Pakistan Journal of Science, 71 (4 Suppl.): 2019, Page 216-221
ISSN: 2411-0930

SEMANTIC BASED SEARCH USING CATALOGUE ONTOLOGY

M. Khan, S. Jan and F.U. Khan

Deparment of Computer Software Engineering ,University Of Engineering & Technology Mardan, Pakistan.
*Corresponding author's E-mail: khan.mehwish@outlook.com

ABSTRACT: As the digital information size grows in every field of life, its efficient retrieval has become the most vital issue. For this purpose, many approaches have been proposed and investigated. The use of semantic web technologies has been proven one of the most successful techniques in many information retrieval fields. In this paper, we have proposed and implemented an ontology-based search mechanism for library catalogues by developing a scalable catalogue ontology. The existing systems search for book titles, using string-based syntactic search which gives a positive result if and only if a user remember the exact title or other related data. In our proposed system, the knowledge in our created ontology is used in order to search the most relevant titles even if a user does not know anything about the data related to books. In this regard, we have created a catalogue ontology where titles are classified using major domain areas and then subdomains subsequently. The ontology considers all the important concepts, data and object properties, their relationships and possible restrictions. The ontology is implemented in a search system and extensive and diverse experiments were carried out. In this way, user is able to search books based on categories to get only relative searches instead of irrelevant searches. The results show that ontology based search outperform the syntactic systems. The comparisons are shown in terms of precision, recall and F-measure.

Keywords: Semantic, Ontology, Book management system, digital libraries.

INTRODUCTION

In existing systems, search has been done by exactly matching of titles or category and is based on rational database. So, searching book titles with the same meanings is difficult and it is also not possible to relate same contents with different titles. This syntactic search results in a lot of irrelevant results and we have to skim retrieved data to get our required data. This whole process takes a lot of time. So, there should be a way through which we can organize data in such a manner that users can search the content easily and accurately. By using semantic concepts we can retrieve the exact required data. The proposed approach goes for allowing searchers to use the semantics of their data needs in an unambiguous manner and get exactly the same results what they want. The main purpose is to develop ontology for library catalogue to make our search more easy and accurate. The semantic organizes information and data in structures called "ontologies". Data, that is connected, is put away as taxonomy. Ontologies are used for demonstrating a unique modelling of the real world problem and semantic concepts (Giri, 2011). This makes it easy for search engines to retrieve all the relevant data by using the semantic concepts. Semantic Web vision is about aiding resource discovery by creating tools to help searcher's refine and develop their searches, and to aid in the navigation of search results.

This study focuses on making a library catalogue which enables users to search books on the basis of categories. The searching result will contain only relative searches instead of irrelevant results. Hence searching will be easier and relevant for the users by using semantic concepts. As this ontology includes subject of various level of specificity, so navigation of the contents across the ontology will be done based on subject category and hence user can easily browse the broader and narrower terms. Hence, we can say that searching information from a semantic portal can benefit from more intelligent search-operations than traditional keyword-based searches. The searching result will contain only relevant instead of irrelevant results. Searching will be easier and relevant for the users by using semantic concepts. As this ontology includes subject of various level of specificity, so navigation of the contents across the ontology will be based on book category. The proposed system is used to get efficient and effective book search experience.

Ontologies describe the relationship between different things and allow definitions of the properties as well. By implementing semantic concepts, we can search the results in proposed system more accurate and efficient. The semantic can define classes, subclasses and relationships with keyword that are searched at the same time. Currently used library systems are not able to check consistency between classes. Ontologies overcome this problem efficiently and effectively with the help of reasoning services. This paper shows that ontology based library classification schemes are better than relational database library management system. For the evaluation of results or syntactic search ordinary search engine is used and for that of semantic search SPARQL queries are

used to fetch desired results. SPARQL is used to both query and analyze data. These results are then evaluated by using precision, recall and F1 score.

MATERIALS AND METHODS

Machines are unable to understand and interpret the meaning of the information in natural-language form and that is how most of the Web information is represented nowadays. A solution to this problem is provided by the third basic component of the Semantic Web, collection and organizing of information called ontologies. Semantic is an intelligent and meaningful approach to search exactly what we want. It describes the things in a way that the computer can easily understand. Ontology plays an important role in achieving goals of semantic i.e. how to use, reuse and process knowledge that can be interconnected across applications systems.

There are several studies where ontologies have been used to manage and process the knowledge (Tran *et al.*, 2007; Kara, *et al.*, 2012; Jan *et al.*, 2011). In few domains, the use of ontologies has shown great results. Classification based ontologies can be very helpful for practical implementation of semantic web. This fact was proved by Giunchiglia et al. by converting the generic classification schemes into OWL ontology. ScholOnto, is an ontology-based digital library server used to support scholarly interpretation and discourse. It enables researchers to describe and debate through a semantic network. It proposes that researchers enrich their texts with nodes and links which they add to an evolving semantic network reflecting the relevant literature.

A browsing and searching personalization system for digital libraries based on the use of ontologies for describing the relationships between all the elements which take part in a digital library scenario. It is now widely believed that ontologies have an important role in achieving the goal of machine understandable, also known as semantic web. Several methodologies have been proposed to develop ontologies. A number of ontology engineering methodologies have been proposed, still the field lacks widely accepted and mature methodologies. There is not even a single completely mature methodology (Iqbal *et al.*, 2013).

British Library developed a linked data instance of the British National Bibliography (BNB) to increase commitment from the UK Government since 2009 and for opening up public data for re-use. The main purpose is to incorporate the benefits of linked data in British National Bibliography (Deliot, 2014).

It is not important either the data store is in the form of MADS, MARC, RDF, or XML, it can be converted into another form by using different techniques. A centralized database can be created from the various files available in different library communities.

Using semantic concepts to the digital library delivers the significant information and data that may be accessed without any human interaction. T. Krishna and et.al. Proposed a cost effective web enabled organization document repository framework consisting of Digital Library and related application. They used open source tools to implement a prototype of their model which gives a cheaper way to introduce Digital Library based document repository system using efficient knowledge management. After the emergence of semantic web, ontology was the only efficient way for knowledge management, sharing and processing in verity of situations and applications.

Main purpose of Ontologies is implementing domain knowledge in a generic manner and offers a generally approved understanding of a domain. While developing ontologies one should have to face the difficulty in engineering as well as in the maintenance of ontology. There is not only a single correct way to develop ontology, so the best solution always depends on the type of application for which we are going to develop an ontology. For ontology development we use different tools like protégé. Protégé is used to define classes, class hierarchies, properties and the relationship between properties and relationship among classes. Different tools are used for the development implementation and evaluation of ontology. One of these tools is Jena API which has java, PHP and .net versions. Jena2 provides integrated implementations of the W3C Semantic. Another version is Jena.net. Jena .NET is used for developing semantic applications in C#.

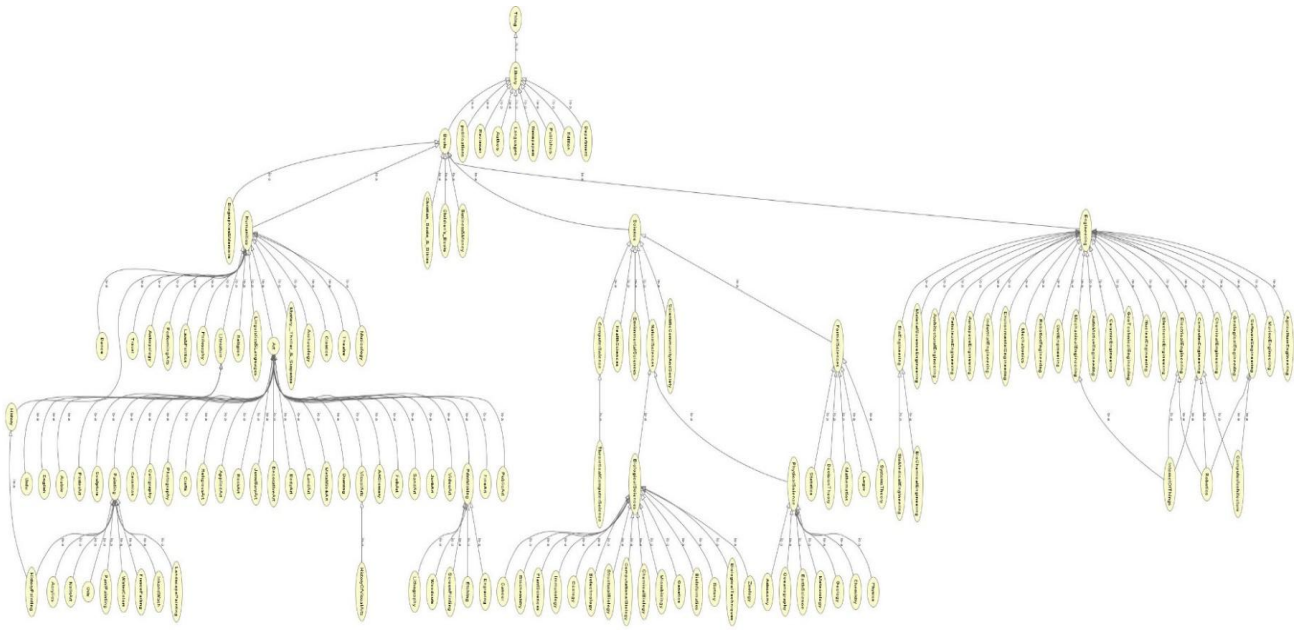

Figure-1: Pictorial Representation of catalogue ontology

Ontology consists of concepts, relationships, attributes and individuals. Concepts are abstract groups or types of objects. Relationships are relations between objects. Attributes are basically concepts that belong to classes. Individual usually representing real-world objects. They contain values (attribute instances) and are connected with relationship instances.

Ontology is created using protégé having API version 3.4.2. Ontology development focuses on the creation, editing and management of ontology. The language used for the development of ontology is OWL (web ontology language) which is knowledge representation language for creating ontologies. OWL is an extension of RDFS to use more expressive semantics in ontologies. Web Ontology Languages are built upon a standard of the World Wide Web Consortium called Resource Description Framework (RDF). The OWL family has evolved to include many syntaxes and specifications, and has attracted a great deal of interest in fields such as medicine and academia.

Ontology is developed by following the World Wide Web Consortium's (W3C) XML standards. Ontologies contain concept, data, object properties, relationships, and some restrictions. Catalogue Ontology consists of 515 axioms, 141 classes, 16 object properties, 23 data properties, and 252 logical axioms. Top down concept is used in our ontology starting from most general classes and then going to the most specific classes. The specific classes are subclasses of any of the general classes and one class may be the sub class of one, two or more classes. This hierarchy defines the general domain which we want to model. This hierarchy explains that instance of subclass will be consequently member of superclass. Book categories are divided into classes and sub classes on the basis of their relevancy. These classes share common properties if they belong from the same category. Classes are further divided into sub classes. For example, “Dance” is sub class of “Humanities” which is in turn subclass of more general class “Book”. Properties for each class have also been defined. For example, “Humanities” class has properties “has Author”, “hasVersion”, “hasPublisher” and some others like these. All the subclasses of “Humanities” also have all of its properties. Figure 1 shows pictorial representation of ontology indicating that how class are related with each other and what relations exist among them.

Ontology is processed using SPARQL queries. Writing a query using SPARQL requires the knowledge of the domain ontology as well as the syntax of the language. Therefore, Semantic Researchers works on simplifying the process of query writing for the end-user. The semantic query has four categories, i-e keyword-based, form based view-based and natural language-based systems.

The user will enter the book category as keyword to search the all the relevant books. SPARQL queries are executed to get the results. Then these results will be compared with syntactic search. Same dataset will be used for both semantic and syntactic searches. Data, used in the ontology, is added in a database and then queries are run against this data. For example, user will use “Humanities” as keyword to get all the relevant books associated with this category.

Ontology is evaluated using SPARQL queries. Being a semantic based query language SAPARQL will displays required results very efficiently and hence highly

scalable. The queries are used to find the sibling, super and sub categories from the list of book catalogue using different combinations. A dataset of 75 randomly chosen keywords are used in each case to evaluate that how efficient and accurate our search system is. Same data set is used for syntactic search too. A simple interface is developed using java. This traditional search will only return exactly matches with keyword and skip many relevant data. At the end we will evaluate our search by considering F1 measure. The formula used to calculate F1 measure is:

F1 = 2 * (precision * recall) / (precision + recall)

The whole process starts with identifying and defining classes. After defining all possible classes, these classes are arranged in a hierarchy depending on subject categories and sub categories. At this level we have refined our ontology by rearranging classes into sub classes and super classes. After the development of ontology we have implemented it using SPARQL queries. Finally the results which we got after running each query are analyzed using F1 score. The whole process is shown in figure 2.

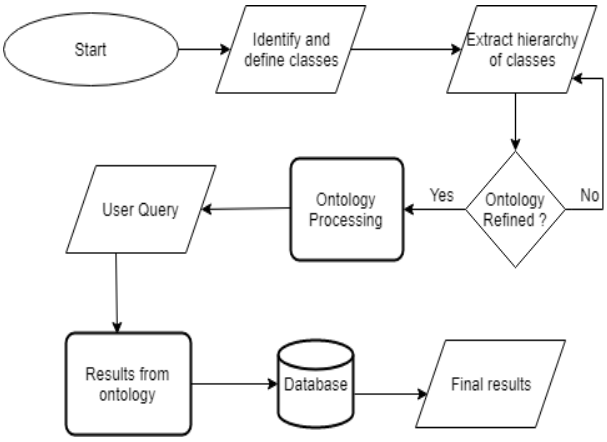

Figure-2: Process presentation of semantic search

RESULTS AND DISCUSSION

The results that we get from both semantic and syntactic searches are evaluated and compared using precision, recall and F1 score. These evaluation techniques are widely used in information retrieval systems as reported in most of the relevant literature like (Jan *et al.*, 2012). In this paper, we consider Precision is the ratio of correctly retrieved book titles to the total predicted positive titles. Recall is the ratio of correctly predicted positive book titles to all the titles in actual class. F1 score is average of precision and recall. Precision tells us how accurate our results are, recall tells us how good this model is for our search while the F1 score takes both false positives and false negatives into account. Naturally it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if we have an uneven class distribution.

To properly evaluate the efficiency of our proposed mechanism, we designed three groups i-e Group 1, Group 2, Group 3 and Group 4.

- **Group-1:** It represents the results for the combination of super class and sibling classes. Simply, the search mechanism considers and retrieves titles which match the super and sibling classes.
- **Group-2:** It represents the results for the sub class. Simply, the search mechanism considers and retrieves titles which match the sub class.
- **Group-3:** It represents the results for the super class. Simply, the search mechanisms consider and retrieve titles which match the super class.
- **Group-4:** It represent the results for the combination of super class, sub class and sibling classes. Simply, the search mechanisms consider and retrieve titles which match the super, sub and sibling classes.

We run a related query for each group for all the 75 randomly chosen keywords. For each keyword results we calculated precision and recall. After calculation of precision and recall, we calculated average precision and average recall for each group. From these average precision and recall we calculated F1 score which is actually required for evaluation of results.

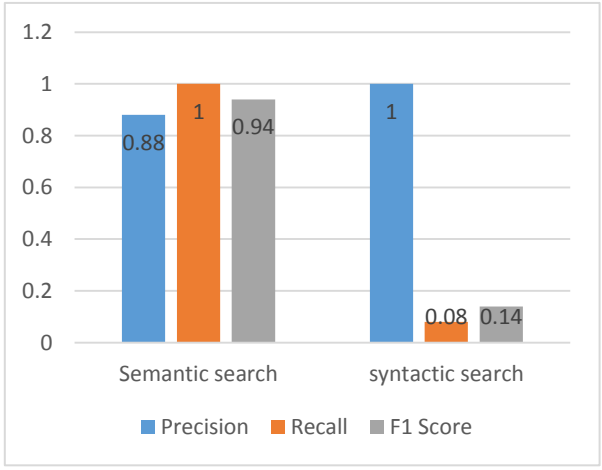

Figure-3: Semantic Search for Group 1

Figure 3 shows that in case of semantic search average precision was 0.88, meaning that 88% of relevant book categories has been identified. Only 12% false positive values has been identified. In this case recall is 1 that is all the relevant book categories has been identified. There was no relevant book category which remain unidentified. F1 score for this query is 0.94. This means that 94% of accuracy we get after running test for group 1.

Figure 3 also shows that in case of syntactic search average precision is 1 while that of recall and F1 score are 0.08 and 0.14 respectively.

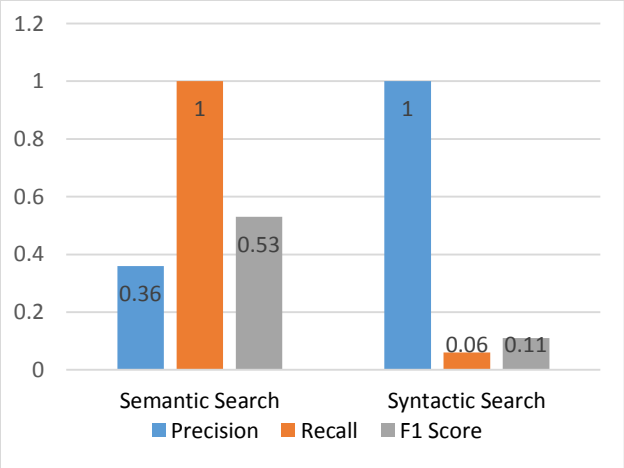

Figure-4: Semantic search for Group 2

For this query when comparison is made between semantic search and syntactic search, it is concluded that in this case semantic search is much better than syntactic search. Semantic search has F1 score is 0.94 and syntactic search has F1 score 0.14. There is huge difference between relevancy of results, means that in case of semantic search query results in low false positives and low false negatives as compare to syntactic search

Figure 4 shows that in case of semantic search average precision was 0.36, means that 36% of relevant book categories has been identified. Only 67% false positive values have been identified. In this case, recall is 1 that is all the relevant book categories have been identified. There was no relevant book category which remained unidentified. F1 score for this query is 0.53. This means that 52% of accuracy we get after running test for group 2.

Figure 4 also shows that in case of syntactic search average precision is 1 while that of recall is 0.06 and F1 score is 0.11.

When comparison is made between semantic search and syntactic search, it is concluded that in this case semantic search is much better than syntactic search. Semantic search has F1 score is 0.53 and syntactic search has F1 score 0.11, means that in case of semantic search query results in low false positives and low false negatives as compare to syntactic search.

Figure 5 shows that in case of semantic search average precision is 0.52, means that 52% of relevant book categories have been identified. Only 48% false positive values have been identified. In this case, recall is 1 that is all the relevant book categories have been identified. There was no relevant book category which remained unidentified. F1 score for this query is 0.68. This means that 67% of accuracy we get after running test for group 3(Semantic search with super class). When we see syntactic results, precision is 1, recall is 0.95 and F1 score is 0.97 as shown in figure 5.

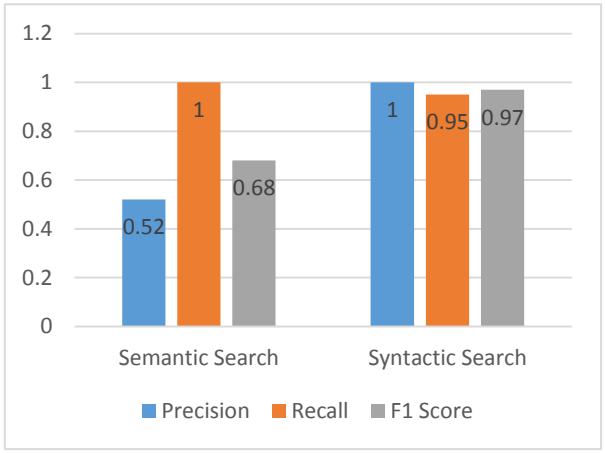

Figure-5: Semantic Search for Group 3

When comparison is made between semantic search and syntactic search, it is concluded that in this case syntactic search is better than semantic search. Semantic search has F1 score is 0.52 and syntactic search has F1 score 0.97, means that in case of syntactic search query results in low false positives and low false negatives as compare to syntactic search.

In this case, F1 score for syntactic search is better than semantic search because in case of syntactic search every keyword identified accurately for all tested super classes.

Figure 6 shows that in case of semantic search average precision was 0.87, means that 87% of relevant book categories have been identified. Only 13% false positive values have been identified. In this case, recall is 1 that is all the relevant book categories have been identified. There was no relevant book category which remained unidentified. F1 score for this query is 0.93. This means that 93% of accuracy we get after running test for group 4. Figure 6 also shows that in case of syntactic search average precision is 1 while that of recall and F1 score are 0.07 and 0.13 respectively.

For this query when comparison is made between semantic search and syntactic search it is concluded that in this case semantic search is much better than syntactic search. Semantic search has F1 score is 0.93 and syntactic search has F1 score 0.13. There is huge difference between relevancy of results, means that in case of semantic search query results in low false positives and low false negatives as compare to syntactic search.

From the all above discussion, it is concluded that Group 1 got the best semantic results i-e having F1 Score 0.94 meaning that 94% accuracy we got in this

case. The worst results are fetched in case of Group 2 where only 53% of accuracy is attained.

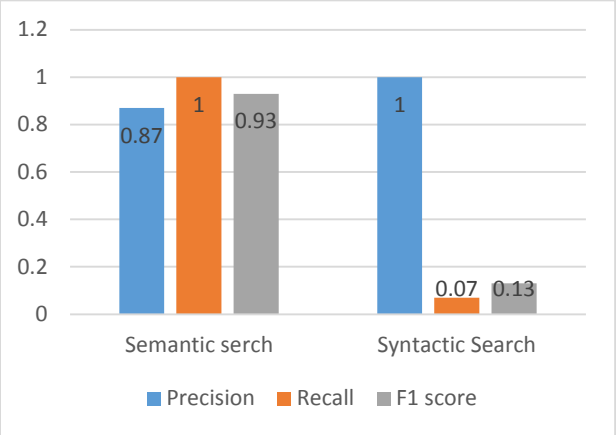

Figure-6: Semantic Search for Group 4

Conclusion: Semantic search implemented here is an efficient and accurate way to search books from library catalogue. This technique is much more different and unique than ordinary searching techniques. The project is successful as it gave us much more significant results for semantic search (when average is taken) than common library application softwares. In syntactic search we got only one exact matched result while in case of semantic search all possible related book categories have been identified and retrieved. Syntactic search is a matching search query based upon the actual words (Keywords) the searcher typed into the engine. This search would be exact and phrase matched. Semantic search is matching search queries based upon the intent of what the searcher typed into the engine. This is broad match and gives us all possible relevant results.

For Group 3, some contradictory results are fetched that is F1 score of syntactic search is greater than that of semantic search. As this happened for only one query so we can say that overall more relevant results are retrieved for semantic search and hence overall we get more efficient results for semantic search.

In future, various more attributes are added to the ontology and more combination of queries can be applied. Similarly, this ontology can be refined by working on properties and relations among different classes. Some developments can be done by adding some more book categories and then refine it accordingly so that it can be used for a book library on large scale. There are still a lot of things to study and explore.

In future, this ontology can be integrated with the help of the OWL API in Net Beans/Eclipse where a user will be able to query the new friendly interface with accurate and efficient semantic search.

REFERENCES

Candela, G., P. Escobar, R.C. Carrasco and M. Marco-Such (2018). Migration of a library catalogue into RDA linked open data. Semantic Web, 9(4), 481-491.

Deliot, C. (2014). Publishing the British national bibliography as linked open data. Catalogue & Index 174, 13–18.

Giri, K. (2011). Role of Ontology in Semantic Web. DESIDOC Journal of Library & Information Technology, 31(2), 116–120.

Imam, F.T., S. Larson, J.S. Grethe, A. Gupta, A. Bandrowski and M.E. Martone (2012). Development and use of ontologies inside the neuroscience information framework: a practical approach. Frontiers in genetics, 3, 111.

Iqbal, R., M.A.A. Murad, A. Mustapha and N.M. Sharef (2013). An Analysis of Ontology Engineering Methodologies: A Literature Review. Research Journal of Applied Sciences, Engineering and Technology, 6(16), 2993–3000.

Jan, S., M. Li, G. Al-Sultany, H. Al-Raweshidy and I.A. Shah (2011). Semantic File Annotation and Retrieval on Mobile Devices. Mobile Information Systems, 7(2), 107–122.

Jan, S., M. Li, H. Al-Raweshidy, A. Mousavi and M. Qi (2012). Dealing With Uncertain Entities in Ontology Alignment Using Rough Sets. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6), 1600–1612.

Kara, S., Ö. Alan, O. Sabuncu, S. Akpınar, N.K. Cicekli and F. N. Alpaslan (2012). An ontology-based retrieval system using semantic indexing. Information Systems, 37(4), 294–305.

Sarasua, C., E. Simperl, N. Noy, A. Bernstein and J.M. Leimeister (2015). Crowdsourcing and the semantic web: A research manifesto.

Subramaniyaswamy, V., G. Manogaran, R. Logesh, V. Vijayakumar, N. Chilamkurti, D. Malathi and N. Senthilselvan (2019). An ontology-driven personalized food recommendation in IoT-based healthcare system. The Journal of Supercomputing, 75(6), 3184-3216.

Tran, T., P. Cimiano, S. Rudolph and R. Studer (2007). Ontology-Based Interpretation of Keywords for Semantic Search. The Semantic Web Lecture Notes in Computer Science, 523–536.

Zhang, L. and J. Li (2011). Automatic generation of ontology based on database. Journal of Computational Information Systems, 7(4), 1148-1154.