Falcon-S: An Ontology-Based Approach to Searching Objects and Images in the Soccer Domain

Honghan Wu, Gong Cheng, and Yuzhong Qu

School of Computer Science and Engineering,
Southeast University, China
{hhwu,gcheng,yzqu}@seu.edu.cn
http://xobjects.seu.edu.cn/

Abstract. While the Semantic Web technology has reached a considerable achievement, the main part of the Web is still far from semantic. To achieve semantic search on the current Web, we developed Falcon-S, a Semantic Web application which indexes soccer images semantically and provides ontology-driven searching and browsing mechanisms. Especially a graph-based user interface is provided to customize semantic queries for domain objects. In addition, the meanings of the image contents are captured and utilized to improve performance. From our experiments and the evaluations, it can be figured out that before the Semantic Web eventually creates order out of chaos on the Web, an ontology-driven approach applied to traditional search technology is a promising way to bring semantics to current Web in the searching area.

1 Introduction

Falcon-S is a Semantic Web application focusing on the soccer domain. An ontology and a knowledge base for the soccer domain are built by integrating the information from the official web sites of FIFA[1] , UEFA[2] , etc. It crawls the World Wide Web for images on the soccer domain, analyses their context and indexes them on the objects (e.g., players, teams) in the knowledge base. Falcon-S provides functionalities for users to search and explore objects and images in the soccer domain, and some services are beyond what traditional search or query applications can provide. Fig. 1 illustrates the architecture of Falcon-S.

The rest of the paper is organized as follows: Section 2 introduces the three services of Falcon-S. Section 3 presents the features of Falcon-S. Evaluations are presented in Section 4. A conclusion is given in Section 5.

2 Services

At present, Falcon-S provides three kinds of services, i.e. basic search, semantic search and semantically exploring. These services will be briefly described in this

[1] http://www.fifa.com/
[2] http://www.uefa.com/

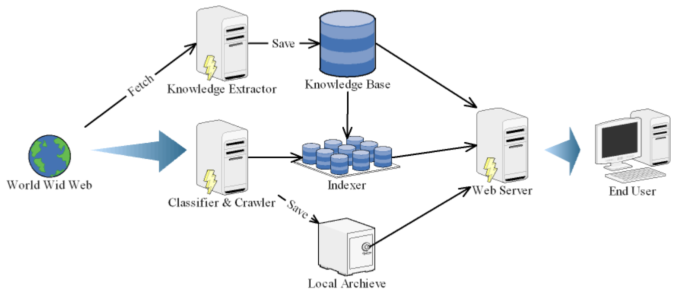

Fig. 1. Falcon-s Architecture

section. The underlying techniques that support these services are discussed in the next section.

2.1 Basic Search

Similar to traditional search application, the user can use one or more terms to search for soccer images. The add-on of Falcon-S in the basic search is finding the objects associated with the search terms and then extracting semantic information associated with the found objects from the knowledge base. If one term denotes one or more domain objects, the objects will be listed. For each identified object, the detailed information and objects that have semantical association with it can be expanded out (Fig. 2).

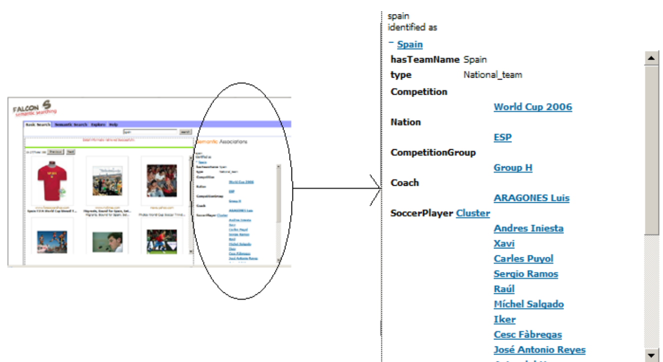

Fig. 2. Basic search

2.2 Semantic Search

The semantic search is based on the semantic associations between objects which will be discussed later. A graph based user interface is provided to compose the compound associations which are also called as semantic queries. When running a semantic query, corresponding objects and related images of them will be returned as results. Fig. 3 shows the edit panel of semantic queries. The user can create new semantic queries or modify existing ones created by others, such as “Find me the players who are teammates of Ronaldo and also play the same position of him” (See the association graph in figure 2). Detailed information about how to use the user interface can be accessed from the help page of Falcon-S (http://www.falcons.com.cn/help.html).

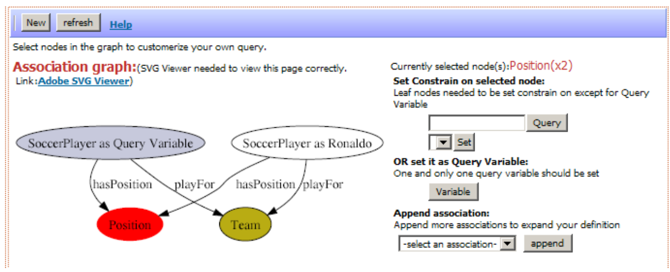

Fig. 3. Semantic query edit panel

2.3 Explore

Guided by concepts from domain ontology, the user can select an individual in the domain to start exploring. Semantically associated objects are listed under the individual. Through these linked objects the user browses information guided by the domain ontology. So, Falcon-S exhibits an ontology-driven approach to semantically exploring domain objects and associations among them.

3 Features

3.1 Customized Semantic Search

In a knowledge base, objects are connected by properties. Hence most semantic applications use the property as the unit to find semantically related objects to the end user such as [1]. However some properties may be too fine-grained or even meaningless to the end user. In Falcon-S the units of associations between entities are pre-defined meaningful paths consisted of one or more properties, which we call semantic associations.

There are two kinds of semantic associations in Falcon-S. The first is the basic associations which are defined manually as paths between classes. These associations are basic but most important relations in the soccer domain, such as: playFor, hasNationality, hasCoach, etc. The basic association can be viewed as the notion of *Semantic Connectivity* which was defined by Boanerges Aleman-Meza, et al in [2] [3] .

The second type of associations is what we call the compound associations which are composed with the basic associations. While basic associations are fundamental and commonly accepted, the interestingness of complex associations varies heavily to different persons. We provide a mechanism for the end users to customize their own associations as queries. When basic associations are paths connected two entities, compound associations are graphs consisting of a set of connected nodes. It is notable that the arcs in these graphs are not properties in the knowledge base but the pre-defined basic associations.

Technically, these compound associations are eventually converted into RDF query statements (we use SeRQL now) which will be searched upon the knowledge base. After the searching, related objects i.e. instances of the query variable will be returned. Eventually these objects are used to search in the semantic indexing to get all related images.

3.2 Mapping search terms to objects by using contextual information

Falcon-S constructs a term-object index from the knowledge base. Searching terms will be matched first in this index to find corresponding objects. If it is a multi-term search, the search string will be used in the identification process for domain objects. For example, a single term “ronaldo” would be identified as “Brazil Ronaldo” as default. When the query string is “Manchester ronaldo” or ” “Portugal Ronaldo”, the term “ronaldo” would be identified as “Cristiano Ronaldo” because his context information contains “Manchester” and “Portugal” (In facet, he plays for the club of Manchester United as well as the national team of Portugal).

3.3 Indexing images by domain objects

Traditional search engines index the crawled documents in the Vector Space Model[4] which has several limitations. First the independent assumption of terms makes this model inadequate to capture the semantics between terms [5]. Secondly, the missing of machine understandable semantics makes it impossible to answer queries that can not be expressed in combination of terms. For example, one user may want to find the images of top ranked forwards who participated in the FIFA World Cup 2006.

Provided with the soccer domain ontology and a knowledge base, we do have the semantics of terms explicitly rather than untouched or latent. Hence we can address these problems in an ontology-driven way in Falcon-S. In addition to traditional inverted term-document index, we created an object-image index in which the image is indexed by corresponding objects’ IDs, depicted in Fig. 4 as the lines between the document and knowledge layers.

In the indexing process, the contextual information of objects and images is used. For an image, the context information is those texts that surround it in the Web pages, such as the alternative text. The context information of an object consists of the information of objects that are related to it through

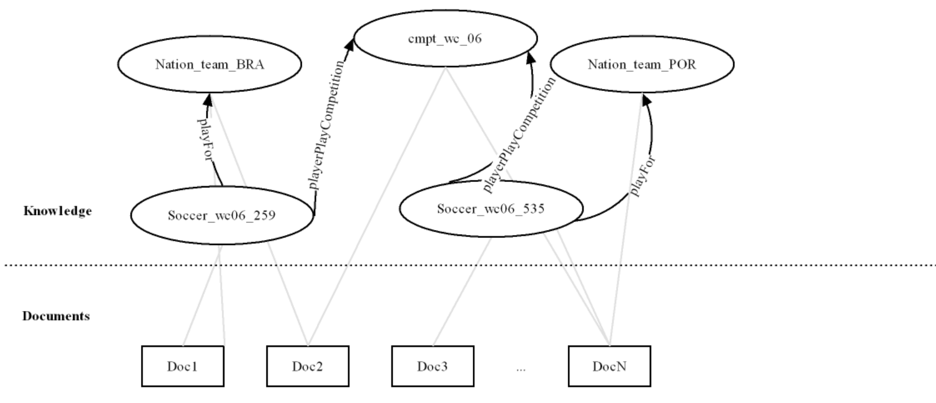

Fig. 4. Falcon-s indexing model

specific associations. These associations is a subset of basic semantic associations (refer to previous subsection). They are manually selected from domain ontology to keep the context information discriminable among objects. For example, in Fig. 4, the association of playFor is selected instead of playerPlayCompetition because competitions are much less discriminable than the teams he plays for.

3.4 Enhancing the search by the image content

In Falcon-S, we introduced an ontology-based approach to bridge the semantic gaps[6] between low-level feature (we only use color feature) and the high-level concept. The idea comes from the fact that in the soccer domain jerseys are most important sources for a human to identify teams. Moreover each team has a rigid definition about the colors of their jerseys. That means the color feature of an image is an important source for measuring the relevance of the image to a soccer team. Therefore if we can apply this knowledge in the searching process, the precision should be improved.

The knowledge of correspondences between domain objects (i.e. teams) and the colors of their jerseys is captured into the knowledge base. When indexing images, the dominance colors of images are stripped and added into the indexes. In both processes, color histogram analysis is used to generate the representing colors.

Eventually, when searching, if one of the terms denotes a domain object who has color knowledge, color queries are generated automatically and combined with original queries. A score factor is added into the score formula in the combination. Hence the hits that match the colors will have higher score than those do not.

4 Evaluation

4.1 Customized semantic search

We proposed a graph-based interface to customize semantic queries for the end user. Related works such as [7] [8] proposed graphical user interfaces for building complex queries over knowledge bases. Four criteria are used for the evaluation of the three as follows.

- **-Intuitiveness** In both [7] and [8], the whole structure of queries is not represented in a graph. [8] only provides a graph based interface for nodes manipulation. Hence in both case users need to map the query graph in his mind to the user interfaces. The queries are intuitively drawn as graphs in our interface, which brings better usability.
- **-Structure of composed query** Both our interface and [7] support the definition of queries whose data models are graphs. In [8], only tree-structured queries can be defined.
- **-Reuse of composed query** Defined queries can be saved and loaded for reuse in [7] and ours. [8] focuses on browsing. Hence no mechanism of reuse is provided.
- **-Availability** [7] and ours are web-based which has higher availability than standalone tools like [8].

4.2 Semantic indexing evaluation

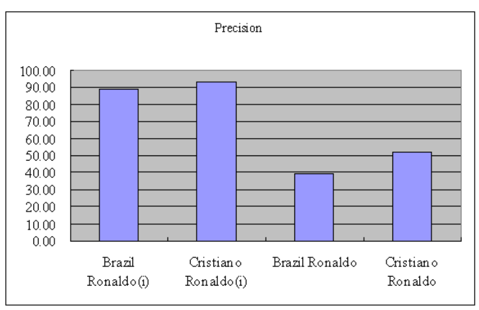

Fig. 5. Precision of search for "Ronaldo"

In this subsection we evaluate the indexing model. First we search the traditional term-document index with the keyword "Ronaldo". The result set (totally 280 hits) includes both "Brazilian Ronaldo" and "Cristiano Ronaldo". If user's intention is the first one, the precision is depicted as the third bar in Fig. 5. The forth bar represents the precision when the intention is "Crinstiano Ronaldo". The same search is fired to our semantic indexing. The result set is divided into two clusters representing the two "Ronaldo"s respectively. The correct image count of these two clusters is manually checked. The numbers are 111 and 145 respectively. The precisions of the two clusters are showed in Fig. 5 as the first and the second bars.

The statistics shows that the semantic indexing can cluster the polysemy terms into different concepts perfectly with the help of discriminable context information.

4.3 Ranking with colors

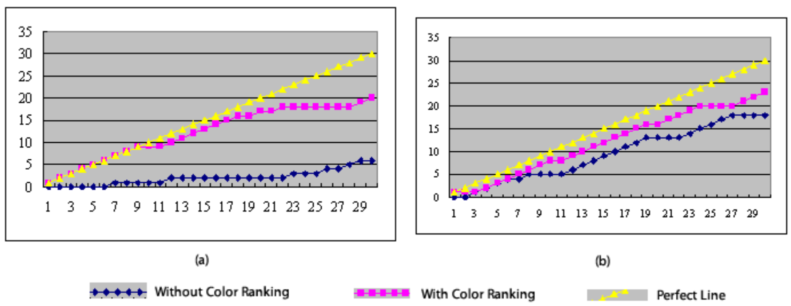

Fig. 6. (a)Search result of "Barcelona" (b) Search results of "Spain"

We select three teams to evaluate the ranking with colors. They are National Team of Spain, Club Team of Barcelona and Club Team of Juventus. The color characteristics of these teams are quite different. The color of Spain is distinguish (red), the colors of Barcelona are multiple (red and blue) and the colors of Juventus is indiscriminating (white and black which are dominances in most images). Before we start the evaluation, we set the context similarity threshold to a lower value to avoid the disturbance on the color evaluation.

The precisions are checked for both of the indexes with and without color as shown in Fig. 6(a) and (b). The X axis is the number of images checked. The Y axis is the number of relevant images in the checked images. Thereby the yellow line is the perfect line which represents the precision is 1. For "Juventus", we found that the ranking orders in both case are the same.

From the statistics, we can conclude that for those teams whose colors are not very indiscriminating the color ranking can improve the performance significantly.

5 Conclusion

Falcon-S extends the traditional search applications by an ontology-driven approach. The domain ontology and knowledge base are represented in OWL-Lite, and Sesame and OWLIM[3] are used in the implementation of Falcon-S. AJAX

[3] http://www.ontotext.com/owlim/

and SVG technologies are used to achieve a user-friendly interface. Besides, the information sources are from the World Wide Web. Continuous crawling and indexing jobs are running on different servers. Hence our information sources are from diverse owners, heterogeneous and never complete.

At the conceptual level, the ontology-driven approach and the knowledge base make Falcon-S having the following features:

- Searching objects by user-defined semantic queries
- Mapping search terms to objects by using contextual information
- Indexing images by domain objects
- Enhancing the search by multimedia content

The above features all together make the users of Falcon-S having a different experience in searching and exploring objects and images in specific domain, as compared with in using traditional search or query applications without Semantic Web technologies. However, more researches are deserved to pursue a general semantic search instead of domain-specific semantic search.

Acknowledgement

The work is supported partially by the 973 Program of China under Grant 2003CB317004, and in part by the JSNSF under Grant BK2003001. We would like to thank five graduate students in our research group, Lianghui Shi, Dan Li, Jun Ye, Yuanyuan Zhao and Wenhao Xu, for their work on implementation and experimentation. We are also grateful to Xiang Zhang, a PhD student in our group, for his valuable suggestion on this work.

References

1. Guha, R., R. McCool, and E. Miller, Semantic search. Proceedings of the twelfth international conference on World Wide Web, 2003: p. 700-709.
2. Aleman-Meza, B., et al., Context-Aware Semantic Association Ranking. Proceedings of SWDB, 2003. 3: p. 3.
3. Anyanwu, K. and A. Sheth, -Queries: enabling querying for semantic associations on the semantic web. Proceedings of the twelfth international conference on World Wide Web, 2003: p. 690-699.
4. Salton, G. and M.J. McGill, Introduction to Modern Information Retrieval. 1986: McGraw-Hill, Inc. New York, NY, USA.
5. Deerwester, S., et al., Indexing by latent semantic analysis. Journal of the American Society for Information Science, 1990. 41(6): p. 391-407.
6. Rui, Y., T.S. Huang, and S.F. Chang, Image retrieval: current techniques, promising directions and open issues. Journal of Visual Communication and Image Representation, 1999. 10(4): p. 39-62.
7. Athanasis, N., V. Christophides, and D. Kotzinos, Generating On the Fly Queries for the Semantic Web: The ICS-FORTH Graphical RQL Interface (GRQL). Proc. of ISWC, 2004.
8. Catarci, T., et al., An ontology based visual tool for query formulation support. Proc. of ECAI, 2004.