Ibn Al Haitham Journal for Pure and Applied Science

Journal homepage: http://jih.uobaghdad.edu.iq/index.php/j/index

Surveying the Organization of Digital Contents in the Internet Environments Using Ontological Approaches

Geehan Sabah Hassan

College of Education for Women University of Baghdad

gee.fah@gmail.com

Article history: Received in 14 January 2019, Accepted in: 28 May 2019, Publish September 2019

Doi:10.30526/32.3.2294

Abstract

Ontology is a system for classifying human knowledge according to its objective characteristics and hierarchical relations through building clusters or that bear common characteristics. In digital environments, it is a mechanism that helps regulate a vast amount of information by achieving a complete link between sub-thematic concepts and their main assets. The purpose of this study is to survey the previously conducted studies that use ontology in organizing digital data on social networking sites, such as the search engines Yahoo, Google, and social networks as Facebook and their findings. Results have shown that all these studies invest ontology for the purpose of organizing digital content data, especially on the Internet sites.

Keywords: Ontology Approaches, Digital Contents, Online Searches, Sentiment Analysis, Naïve Bayes Classifier, Twitter.

1. Introduction

There is no doubt that the tremendous development in the field of digital publishing has contributed to the emergence of a huge amount of information available on the Internet. Today, most search engines offer the possibility for internet users easily to search and retrieve information without the need to be professional in dealing with these search engines. The real problem today does not lie in the ability of users to access this information quickly, but in the very long time required in the processes of viewing, filtering and in choosing the appropriate application. Moreover, today there is also a real desire for Internet users and owners alike to create more flexible methods by developing sites that allow users to virtually build relationships and friendships. Such a thing can be met based on common personal characteristics, as well as the continual desire of academic and academic institutions to employ internet and its various applications in the field of education.

Thus, there is a need to know about the best tools and means by which digital knowledge can be organized in a way that simulates the process of organizing and categorizing books on the shelves of libraries based on different criteria. In digital environments, there is also a need to enable search engines to manage, organize, and classify the digital contents of sites, based

on the substantive convergence of knowledge and personal characteristics of humans. Today, there is an attempt to survey Ontology, as a concept, and its role in organizing digital contents on the Internet through examining the previously conducted studies and their arrived at findings.

2. Concept of Ontology

Ontology is defined as an unequivocal formal determination of normal concepts [1]. It is a branch of theory that investigates the sorts and structures of items, properties, occasions, and procedures of relations in each region of reality. Portrayals of ideas and connections can exist as a specialist or a group of operators. This definition can be predictable when utilizing Ontology in the arrangement of definitions. Many studies have been invested Ontology to provide solutions in both social and common sciences. In addition, numerous kinds of Ontology editors (e.g., Protégé-OWL) can be utilized to construct ontology. Hence, the productive utilization of ontology editors amid improvement is essential while performing ontology representation [2,3].

3. Results of Ontology in the Organization of Digital Content

The Internet environment can be seen as an ideal environment for applying the concepts of ontology. This is because a vast amount of information is available through it to the extent that this amount outperforms the sources of information available in the world's largest library. Another important issue is that the digital contents on the internet involve a variety of texts, images, audio, and video recordings. They further facilitate accessing multimedia applications that combine all these forms into one thematic template. Therefore, the digital environments can accommodate a variety of ontology applications. Perhaps, the most prominent environment among them is the research guide [Yahoo], which enriches the concept of [Ontology] by grouping similar things in blocks of objectivity.

Ontology has been highly recommended for the purpose of organizing the digital contents on the Internet and the classification of web pages objectively to narrow the search circle within the subject limits. This is because it ensures a faster and higher accuracy in retrieving results. Here, one needs to shed light on the mechanism that the search engine [Yahoo] is different from that of [Google] where the former depends on the logic of the subsequent link between the substantive terms. This means that the responsibility of linking the subjects together depends on the user himself/herself. In other words, the construction of the substantive cluster depends on the search strategy implemented by the beneficiary. For example, querying the term 'varies' in the search engine [Google] goes directly to the computational algorithm unless the user links the term to a particular disease. In [Yahoo], health can be defined as a general classification of a subject; accordingly, the search for that term itself will result in the retrieval of its property in connection with trauma only. This is due to an earlier link between the thematic transferences and their general classifications. Besides, the search engine is limited to the field of health only and neglects the rest. **Figure 1.** shows the diagram of working mechanism of the Yahoo engine.

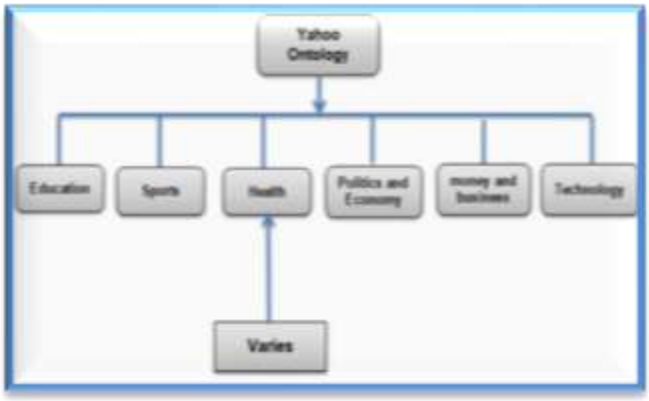

Figure 1. Mechanism Work of Yahoo Engine.

4. Ontology in Social Networking Sites

One of the important applications of [Ontology] in social networks is [Facebook]; it invests its principles in building groups based on the common interests between the owners of the accounts on the sites of those networks and the categories of the participants. Such a group is usually done according to the characteristics, attributes and features of the process of collection. There is no doubt that ontology helps control a large number of pages of subscribers in Facebook, for example, and maintains the method of organization to broadcast and circulate the sites despite being magnitude [4]. If one ponder upon the way friends are being nominated for you on Facebook, one will find out that there is a relationship of some kind that actually connects between or among people. It is very simple and complex at the same time. A case in point is that every subscriber in Facebook can identify himself/herself with multiple features, starting from the name and ending with identities and interests. Therefore, the philosophy of social networking sites in building relationships between participants is based on the principle of common characteristics. It is the principle of social relations in natural environment as well. This is because relations usually arise from the existence of a relative relationship with respect to education, occupation, hobby, interest, etc. This paper is to review previous studies on the application of ontology in social networking sites: In [5]. An ontology-based clustering of websites for grouping papers or proposals, and assigning that grouped proposal to reviewers systematically was examined. It was shown that ontology facilitates text-mining and some text extraction techniques to cluster based on similarities. It further helps assign these papers or proposals to reviewers and obtain text summarization. A principled approach is proposed to develop an intelligent information system by analyzing the unstructured repair verbatim data. Consequently, fault diagnosis ontology was constructed to find out the fault, and extract the irrelevant information. The three text mining techniques can be used for creating an intelligent information system. The proposed method is to analyze text summarizations and plagiarism analysis for the purpose of finding out the efficiency i.e. time complexity and increasing the performance of the system using a cluster-based approach. As for [6]. An approach was made towards semantic web mining data, where Semantic Web includes efforts to build a new structure World Wide Web (WWW) that enhances contents with official connotations. This will enable automated agents to reason about web contents, and perform smarter tasks on behalf of users with a focus on clustering objects described by the metadata-based ontology. Such a method has been empirically evaluated on the basis of World Facebook which was later easily converted into

an over-the-top meta-information. Objects are described by ontology-based metadata that act as inputs to measure similarity. Such an approach is based on similarities using the ontology structure of an instance and the metadata structure being installed. Objects require measuring the similarity that is calculated between objects. This study used three types of similarities: Taxonomy similarity, relation similarity and attribute similarity. The results showed that the clustering method was able to detect known groups from countries, such as Scandinavia or the Middle East. In [7]. Another approach was proposed in view of semantics to evaluate information retrieval systems. The objective was to increment specific inquiry apparatuses, and to enhance how to surf the Internet using Google, Bing, and Yahoo. Such a study was an attempt to provide an effective way to browse web contents. This kind of tool does not consider the semantic-driven question terms and documented words. Furthermore, measuring the commitment of considering semantics in the request of results was limited to the search engines. In this paper, the specialists demonstrated outcomes as indicated by the following two unique techniques: First, the default arranges data as recommended by the search engines which are called "classical rankings", whereas the second request made by the framework proposition of this investigation is booking, which is ranked by the ontology- based approach, "semantic ranking". This test is planned to gauge client's fulfillment by comparing the same set of queries in both types of result rankings. The proposed approach test for the assessment of web search tools has ended up being appropriate to genuine hunt instruments. The outcomes demonstrated that semantic assessment is a promising method to enhance the execution and conduct search engines to finally obtain significant outcomes. In [8]. a focus was dedicated to the use of ontology in the search of web documents to capture all concepts or terms necessary for the scope. Knowing the scope, and the terms of the search can make the search accurate and effective on the Web. Accordingly, such a ontology- based technique was invested by the researchers of this study for the purpose of data detection. Ontologies is predicted to play an important role in the Semantic Web in the future. This is because it helps organize information about ranges to facilitate not only the retrieval of information about a specific domain, but also the logic and derivation of new information that can be used by search engines and agents programs. What made scientists use ontology is their ability to capture important terms and their hierarchical relationships to a particular domain, which encourages the use of this information to detect Web documents that fall within the same domain. This is because the results of this work were distinguished. This study aims to reveal the domain and annotation of HTML web documents and improves the search engine on the Internet. The basic idea is that each concept is represented as a class. More specialized concepts or classes are derived from children or from more general concepts or classes to form a hierarchical structure. Thus, ontology is a repository of all possible terms; a collection of terminologies can be a vital resource for the application when detecting a document scope of the web, commentary or research. It will be more useful if one can process ontology to derive a chart table that stores all possible terms from the corresponding field with some additional information for each term. With the development of Web 2.0 and its related advancements (social networks, blogs, wikis etc.), the current blast in the use of small-scale blogging administrations, and especially Twitter, has moved the focus to investigate smaller scale blogging posts and tweets. There exist different machine learning-based methodologies that perform opinion investigation on tweets, with the downside that they regard each tweet as one uniform proclamation, and dole out a slant score to the post all in all. Where in [9]. the

researchers propose the deployment of ontology-based strategies to decide the subjects examined in tweets, and separate each tweet in terms of viewpoints significant to the subject. The outcome is the task of an assessment score to each unmistakable perspective. A pattern situation gives that arrangements with the domain of a prevalent item (cell phones) and assesses the unmistakable highlights of each model arrangement. In [10]. a new model was proposed for the local mining of opinions based on common sense information extracted from the Contextual Ontology. In recent years, research on emotion analysis of public information from social networking sites has increased dramatically. Data available in social networking sites is one of the most effective and accurate sources for determining the general sentiments of any product /service. The proposed methodology allows the interpretation and use of data extracted from the Twitter site to determine the general views. This paper includes a specific site, male and female specific and a certain popular concept of the product. All abstractions are used to calculate the result of a year and build an automated learning model that classifies the user's views as positive or negative. In [11]. a model based on an implicit (though crucial) realization of emerging semantics was adopted. This means that meaning is necessarily dependent on a community of agents. Inspired by social labeling mechanisms, the semantic social networks are represented in the form of a triangular diagram of personal associations, concepts and instances, and the expansion of the traditional concept of theology (concepts and situations) with the social dimension. In this paper, the researchers presented how lightweight concepts and social networks of people of this model were produced through simple graph conversions. These effects were presented based on two broad independent groups. By evaluating the results compared to the emerging Ontologies (as a result of a social-based cloning process based) against the results of the traditional method of ontological extraction on the basis of co-occurrence, the effectiveness of the proposed modern model was shown. The results prove that using ontology to extract knowledge from search engines is more likely to be accepted as accurate by the community itself. In this vein, some studies combined the possibilities of searching, browsing and visualization in the distributed information system, as in [12]. In this proposal, researchers describe a distributed information system for a proxy. The Association plays a central role in the system. It produces descriptive information for a local site, which is used as a guideline for distributing research, navigation, and visualization. Therefore, high-quality metadata leads to a more useful information system. The mapping approach also introduced an ontology-based site, which uses space carriers to classify texts. The evaluation experiments proved the effectiveness of the classification approach for this proposal. The classification was also more accurate compared to the previous studies that used Naïve-Bayes implementation from Carnegie-Mellon University (McCallum, 1998), and found that the method proposed was more accurate.

5. Conclusion

In this study, a review of the application of ontology and its role in organizing digital contents in the Internet environment, especially the social networking sites was presented. The study further displayed the previous studies that dealt with this subject in question and their findings. It was explained how to use ontology in the organization of texts on the Internet sites, and how each study used anthology in a different way. However, all of these studies showed the superiority of ontology approaches when organizing the digital contents of data websites, such as Yahoo, Google, etc.

References

1. Gruber, T.R. A translation approach to portable ontology specifications. *Knowledge acquisition*.**1993**, *5*, *2*, 199-220.
2. Shoaib, M.; Yasin, M.N.; Hikmat, U.K.; Saeed, M.I.; Khiyal, M.S.H. Relational WordNet model for semantic search in Holy Quran, in *Emerging Technologies*, *ICET 2009. International Conference on*.**2009**, 29-34.
3. Saad, S.; Salim, N.; Zainal, H.; Muda, Z. A process for building domain ontology: an experience in developing solat ontology. *Electrical Engineering and Informatics International Conference*.**2011**, 1-5.
4. Tserpes, K.; Papadakis, G.; Kardara, M.; Papaoikonomou, A.; Aisopos, F.; Sardis, E.; Varvarigou, T. Ontology for Social Networking Sites Interoperability. Proceedings of the International Conference on Knowledge Engineering and Ontology Development. **2012**, 245-250.
5. Komal Rajput, Narendra Kandoi. An ontology-based text-mining method to develop an intelligent information system using cluster based approach. *International Conference on Inventive Systems and Control* (ICISC),**2017**.
6. Maedche, A.; Zacharias, V. Clustering ontology-based metadata in the semantic web. *European Conference on Principles of Data*– Springer.**2002**, 348-360.
7. Bouramoul, A.; Kholladi, M.K.; Doan, B.L. An ontology-based approach for semantics ranking of the web search engine results. In Multimedia Computing and Systems *(ICMCS) 2012 International Conference on IEEE*.**2012**, 797-802.
8. Bin Husain, J. *An ontology-based application to detect, annotate and search web documents Ph.D*. dissertation, Concordia University, **2004**.
9. Kontopoulos, E.; Berberidis, C.; Dergiades, T.; Bassiliades, N. Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*. **2013**, *40*, *10*, 4065-4074.
10. Mika, P. Ontologies are us A unified model of social networks and semantics. *Web semantics: science, services and agents on the World Wide Web*.**2007**, *5*, *1*, 5-15.
11. Zhu, X.; Gauch, S.; Gerhard, L.; Kral, N.; Pretschner, A. Ontology-based web site mapping for information exploration. In *Proc. 8th Intl. Conf. on Information and Knowledge management (CIKM'99)*, **1999**.
12. Jain, A.; Jain, M. Location based Twitter Opinion Mining using Common-Sense Information. *Global Journal of Enterprise Information System*.**2017**, *9*, *2*, 28-32.