Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 11, November 2014, pg.325 – 330

RESEARCH ARTICLE

ONTOLOGY BASED WEB CRAWLER FOR MINING SERVICES INFORMATION DISCOVERY

S.SASIREGA, Mrs. A.Jeyachristy
PG STUDENT, ASSISTANT PROFESSOR
NPR COLLEGE OF ENGINEERING AND TECHNOLOGY, TAMIL NADU, INDIA
EMAIL: sasi.rega25991@gmail.com

***Abstract:** The Mining process is based on ontology but the problem on ontology is based on mining process using crawler it cannot mine user queries without keywords in meta directory. To resolve this issue, to propose two kind of techniques used for information discovery for user query. First Ontology based web crawler algorithm and Combined Tag and Value Similarity for Data Extraction algorithm. The Ontology based web crawler algorithm used to mining user queries by using separate service. The use of algorithm like page rank and other importance-metrics have scheduled a new approach in prioritizing the URL queue for downloading higher relevant pages. The Combined Tag and value similarity for Data Extraction technique is easily identify heterogeneous databases the user criteria will satisfy with the help of meta directory. The large amount of available information on the Web makes it hard for users to locate resources about particular topics of interest. Traditional search tools, e.g., search engines, do not always successfully scope with this problem, which is, helping users to seek the right information. In the personalized search domain, focused crawlers are receiving increasing attention, as a well-founded alternative to search the Web. Unlike a standard crawler, which traverses the Web downloading all the documents it comes across, a focused crawler is developed to retrieve documents related to a given topic of interest, reducing the network and computational resources. To presents an overview of the focused crawling domain and, in particular, of the approaches that include a sort of adaptivity. That feature makes it possible to change the system behavior according to the particular environment and its relationships with the given input parameters during the search.*

Index Terms— Mining process service, web crawler, meta directory, ontology directory, service information discovery

1. INTRODUCTION

An Ontology is a Specification of a Conceptualization. An Ontology provide a shared vocabulary, which can be used to model a domain, that is, the type of objects and/or concepts that exist and their properties and relations. A conceptualization can be defined as an intensional semantic structure that encodes implicit knowledge constraining the structure of a piece of a domain. Ontology is a (partial) specification of this structure, it is usually a logical theory that expresses the conceptualization explicitly in some language. Conceptualization is language independent, while ontology is language dependent.

Ontology is important for the purpose of enabling knowledge sharing and reuse. An ontology is in this context a specification used for making ontological commitments. Practically, an ontological commitment is an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent (but not complete) with respect to the theory specified by an ontology. Agents then commit to ontologies and ontologies are designed so that the knowledge can be shared among these agents. After an ontology is developed, it is used, reused, and related to other ontologies, and also needs to be maintained. These tasks may be easier when an ontology is designed with these tasks in mind. For example, building ontology on an shared upper ontology and using a modular design usually means easier use and maintenance. In this chapter we describe operations on ontologies, relations between ontologies, and a classification of ontologies.

It is possible that one application uses multiple ontologies, especially when using modular design of ontologies or when we need to integrate with systems that use other ontologies. In this case, some operations on ontologies may be needed in order to work with all of them. We will summarize some of these operations. The terminology in this areas is still not stable and different authors may use these terms in a bit shifted meaning, and so the terms may overlap, however, all of these operations are important for maintenance and integration of ontologies. Merge of ontologies means creation of a new ontology by linking up the existing ones.

Alignment is a process of mapping between ontologies in both directions whereas it is possible to modify original ontologies so that suitable translation exists (i.e., without losing information during mapping). Thus it is possible to add new concepts and relations to ontologies that would form suitable equivalents for mapping. The specification of alignment is called articulation. Alignment, as well as mapping, may be partial only.

Refinement is mapping from ontology A to another ontology B so that every concept of ontology A has equivalent in ontology B, however primitive concepts from ontology A may correspond to non-primitive (defined) concepts of ontology B. Refinement defines partial ordering of ontologies.

Unification is aligning all of the concepts and relations in ontologies so that inference in one ontology can be mapped to inference in other ontology and vice versa. Unification is usually made as refinement of ontologies in both directions.

Integration is a process of looking for the same parts of two different ontologies A and B while developing new ontology C that allows to translate between ontologies A and B and so allows interoperability between two systems where one uses ontology A and the other uses ontology B. The new ontology C can replace ontologies A and B or can be used as an interlingua for translation between these two ontologies. Depending on the differences between A and B, new ontology C may not be needed and only translation between A and B is the result of integration. In other words, depending on the number of changes between ontologies A and B during development of ontology C the level of integration can range from alignment to unification.

Inheritance means that ontology A inherits everything from ontology B. It inherits all concepts, relations and restrictions or axioms and there is no inconsistency introduced by additional knowledge contained in ontology A. This term is important for modular design of ontologies) where an upper ontology describes general knowledge and a lower application ontology adds knowledge needed only for the particular application. Inheritance defines partial ordering between ontologies.

The System is identifying user query result with the help of semantic web and ontology based. Exact Matching and generating result rank basis. This could be useful to every search process based data mining, crawling the valid information from data repository by using Meta Directory. It is definitely a platform they are comfortable with communicating on.

2. LITERATURE SURVEY

2.1 Self-Adaptive Semantic Focused Crawler for Mining Services Information Discovery

The service users may encounter three major issues – heterogeneity, ubiquity, and ambiguity, when searching for mining service information over the Internet. To present the framework of a novel self-adaptive semantic focused crawler – SASF crawler, with the purpose of precisely and efficiently discovering, formatting, and indexing mining service information over the Internet, by taking into account the three major issues. This framework incorporates the technologies of semantic focused crawling and ontology learning, in order to maintain the performance of this crawler, regardless of the variety in the Web environment. The innovations of this research lie in the design of an unsupervised framework for vocabulary-based ontology learning, and a hybrid algorithm for matching semantically relevant concepts and metadata. To design a mining service ontology and a mining service metadata schema to solve the problem of self-adaptive service information discovery for the mining service industry.

2.2 Focused crawling for automatic service discovery, annotation, and classification in industrial digital ecosystems

Digital Ecosystems make use of Service Factories for service entities' publishing, classification, and management. However, before the emergence of Digital Ecosystems, there existed ubiquitous and heterogeneous service information in the Business Ecosystems environment. Therefore, dealing with the preexisting service information becomes a crucial issue in Digital Ecosystems. In order to resolve this issue, to present a conceptual framework for a semantic focused crawler, with the purpose of automatically discovering, annotating, and classifying the service information with the Semantic Web technologies.

2.3 A framework for discovering and classifying ubiquitous services in digital health ecosystems

A digital ecosystem is a widespread type of ubiquitous computing environment comprised of ubiquitous, geographically dispersed, and heterogeneous species, technologies and services. As a sub-domain of the digital ecosystems, digital health ecosystems are crucial for the stability and sustainable development of the digital ecosystems. Since the service information in the digital health ecosystems exhibits the same features as those in the digital ecosystems, it is difficult for a service consumer to precisely and quickly retrieve a service provider for a given health service request. Consequently, it is a matter of urgency that a technology is developed to discover and classify the health service information obtained from the digital health ecosystems. A survey of state-of-the-art semantic service discovery technologies reveals that no significant research effort has been made in this area. Hence, To present a framework for discovering and classifying the vast amount of service information present in the digital health ecosystems. The framework incorporates the technology of semantic focused crawler and social classification. A series of experiments are conducted in order to respectively evaluate the framework and the employed mathematical model.

2.4 Semantic web services in factory automation: Fundamental insights and research roadmap

One of the significant challenges for current and future manufacturing systems is that of providing rapid reconfigurability in order to evolve and adapt to mass customization. This challenge is aggravated if new types of processes and components are introduced, as existing components are expected to interact with the novel entities but have no previous knowledge on how to collaborate. This statement not only applies to innovative processes and devices, but is also due to the impossibility to incorporate knowledge in a single device about all types of available system components. To proposes the use of Semantic Web Services in order to overcome this challenge. The use of ontologies and explicit semantics enable performing logical reasoning to infer sufficient knowledge on the classification of processes that machines offer, and on how to execute and compose those processes to carry out manufacturing orchestration autonomously.

3. CONCLUSION

The Ontology based Web Crawler finding exact mining in information. All the Available service are validated before the identification. It provides the rank result. It is automatically extracting QRRs from a query result page. CTVS employs two steps for this task. The first step identifies and segments the QRRs. We improve on existing techniques by allowing the QRRs in a data region to be noncontiguous. The second step aligns the data using re-ranking method. This employs semantic similarity to improve the quality of search results. We fetch the top N - results returned by search engine, and use semantic similarities between the candidate and the query to re-rank the results. The crawler, for service information discovery in the mining service process, by taking into account the heterogeneous, ubiquitous and ambiguous nature of mining service information available over the Internet.

REFERENCES

[1] H.Dong, and F.K.Hussain, "Self-Adaptive Semantic Focused Crawler for Mining Services Information Discovery," IEEE Trans. Ind. info., vol. 10, no. 2, may 2014.

[2] H.Dong, and F.K.Hussain, "Focused crawling for automatic service discovery, annotation, and classification in industrial digital ecosystems," IEEE Trans. Ind. Electron., vol. 58, no. 6, pp. 2106–2116, Jun. 2011.

[3] H. Dong, F. K. Hussain, and E. Chang, "A framework for discovering and classifying ubiquitous services in digital health ecosystems," J. Comput. Syst. Sci., vol. 77, pp. 687–704, 2011.

[4] J. L. M. Lastra and M. Delamer, "Semantic web services in factory automation: Fundamental insights and research roadmap," IEEE Trans. Ind. Informat., vol. 2, no. 1, pp. 1–11, Feb. 2006.

