



*Journal of Advances in
Science and Technology*

*Vol. IV, Issue No. VIII,
February-2013, ISSN 2230-
9659*

MAGNIFICENCE OF COGNITIVE ONTOLOGY FOR SEMANTIC INFORMATION RECOVERY

AN
INTERNATIONALLY
INDEXED PEER
REVIEWED &
REFEREED JOURNAL

Magnificence of Cognitive Ontology for Semantic Information Recovery

Suman Choudhary¹ Dr. Vijay Parkash Agarwal²

¹Research Scholar, Deptt. Of Computer Science, Singhania University, Pacheri Bari, Jhunjhunu (Raj.).

²Assistant Professor, Deptt. Of Comp. Sc. & App., Kurukshetra University, Kurukshetra (HR)

Abstract – Information is Knowledge: Knowledge is Wealth .The combinations of concept with its associative relation, in a quantitative sense is said to be valid information. Relevant information retrieval through cognitive process is our main objective of this paper. The specification of Concept with its relation in an organized way is said to be ontology. Here we use Cognition based ontology for information retrieval which can be implemented in a semantic based information retrieval system.

The main contribution of this paper has been organized in three phase. One is the creation of domain ontology where, here we are using event based sports ontology. The second phase is the conversion of HTML pages through a designed Web Crawler to an entity based XML page and the third phase is the mapping of the entity XML to sports domain ontology. This domain ontology can be used for the refined Semantic information retrieval system.

Keywords: Ontology, NLP, Syntactic Analysis, Semantic Analysis, OWL, Machine Learning

INTRODUCTION

Information retrieval is one of the active research areas as the number of users, searching for knowledgeable information and also the amount of information available on internet is increasing exponentially in day-to-day basis. So, we are in need of intelligent information retrieval system.

A retrieval system will be more effective if the machine understands the user query .Thus we required a semantic based information retrieval system.

The sole part of semantic based information retrieval system is based on Ontology. Ontology is an organized way of representing the entities and their relationship in a domain. Ontology formally defines different concepts of a domain and relationships between these concepts.

The recent Web Ontology Language (OWL) [14] has become a popular standard for data representation and exchange. The OWL supports the representation of domain knowledge using classes, properties and instances for the use in a distributed environment as the World Wide Web.

The main part of any semantic information retrieval is the recognition of what the user wants to search for. A human response to a query through his cognitive

process. Cognition is the process of endless decision making without any perception but through semantic knowledge experience and sense.

In this paper we have designed a framework for semantic information retrieval system using cognitive ontology concept. This paper is organized in such a way that first the overall framework has been explained, next the specific domain ontology is briefed, then the modules of the framework has been explained.

REVIEW OF LITERATURE

In this section, several strategies for deriving Ontology's [20] from heterogeneous XML data sources have been disused. Some approaches targeted either more on a general mapping between XML and RDF others aim at mapping XML Schema to OWL. OWL and RDF are much of the same thing, but OWL is a stronger language with greater machine interpretability than RDF. OWL comes with a larger vocabulary and stronger syntax than RDF.

In [4] the author discussed some XML rule pattern, if the pattern matches it will be converted to the mapped OWL attribute and in [8] discuss the way of converting the HTML to OWL using table. They consider the TABLE tag of HTML page and tried to convert to OWL both of this paper result won't produce any semantic to the ontology. In [9] they tried

to convert the HTML to OWL using the FRAME set tags they also tried to incorporate UML to identify the class and subclasses. In [10] the conversion is done by first annotating the web page. The annotation they consider is the semantic annotation thus they tried to provide semantic of the page. They use the tool called GATE to analyze the semantic through natural language processing. In [13] the conversion is take place using the tag and they used GRDDL tool for conversion. In [5] they extend the XML to the existing OWL using a tool JXML to OWL. In [11] the reusability of created OWL from HTML from forms has been discussed. The overall categories of Ontology based search engine as explained is discussed in [12].

COGNITIVE ONTOLOGY BASED INFORMATION RECOVERY FRAMEWORK

To design an information retrieval system as similar to that of human mind is a challenging research area which comes under computer vision. In [15] [16] the concept of semantic information retrieval using the idea of ontology has been used.

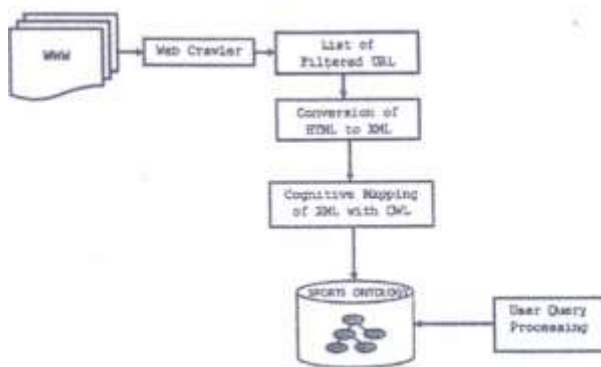


Figure 1 Proposal Framework

Figure 1 shows the overall view of the proposed Cognitive Ontology based information retrieval system

The Sports Ontology is a repository of the domain ontology created for our work. Here we have created ontology by considering 10 events. The information from World Wide Web is been collected through a programmed web spider which is also called as a web crawler. It is programmed in such a way that it would filter out only most of the listed 10 sports related event web pages. Those pages are given to a conversion process of html to xml converter from which the entity related XML is created. Then the intelligent process of mapping the generated XML with our Sports domain ontology is processed. So, for a user query the required result are navigated through the mapped ontology.

In following sub-section the creation of sports event ontology, HTML to XML conversion and the Cognitive mapping of XML to OWL mapping is discussed in detail.

A. Sport Event Ontology

Ontology is an organized way of representing the entities and their relationship in a domain. This ontology can be represented using OWL the Web Ontology Language. At first few decades after introducing the concept of Semantic web technologies Resource Description Framework, RDF [1] is used to define the ontology of an domain, where the hyperlink in a html page is replaced Then in second decade of ontology the concept of RDFS that is RDF Schema is used which introduce the concept of creating associate relationship between two entities. Then by 2004[2] W3C recommend OWL which is the combination of RDF, RDFS and some of cardinal expression which is missing out in RDFS which also help us to build more complex domain ontology. Fig.2 shows the structure of OWL



Figure 2: OWL language structure

The sports event consider to create the sports ontology are cricket, croquet, marathon, polo, rowing, Mountain climbing, sailing, ski, and tennis and volley ball.

The sports event ontology is designed by classifying into two different classes at the top level which is shown in Fig.3

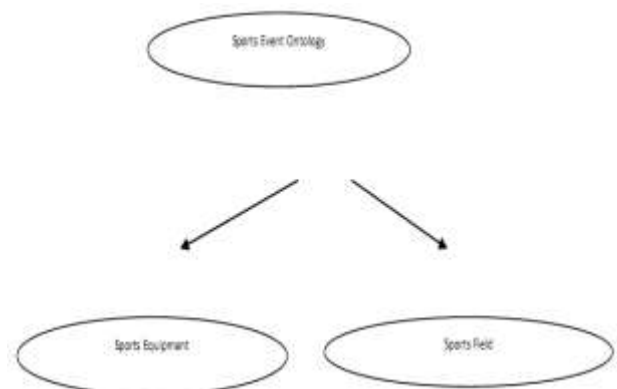


Figure 3: Sports Event Ontology- level 1

The Sub-classes of sports field class are:

- Grass field
- Water field
- Road field
- Indoor stadium
- Rocky field
- Snow field

The Sub-classes of sports equipment class are:

- Ball
- Bat
- Boat
- Hoop
- Mallet
- Mast
- Net
- Oar
- Pole
- Racquet
- Rope
- Stump
- Skis

All the above entities are defined as classes and its sub classes. The 10 events are specified as individuals as they have to be related to all the two major classes. In Table 1 shown the Individual (instants) created for the sub class of the class Sports Field is specified. In Table 2 the individual created for the sub class Sports Equipment is shown

TABLE I

SPORTS FIELD CLASS - INDIVIDUAL RELATION

Classes	Individual
Grass field	Cricket
	Croquet
	Polo
Water field	Sailing
	Rowing
Road field	Marathon
Rocky field	Mountain Climbing
Indoor Stadium	Tennis
	Volley ball
Snow field	Ski

TABLE II

SPORTS EQUIPMENT CLASS - INDIVIDUAL RELATION

Classes	Individual
Ball	Cricket
	Croquet
	Polo
	Tennis
	Volley ball
Bat	Cricket
Boat	Sailing
	Rowing
Hoop	Croquet
Mallet	Croquet
	Polo
Mast	Sailing
Net	Tennis
	Volley ball
Oar	Rowing
Pole	Ski
Racquet	Tennis
Rope	Mountain
Stump	Cricket
	Croquet
Skis	Ski

With this class, sub-class and individual hierarchy organization we have to specify the object property of each individuals. The identified object properties are listed in Table 3

TABLE III

SPORTS EVENT ONTOLOGY INDMDUAL'S OBJECT PROPERTIES

Class	Individual
Cricket	has_cricketball / has_cricketbat
Croquet	has_croquetball / has_croquetmallet
Polo	has_polo / has_polomallet
Rowing	has_rowinggoar / has_rowingboat
Sailing	has_sailingboat / has_sailingmast
Mountain Climbing	has_rope
Ski	has_skipolo / has_skiskies
Tennis	has_tennisball / has_tennisnet
Volley Ball	has_volleyball / has_volleyballnet

B. HTML TO XML CONVERSION

The first phase of this Framework is the conversation of all the web page collected from the designed web crawler to a standard XML files with name entity concept. Name Entity Recognition is a concept of Natural Language Processing. In short it is called as NER. The main technology [13] used here are patterns and Lexicons. For the given Corpus text NER classifies the entity as Person Name, Organization Name, Location and Miscellaneous (Date, Time, Number, Percentage, Monetary expression, Number expression and Measurement expression)

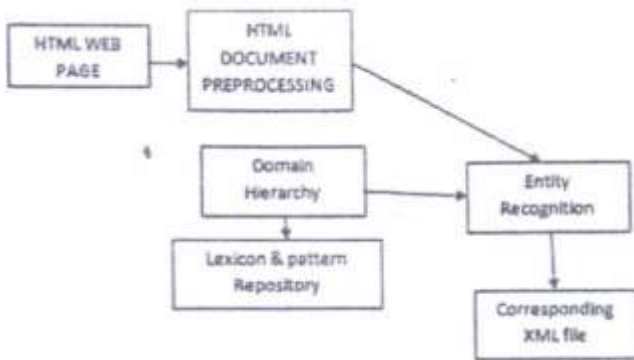


Figure 4; Entity based XML Conversion

Fig. 4 shows the general framework for converting HTML document to XML using Name Entity concept this technique is derived from [3].

In this process a domain hierarchy is created manually as per the domain knowledge, where the possible knowledge on the entities listed is hierarchy organized. When an HTML page is given as input to this process the unwanted advertisement and post are all removed with the help of DOM of the HTML page. With the preprocessed page the entities are all matched through the lexical pattern with the help of Domain Hierarchy. Thus identified pattern are list under their corresponding tag such as <organization> with the instant tag.

C. XML to OWL Mapping

In general and well formed XML can be converted or mapped to ontology through a syntactic way. In [4] & [6] the general rule of converting an XML Schema to ontology is explained. An overview of those mapping is tabulated in table 4

TABLE IV

XML TO OWL MAPPING

XSD	OWL
Xsd: elements, containing other elements or having at least one tributes	Owl: Class, coupled with
Xsd: elements, with neither elements nor attributes	Owl:DatatypeProp
Named Xsd: comoplexTyoe	Owl:class
lNamed xsd:SimoleTyoe	Owl: Datatvoe Prooerties
Xsd: minOccurs,xsd:maxOccur	Owl:min Cardinality, owl:max Cardinality
Xsd: sequence,xsd:all	Owl: intersection Of
Xsd: choice	Combination of Owl:unionOf and

In [17] the mapping of XML to OWL has been discussed for a B2B domain where the heterogeneous classes and attributes are employed.

The concept of this mapping is been explain in my paper [7]. The lack in this kind of conversion is that here the XML is not converted in a semantic way [18]. Thus here we designed a framework were we tried to convert the XML in a meaningful way.

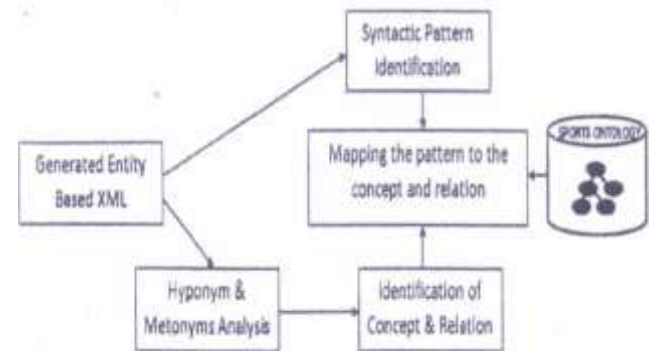


Figure 5: Semantic based XML- OWL Mapping

As per Fig.5, from the generated XML, the OWL classes, sub-classes, Object property and data property are identified as per the pattern explained in Table 4.

In parallel to that hyponyms and metonyms are analyzed. Where hyponyms are, semantically related words. Thus it provide a semantic relationship “is-a” between two related concepts or terms. For example has_cricket_batting_ball and has _ cricket_ bowling ball are semantically related. Meronyms provides the “part-of” relationship between the concepts or terms for example has_croquet_hoop is a part of has_croquet_mallet. Likewise the “part-of” and “is-a” relationship between all the classes and object properties is been analyzed over here thus to provide a semantic knowledge to the semantic search system.

Thus the Semantically identified concepts and it relation is mapped with the syntactically generated pattern. Thus generated XML is mapped with the created sports domain ontology.

Searching the ontology is done by SPRQL [19] or OWL- QL language. Logical descriptive language is one of the strong foundations of ontology. These query language is basically used to search ontology for relevant result of each classes in the ontology is represented by a specific URL.

For the given textual keywords or a sentence the probability reasoning concept is used to analyze the query to extract the main concepts, relation words form the query with that, we can search the Sports ontology for the related URL.

CONCLUSION

The main objective of this paper is the mapping of existing HTML page to the domain Ontology for this an entity based XML conversion and semantic based OWL mapping is used. Such created ontology can be used for semantic search system.

The future part of our work [22] is to employ a multirnode base ontology search [21], where a user can give both text and image as query.

REFERENCES

- [1] Linyang Yu "Introduction to the Semantic Web and Semantic Web Services" Chapman & Hall/CRC Taylor & Francis Group 2007.
- [2] Dean, M. and Schreiber, G., 2004, OWL Web Ontology Language Reference, W3C Recommendation, <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>.
- [3] Jianhan Zhu, Victoria Uren, Enrico Motta "Espotter, Adaptive named Recognition for web browsing" (2004).
- [4] Ivan Bedini, Christopher Matheus and Benjamin Nguyen "Transforming XML Schema to OWL using Patterns" 2010.
- [5] Toni Rodrigues, Pedor Rosa and Jorge Cardoso "Mapping XML to Existing Owl Ontologies" 2006 Hannes Bohring and Soren Auer "Mapping XML to OWL Ontologies" 2005.
- [6] G.Nagarajan, Dr. K.K. Thyagarajan "Linguistic Conversion of Syntactic to Semantic Web Page" The Second International conference on Advances in Computing and Information Technology 2012, Publish on Springer.
- [7] Alessandro Lenci et al "NLP based ontology learning from legal texts. A case study" (2006)
- [8] Yuri A. Tjjerino et al "Towards ontology generation from tables" Kluwer academic publishers (2004).
- [9] Sidi Benslirane et al "Towards ontology extraction from data intensive web sites: An htmlforms based reverse engineering approach" International Arab journal of information technology (2012).
- [10] Debajyoti, Mukhopadhyay et al "A New semantic web services to translate HTML pages to RDF" Int, Conference of IT (2010).
- [11] Hoon Hwangbo et al "Reusing of information constructed in HTML document: A conversion of HTML to OWL" Int. conference on control, automation and systems (2008).
- [12] Kyurnars Sheykh Esmaili, Hassan Abolhassani "A Categorization scheme for semantic web search engines" (2005).
- [13] Jianhan Zhu, Victoria Uren, Enrico Motta "Espotter. Adaptive named Recognition for web browsing" (2004).
- [14] P. Hitzler, M. Krotzsch, B. Parsia, P.F Patel-Schneider, and S. Rudolph, "OWL 2 Web Ontology Language Primer", W3C Recommendation 27 October (2009).
- [15] Mirian Fenandez et al "Semantically enhanced information retrieval: An ontology based approach" Elsevier's Web semantics: Science, Services and Agent on the World wide web. Vol. 9 issue 4, dec (2011), pp: 434-452.
- [16] Soner kara et al "An ontology based retrieval system using semantic indexing" Elsevier's Information systems Vol. 37, issue 4, June (2012) pp 294 – 305.
- [17] Jorge Cardoso and Christoph Bussler "Mapping between heterogeneous XML and OWL transaction representations in B2B integration" Elsevier Data & Knowledge Engineering Vol. 70, Issue 12 Dec(2011) Pg: 1046- 1069.
- [18] Damine, Lacoste et al "An efficient XML to OWL converter" ACM ISEC'11 Proceedings of the 4th Indian software engineering conference (20U) Pg: 145 – 154.
- [19] Rashmi Chauhan et al "Ontology based automatic query expansion for semantic information retrieval in sports domain" Communications in computer and information science volume 305, (2012), pp 422-433.
- [20] G.Nagarajan, Dr. K.K. Thyagarajan (2010) "A Survey On the ethical implications of Semantic Web Technology" Journal of Advance Research in Computer Engineering: An International Journal, ISSN: 0947-4320 Vol4, Nurnber 1, pp: 123-133.
- [21] Yanti Idaya Aspura Mohd Khalid et al "Towards multimodality ontology image retrieval" LNCS Visula informatic: sustaining

research and innovations (2011) Vol. 7067 pp
382 – 393.

- [22] G. Nagarajan & K.K. Thyagarajan “Cognitive ontology enrichment using natural language text captions to enhance semantic information retrieval” paper submitted to Indian Journal of Computer science for (2012) December issue.