2011 3rd International Conference on Information and Financial Engineering
IPEDR vol.12 (2011) © (2011) IACSIT Press, Singapore

Sports Information Retrieval with Semantic Relationships of Ontology

Nwe Ni Aung, Thinn Thu Naing

University of Computer Studies, Yangon

nwe.nna@gmail.com,ucsy21@most.gov.mm

Abstract. In the World Wide Web, Information Retrieval has become an everyday activity. The search engine associated with web browsers are used to find information relating to most domains of human activity. With the development of the Semantic Web, ontology is playing a vital role in many research areas. 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, and may be used to define the domain. Most of Ontology-based information retrieval systems use the concepts mapping. In this paper, information retrieval use semantic relationships between ontology of concepts. These ontology concepts based information retrieval cannot get the relevant semantic information and the retrieved results are not precise, ambiguous and require to clustering. By using these semantic relationships, they can retrieve more relevant information for sports information retrieval and these semantic relationship based queries are more correct results than concepts based queries. In this paper, we present information retrieval from Sports Domain Ontology has been using the First-Order Logic (FOL) rules and retrieve relevant semantic relationships between concepts from it.

Keywords: Semantic Web, Sports Domain Ontology, information retrieval, First-Order Logic

1. Introduction

The information retrieval system builds Sports Domain Ontology that provides to extract sports news from sports related web sites. Especially the football league is the main target of the system. 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 domain ontology have consisted of (32) classes, (20) Object properties and (9) Datatype properties. It uses the OWL DL (Description Logics) language features. For the implementation of the information retrieval model, the fundamental rules are applied. These fundamental rules are used as constraints for the system and based on the structure of the sports domain ontology. Sports Domain Ontology uses to map concepts and retrieve the sports information from the Domain Ontology using with First-Order Logic (FOL) rules.

First-Order Logic is more reasoning references than propositional logic or description logic by using the quantifiers. Reasoning of first-order logic systems are sound and complete and it is adequate for most everyday reasoning [3]. This information retrieval has been used Jena SPARQL (Simple Protocol And RDF

Query Language). Jena is a programmer's application programming interface for Java Semantic Web application. The primary use of Jena library is to help for writing Java code that handles RDF and OWL documents and getting the SPARQL facilities for retrieving sports information from OWL ontology [1]. We use many different kinds of semantic relationships based queries from the view of sports information retrieval system. These are five kinds of queries categorized by sports concepts and totally nine types of queries.

The rest of paper, section 2 presents architecture of Sports Domain ontology and section 3 demonstrates the system architecture. Section 4 describes detailed query information for First-Order Logic and finally concludes the paper in section 5.

2. Architecture of Sports Domain Ontology

The development of the sports domain ontology, define the terms for sports information. It is designed in terms of the requirements for the identification of sports named entities and their relations. Firstly, we identify the hierarchical taxonomy according to the respective sports activities. Hierarchical taxonomy can be concepts, properties and attributes (instances). OWL Sports Domain is shown in Fig 1.

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

The attribute is the property of the concept (class). It plays a role in the modification of words or phrases with concepts and relation between concepts. Region is the geographical place and it consists of many

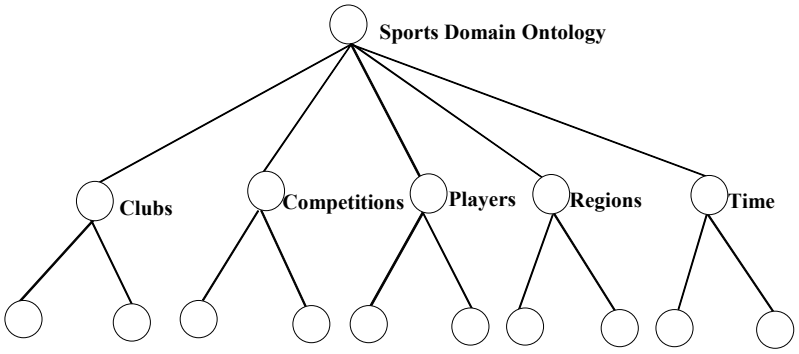

Fig. 1 : Hierarchical taxonomy of OWL Sports Ontology

nations. All countries and most of the dependent territories have their own sports association and sports teams (clubs/squads). Club name is mostly the nation name and sports competition involves many clubs/teams according to the competition of respective 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 currently playing in the hired club and the Ex_Member clubs/teams (previous time playing).

Moreover, the domain ontology contains the sports related objects – hasComposedof, hasDone, hasPerformed, hasPlayedAsExMember, hasMatched, hasQualifiedClubs etc. For each property for Sports Domain Ontology, we must define the class for its role describes. OWL Properties represent relationships between two individuals. There are two types of properties in OWL classes:

Datatype properties, relations between an individual to an *XML Schema Datatype value* or an *rdf literal* value.

Object properties, relations between an individual to an individual [2].

Table 1 shows the properties for the classes of sports related object that includes sports information attributes for competitions.

Table. 1 : Sample for Object Properties and Datatype Properties of Sports Domain Ontology

Properties	OWL Property	Data Type
hasComposedOf	owl:ObjectProperty	undefined
hasNationsName	owl:DatatypeProperty	String
hasExMember	owl:ObjectProperty	undefined
hasInvolved	owl:ObjectProperty	undefined
hasPlayedAsNational -ClubMember	owl:ObjectProperty	undefined
hasParticipated	owl:ObjectProperty	undefined
hasClubsName	owl:DatatypeProperty	String
hasClubsPlaying -Time	owl:DatatypeProperty	Time
hasPerformed	owl:ObjectProperty	undefined
hasPlayingtime	owl:ObjectProperty	undefined
hasPlayingEndTime	owl:DatatypeProperty	Time
hasTakenPartOf	owl:ObjectProperty	undefined
hasPlayedAsNational -ClubMember	owl:ObjectProperty	undefined
hasPlayerName	owl:DatatypeProperty	String
hasNationality	owl:DatatypeProperty	String
hasConcerned -Regions	owl:DatatypeProperty	String

3. The 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 and relation between concepts, etc. The second step identifies the hierarchical taxonomic terms. The clubs name can be Chelsea club, Everton club, or Barcelona club etc. These clubs also have their respective football competitions or football leagues. 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.

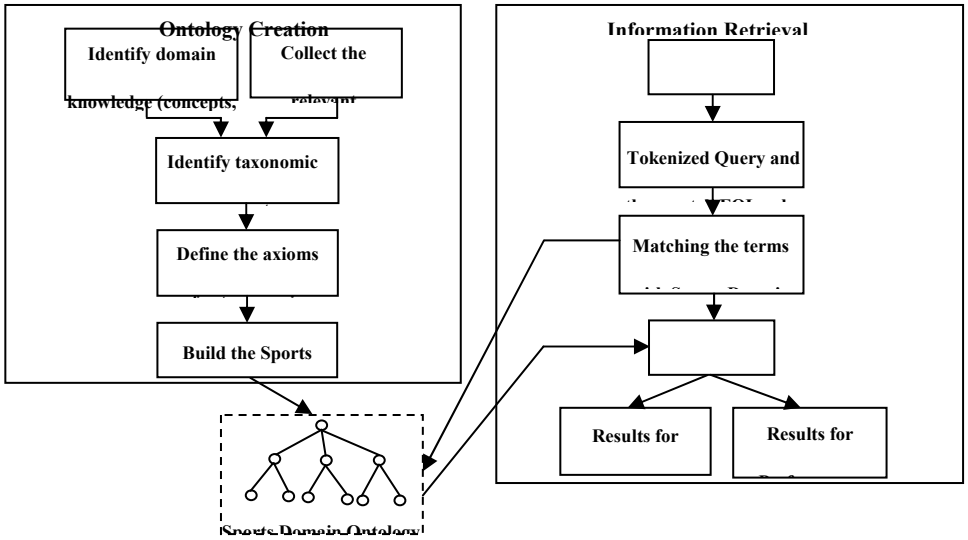

Fig. 2 : The system architecture of Ontology Creation and Information Retrieval

The second part of proposed system is Information Retrieval. The first step is to accept the query from the users. The second step is to tokenize these queries, match the First-Order Logic (FOL) rules. After getting the concepts of query, the third step is mapping with sports domain ontology terms. Finally, retrieves the results from the domain ontology. Example queries show the next section. 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 Fig 2.

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

Query information of our system; use the First-Order Logic (FOL) rules to write queries for user requested terms from the sports domain ontology. Firstly, identify the retrieved terms and these are applied as symbols for first-order logic. This system consists of five concepts, but regions concepts have the sub regions (nations) concepts. So, the information retrieval system has total in 6 symbols words. We define the terms as follows:

cl : define the terms for sports club_name (team_name or squad_name)

p : define the terms for player_name

c : define the terms for competition_name

r : define the terms for region_name

n : define the terms for nation_name

t : define the terms for competition playing date/time (year)

This query type gives the results for the clubs names involve in the specific nation. For example:

Query 1: "The football clubs names involve in Spain"

$$\forall cl,n \ \text{involve}(cl,n) \Rightarrow [\exists \ cl,n \ \neg(cl=n) \wedge \ \text{club\_name}(cl) \ \wedge \ \text{nation\_name}(n)]$$

Information retrieval system of Sports related Query - 1 can answer most of the sports information that can give the required information based on the semantic relationships. This query type retrieves the desired sports related objects "Clubs name and Sub region (nation) name" and then gives the results of the semantic relationships – *hasInvolved*.

This query type can retrieve the football clubs results for the specific football league (or) competition. For instance query:

Query 2 : "The football-clubs names corresponding with English FA Premier League"

To satisfy this query, the system matches the following condition and then gives the results.

$$\forall cl, c \ \text{corresponding}(cl,c) \Rightarrow [\exists \ cl,c \ \neg \ (cl=c) \wedge \text{club\_name}\ (cl) \wedge \text{competition\_name}\ (c)]$$

Query - 2 likes most of the sports related information that can give the required information based on the given clubs names. This query type can extract the desired sports related objects "Clubs name and competition name" that related the semantic relationships – *hasParticipated*.

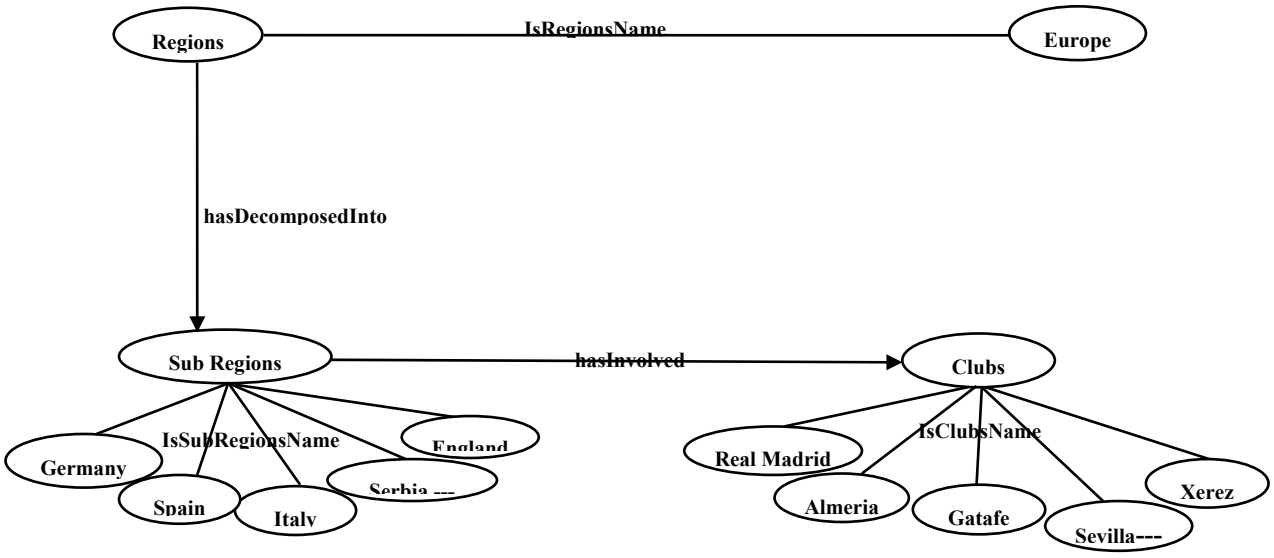

Fig. 3 : Conceptualization scheme of the retrieved results of Query 1

This type of query can answer the qualified football clubs results for FIFA World Cup in a specific competition time. Example query:

Query 3 : "The qualified clubs name in 2002 FIFA World Cup"

$\forall$cl,t,c qualifiedclubs(cl,[t,c]) $\Rightarrow$ [$\exists$ cl,t,c $\neg$ (cl=t) $\wedge$ $\neg$ (cl=c) $\wedge$ $\neg$ (t=c) $\wedge$ club_name (cl) $\wedge$ date/time(t) $\wedge$ competition_name (c)]

Sports related Query - 3 can give the results of semantic information that may be specific both clubs and competition concepts and can answer more precise information. This query can retrieve the desired sports related objects "Clubs name, date/time and competitions name" that related the semantic relationships – *hasQualifiedClubs*.

This type of query can give the qualified time of FIFA World Cup for user requested club. For instance:

Query 4 : "The date/time qualified in FIFA World Cup for Portugal squad"

This query can retrieve the desired sports related objects "date/time, Competition name and Club name" that related the semantic relationships – *hasQualifiedTime*. To satisfy this query, the system matches the following condition and then gives the results.

$\forall$t,c,cl qualifiedtime(t,[c,cl]) $\Rightarrow$ [$\exists$ t,c,cl $\neg$ (t=c) $\wedge$ $\neg$ (t=cl) $\wedge$ $\neg$ (c=cl) $\wedge$ date/time (t) $\wedge$ competition_name(c) $\wedge$ club_name (cl)]

This query type can retrieve the qualified football clubs results in FIFA World Cup for the user requested time and region. For example:

Query 5 : "The qualified clubs name for Asia region in 2010 World Cup"

To satisfy this query, the system matches the First-Order Logic and then gives the results.

$\forall$cl,r,t,c qualifiedclubs([cl,r],[t,c]) $\Rightarrow$ [$\exists$ cl,r,t,c $\neg$ (cl=r) $\wedge$ $\neg$ (cl=t) $\wedge$ $\neg$ (cl=c) $\wedge$ $\neg$ (r=t) $\wedge$ $\neg$ (r=c) $\wedge$ $\neg$ (t=c) $\wedge$ club_name (cl) $\wedge$ region_name (r) $\wedge$ date/time(t) $\wedge$ competition_name (c)]

This query can answer the sports related objects "Club name, Region name, date/time and Competition name" that related the semantic relationships – *hasInvolved* for a specific region.

This type of query can answer the football clubs names that are concerned with the user requested player and the qualified time of the national club in FIFA World Cup. Example of query:

Query 6 : "The clubs names belong to Kaka' and the qualified time of the national team in FIFA World Cup"

$\forall cl,p,t,c$ belongto([cl,p],[t,c]) $\Rightarrow$ [$\exists$ cl,p,t,c $\neg(cl=p) \wedge \neg(cl=t) \wedge \neg(cl=c) \wedge \neg(p=t) \wedge \neg(p=c) \wedge \neg(t=c) \wedge$ club_name(cl) $\wedge$ player_name(p) $\wedge$ date/time(t) $\wedge$ competition_name(c)]

Sports related Query 6 can give the semantic information that may be specific both player name and competition concepts and can answer more precise information. This can retrieve the desired sports related objects "Clubs name, Player name, date/time and competitions name" that are related the semantic relationships - *hasPlayedAsExMember, hasPlayedAsMember*, *hasPlayedAsNationClubMember* and *hasQualifiedTime*.

This type of query can give the football clubs names related to the specific football player. For instance:

Query 7 : "The footballclubs names concern with David Beckham"

Query 7 can answer the desired sports related objects "Clubs name and Player name" that related the semantic relationships –*hasPlayedAsExMember, hasPlayedAsMember* and *hasPlayedAsNationClubMember.*

To satisfy this query, the system matches with First-Order Logic rules and then gives the results.

$\forall cl,p$ concern (cl,p) $\Rightarrow$ [$\exists$ cl,p $\neg(cl=p) \wedge$ club_name(cl) $\wedge$ player_name(p)]

This query type can retrieve the football clubs names that are concerned with the user requested player and the qualified time the player in FIFA World Cup. For example:

Query 8 : "The football_clubs regard to John Terry and the date-time who played in FIFA World-Cup"

To solve this query, the system matches the following condition and then gives the results.

$\forall cl,p,c,t$ regard([cl,p],[c,t]) $\Rightarrow$ [$\exists$ cl,p,c,t $\neg(cl=p) \wedge \neg(cl=c) \wedge \neg(cl=t) \wedge \neg(p=c) \wedge \neg(p=t) \wedge \neg(c=t) \wedge$ club_name(cl) $\wedge$ player_name(p) $\wedge$ competition_name(c) $\wedge$ date/time(t)]

Query 8 can give the desired sports related objects "Clubs, Players, Competitions and Time" related the semantic relationships – *hasPlayedAsExMember, hasPlayedAsMember*, *hasPlayedAsNationClubMember* and *hasPlayingTime*. This query can give more precise and relevant semantic information for the users.

This query can give the player information of the competitions name, clubs name and the competition playing time by year. For example:

Query 9 : "The competition name by time vying the football player Ronaldinho"

This query can answer the specific types of the semantic relationships – players and clubs, clubs and competitions, competitions and time and players and time that relates to these concepts. In this query, retrieves the results of the semantic relationships - *hasPlayingTime*. To satisfy this query, the system matches the following conditions and gives the results.

$\forall c,t,p$ vying(c,[t,p]) $\Rightarrow$ [$\exists$ c,t,p $\neg(c=t) \wedge \neg(c=p) \wedge \neg(t=p) \wedge$ competition_name(c) $\wedge$ date/time(t) $\wedge$ player_name(p)]

5. Conclusion

This paper presented the Sports Domain Ontology architecture and sports information retrieval with semantic relationships between concepts. We have identified concepts, relations and attributes for sports domain. We have also built the hierarchical taxonomic and non-taxonomic relationships between top level concepts categories. And then, define the terms for building Sports Ontology. The proposed ontology is characterized by qualitative and quantitative including semantic relationships.

These semantic relationships are used for mapping concepts with predefined First-Order Logic rules and retrieve the results according to the mapping concepts. Information retrieval system using semantic relationships based queries can give more semantic meaning results than concepts based queries. This system not only support more comprehensive ontology for sports domain but also share, reuse and process these

domain knowledge among the users. We believe that the information retrieval from Sports Domain Ontology using First-Order Logic (FOL) rules can be efficiently and effectively applied in the user community.

6. References

[1] G.Michael, "RDF, Jena, SparQL and the Semantic Web", *Proc of the 37th annual ACM SIGUCCS fall conference*, St. Louis, Missouri, USA, pp. 131-138, 2009

[2] M.Horridge, H.Knublauch, A.Rector, R.Stevens, C.Wroe, "A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0", August 2004, pp. 16-81

[3] S.Russell, P. Norving, "Artificial Intelligence A Modern Approach", Second Edition, Prentice Hall, 2003