th Workshop on Intelligent Techniques for Web Personalization, Vancouver, Canada, July*.