Information Retrieval from Sports Domain Ontology Using First-Order Logic

Nwe Ni Aung, Thinn Thu Naing
University of Computer Studies, Yangon
nwe.nna@gmail.com,thinnthu@gmail.com

Abstract

Ontologies are set to play a vital role in the "Semantic Web", extending syntactic interoperability to semantic interoperability by providing a source of shared and precisely defined terms. An ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, may be used to define the domain. Firstly, constructs concepts of sports domain ontology for sports named entities and sports related entities (objects), named entities relation (relationships) and entities instances (attribute) for the sports domain terms. In this sports domain ontology, addition to the relationships of concepts in the taxonomy, the non-taxonomic relationships between the concepts have been also built. In this paper, we present information retrieval from Sports Domain Ontology has done by using the First-Order Logic (FOL) and retrieve related concepts from it.

Key words: Semantic Web, Sports Domain Ontology, information retrieval, taxonomy, First-Order Logic

1. Introduction

One of the major weakness of current research on the Semantic Web (SW) is the lack of proper means to represent and reasons with uncertainty. A number of recent efforts from the SW community, the W3C and others have recently emerged to address this gap. Such efforts have the positive side effect of bringing together two fields of research that have been apart for historical reasons, the artificial intelligence (AI) and Semantic Web communities (SW). So, the potential research gains of this convergence is the development of the building ontology with reasoning references. Ontology is a formal, explicit specification of a shared conceptualization [2].

Ontological engineering is a field that studies the methods and methodologies for building ontologies. It studies the ontology development process, the ontology life cycle, the method and methodologies for building ontologies and the tools suited and language that support them. It offers a direction towards solving the interoperability problems brought about by semantic obstacles, such as the obstacles related to the identification of business terms and classes. Ontological engineering is a set of tasks related to the development of ontology for a particular domain [3].

Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web. There is overlap in the usage of the terms data retrieval, document retrieval, information retrieval, and text retrieval, but each also has its own body of literature, theory, praxis, and technologies. IR is interdisciplinary, based on computer science, mathematics, library science, information science, information architecture, cognitive psychology, linguistics, statistics, and physics. Automated information retrieval systems are used to reduce what has been called "information overload". Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible IR applications [4].

Firstly, build Sports Domain Ontology. This may be defined as follows: extract sports news from sports web sites. We choose the sports news, articles and journals as corpus domain and there exist a variety of sports named entities (objects), named entities relations (relationships) and instances (attributes). The ontology creation uses the Protégé-OWL editor. This enables users to build ontologies for the Semantic Web, in particular in the W3C's Web Ontology Language (OWL). And then, the Sports Domain Ontology concepts use as mapping for Sports terms. We first accept the user query and then, mapping with Sports Domain Ontology concepts and retrieve the sports information from the Sports Domain Ontology concepts using with First-Order Logic (FOL).

The rest of paper, section 2 reviews related work and section 3 architecture of sports domain ontology. Section 4 presents the proposed system architecture and section 5 describes the query information for First-Order Logic. Finally, concludes the paper in section 6.

2. Related work

V.Snasel et.al [8] presented a basic method of mapping LSI concepts on given ontology (Word-Net), used both for retrieval recall improvement and

dimension reduction. Experimental results for this method use in a subset of TREC collection, consisting of Los Angeles Times articles. Mapping terms on WordNet hypernyms improves recall, bringing more relevant documents. The LSI filtration enhances recall even more, producing smaller index. J.Saias et.al [6] reviewed the methodology to automatically create ontologies and class instances from documents. The ontology is defined in the OWL semantic web language and it is used by a logic programming framework, ISCO, to allow users to query the Semantic content of the documents. ISCO allows an easy and efficient integration of declarative, object-oriented and constraint-based programming techniques with the capability to create connections with external databases.

A.Wessman et.al [1] proposed the implementation of generalized Framework for an Ontology-Based Data-Extraction System. The nature of framework allows new algorithms and ideas to be incorporated into a data extraction system without requiring wholesale rewrites of a large part of the system's source code. It allows researchers to focus their attention on parts of the system relevant to their research without having to worry about introducing incompatibilities with the remaining components. This framework offers improved modularity and extensibility to support further data-extraction research. The framework is sufficiently developed to support a reimplementation of BYU Ontos that preserves the quality of legacy Ontos while also using modular heuristics code.

J.Paralic [5] described the domain knowledge representation schema in form of ontology. New resources registered within the system are linked to concepts from ontology. Resources may be retrieved based on the associations and not only based on partial or exact term matching as the use of vector model. The ontology-based retrieval mechanism has been compared with traditional full text search based on vector IR model as well as with the Latent Semantic Indexing method.

3. Sports Domain Ontology Architecture

The development of the sports domain ontology, we define the terms for sports information. It is designed in terms of the requirements for the identification of sports named entities and their relations. The sports ontology contains many types of sports – archery, football, baseball, badminton and weightlifting etc. Firstly, we identify the hierarchical taxonomy according to the respective sports activities. Hierarchical taxonomy can be concept, properties and attribute (instance). OWL Sports Domain is shown in figure 1.

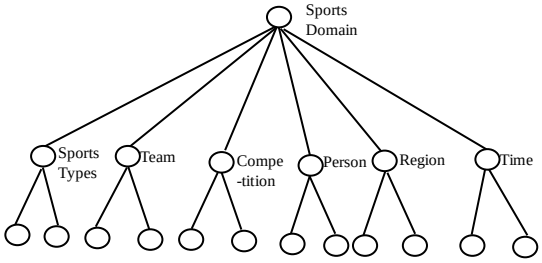

Figure1. Hierarchical taxonomy of OWL Sports Ontology

The domain concepts can be physical or abstract concepts. The physical concepts include material or equipment objects. Abstract concepts are Competition Name, Time, Region, Team Name etc. The relationships involves Sports Domain concepts involve Concepts_Concepts relations.

The attribute is the property of the concepts (class). It plays a role in the modification of words or phrases with concepts and relation between concepts. The sports types consist of many types of sports according to the competition name. For example, the sports types can be different between Games and Para Games. Region is the geographical place and it consists of many nations. All countries and most of the dependent territories in this region have their own sports association and sports teams. Team name is mostly the nation name and sports competition involves many teams according to the competition region. Time is the sports competition time and sports competition has done specific time. Person is the player who compete the competition and some players not only involve the member of national team but also X_member (foreign player) of the foreign hired club.

4. Implementation of proposed system

The proposed system implementation involves two parts. The first is the background theory - First-Order Logic (FOL) and the second is Proposed System Architecture.

4.1 Background Theory - First-Order Logic

We have identified the required terms (Concepts, relations between concepts and attributes) for Sports Domain Ontology. In this step, we construct the knowledge base. We also use the formal axioms; it is the components of the knowledge engineering in First-Order Logic (FOL). It models the sentences that are always true state. If we define the axioms correctly and completely describe the way the world works and the way that percepts are produced. And then, any complete logical inference procedures will infer the strongest possible description of the world state, given the available percepts. The steps of First-Order Logic are:

- identify the task
- assemble the relevant knowledge
- decide on a vocabulary of predicate, functions and constraints
- encode general knowledge about the domain
- encode a description of the specific problem instance
- pose queries to the inference procedure and get answers
- debug the knowledge base [7]

Sports Domain concepts. And then the last step is to retrieve the results. The retrieved results can be two types: query results for required information and results for the performance evaluation of the system. The proposed system architecture is shown in figure 2.

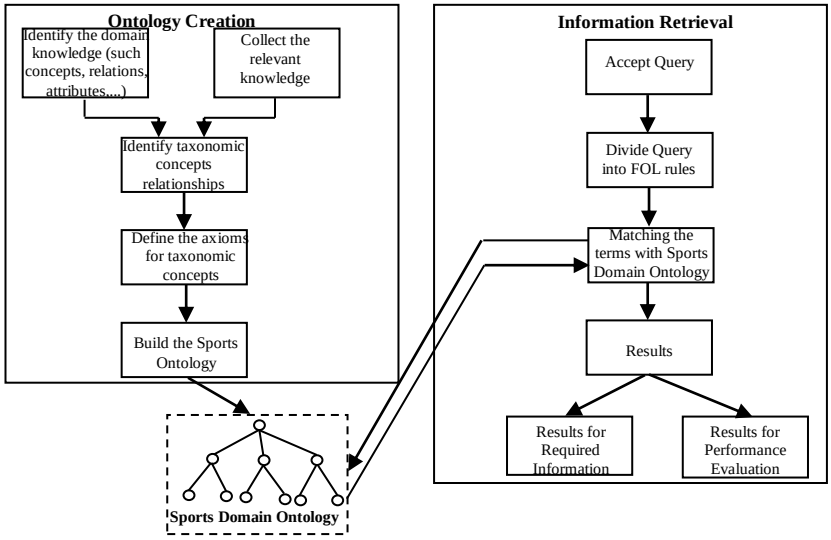

Figure 2. The system architecture of Ontology Creation and Information Retrieval

4.2 Proposed System Architecture

The implementation of proposed system consists of two parts. The first part is Ontology Creation and the second part is information retrieval from the Sports Domain Ontology. The first step of Ontology Creation is to identify the domain knowledge and collects the relevant terms of the domain (concepts, instances, attributes that represent concept properties, relation between concepts, etc. The second step identifies the hierarchical taxonomic terms. According to the hierarchical taxonomy, the sports name can be archery, baseball, badminton etc. The team name can be Brunei team, Indonesia team, or Timor Leste team etc. Others concepts may have their hierarchical concepts. The third step defines the inferences rules for semantic constraints to their relations between sports concepts. The final step is to build the Sports Domain Ontology according to the above steps defined.

The second part of proposed system is Information Retrieval. The first step is to accept the query from the users. Example queries show the next section. The second step is to divide the query into the First-Order Logic (FOL). The third step is the Sports Domain Ontology is used as mapping for

5. Query Information for First-Order Logic (FOL)

Query information of our system; use the First-Order Logic (FOL) to write queries for user request terms from the sports domain ontology. Firstly, we define the retrieved terms and then define the queries. They are:

s : define the terms for sports_name
c : define the terms for competition_name
Y : define the terms for date/time year
r : define the terms for region_name match in the competitions
t : define the terms for team_name

Query 1: The kinds of sports competed in Southeast Asia Games 2007.

$\forall c,Y \quad kinds\_of\_sports(c,Y) \Rightarrow [\exists\, c,Y \ \neg(c=Y) \wedge competition\_name(c) \wedge date/time(Y)]$

Query 2 : The regions name participates in Asian Games.
$\forall r,c$ participates$(r,c) \Rightarrow [\exists r,c \ \neg(r=c) \wedge$ region_name$(r) \wedge$ competition_name$(c)]$

Query 3 : The competitions name match in the European region.
$\forall c,r$ match$(c,r) \Rightarrow [\exists c,r \ \neg(c=r) \wedge$ competition_name $(c) \wedge$ region_name$(c)]$

Query 4 : The football Cup name compete for the Arsenal football team.
$\forall s,t$ compete$(s,t) \Rightarrow [\exists s,t \ \neg(s=t) \wedge$ sports_types$(s) \wedge$ team_name$(t)]$

Query 5 : The football team name compete for the Premier Leagues.
$\forall s,c$ compete$(s,c) \Rightarrow [\exists s,c \ \neg(s=c) \wedge$ sports_types$(s) \wedge$ competition_name$(c)]$

Query 6 : The date/time held in the World Games.
$\forall c$ held$(Y,c) \Rightarrow [\exists Y,c \wedge c \ \neg(Y=c) \wedge$ date/time$(Y) \wedge$ competition_name$(c)]$

6. Conclusion

This paper presented the Sports Domain Ontology architecture and the system implementation issues. We extract the information from online sports news web sites. We also identify concepts, relations and attributes for sports domain. In addition to the hierarchical taxonomy, we have also built the non-taxonomic relationships between top level concepts categories. And then, define the terms for building Sports Ontology.

We build the Sports Domain Ontology for mapping concepts with First-Order Logic terms. And then, retrieve data according to the mapping concepts. In this system, the information provided is of two kinds, Sports Domain concepts and relations between sports concepts. We believe that the information retrieval from Sports Domain Ontology using first-order logic can be applied efficiently and effectively in the community.

References

[1] A.Wessman,S.W.Liddle, D.W.Embley, "A Generalized Framework for an Ontology-Based Data-Extraction System"

[2] Ei Ei Han, Ni Lar Thein, "Concept for Myanmar Language Ontology", ICCA 2007

[3] http://en.wikipedia.org/wiki/Ontology (information_science)

[4] http://en.wikipedia.org/wiki/ Information_ retrieval

[5] J. Paralic, I.Kostial, "Ontology-based Information Retrieval"

[6] J. Saias ,P.Quaresma , "A methodology to create ontology-based information retrieval systems"

[7] S. Russell, P. Norving, "Artificial Intelligence A Modern Approach"

[8] V. Snášel, P. Moravec, J.Pokorný, "WordNet Ontology Based Model for Web Retrieval", *Proceedings of the 2005 International Workshop on Challenges in Web Information Retrieval and Integration (WIRI'05)*