ISSN 2347 - 3983

Volume 8. No. 5, May 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter114852020.pdf

https://doi.org/10.30534/ijeter/2020/114852020

Seminal Paper on Genealogy by using Ontology

Samiksha Sanjay Patankar[1], Dr.A.J.Agrawal[2]

[1]M. Tech Student, Department of Computer Science, RCOEM, Nagpur, Maharashtra, India, samikshapatankar07@gmail.com

[2]Associate Professor, Department of Computer Science, RCOEM, Nagpur, Maharashtra, India, agrawalaa@rknec.edu

ABSTRACT

When one begins research on a topic from seminal papers unfamiliar to them. They face literature survey of problems. They spend considerable amount of time in searching for relevant papers. To reduce their searching time these technique is useful in finding those paper with given titles and keywords. During the literature survey, it is extremely difficult to catch such a research trend correctly because the survey coverage could be limited and the understanding of the topic could be insufficient. First technique is proposed who finds a set of seminal papers on a given subject. The following is a measure of how much paper is influenced by another paper. Then propose a strategy that builds a genealogy of seminal papers using a measure of influence and issue information. Ultimately with this broad approach with a large amount of real world literature data, Show the efficiency and effectiveness of our approach. Second technique is an ontology based search engine helps identify the most effective and useful result in a seminal paper. The result produced by ontology based search engines is based on the literal meaning of the word in the given sentence. It doesn't take the keyword in a given sentence; instead it takes the meaning of the keyword delivered. There are a number of techniques followed in using search ontology search engines. By using ontology search engine we find the seminal papers and identifying their year of conference.

Key words : seminal paper, genealogy, ontology, linear search, binary search.

1. INTRODUCTION

Search is one of the basic functions of Data Processing, used to retrieve specific data from a database. The search may apply to both fixed and random lists of elements. Depending on the type of data structure, a selected the appropriate search algorithm the types of search algorithms used to search an object from a collection of objects. The search algorithm includes both traditional search algorithm i.e. widely used often recommended search algorithms i.e. search high-quality techniques proposed by researchers recently. Although there are many search algorithms, we have them we discussed and compared some useful, practical one's high efficiency in subsequent phases. Therefore, if one grasps the research trend in the topic by a thorough literature survey, a researcher can well predict the future research direction, and thus can perform successful research. However, during the literature survey, it is extremely difficult to catch such a research trend correctly because the survey coverage could be limited and/or the understanding of the topic could be insufficient. For searching new papers researchers face many problems, so someone to start a research on a topic from seminal papers which is unfamiliar to them. When researchers start with a new topic, they face a literature survey problem for the topic. This requires they spend a considerable amount of time for searching for high-quality and relevant papers from conferences, journals, or scholar search engines. First, they may consider good survey papers as a starting point. However, due to their static nature, new research results cannot be covered in them frequently. Second, the researcher may consider browsing papers from premier conferences or journals. However, even with known premier conferences or journals, there are numerous papers to be examined. Also, the coverage of the browsing could be quite limited. Third, the researcher may consider employing search engines such as Cite Scholar, Scopus, Web of Science and Libra. These search engines are very useful in finding those papers with given titles and keywords. However, when one issues a keyword-based query for survey, they list a huge number of search results as candidate papers to be read. A literature survey is even more challenging when one starts with unfamiliar domains such as physics, psychology, chemistry, statistics, and biology for interdisciplinary studies. While conducting a literature survey, it is also important to understand a research trend in a certain topic. A research trend gives us insight in the topic, i.e. what kind of research has been done before and what the hot issues are at this time.

2. RELATED WORK

On constructing seminal paper genealogy paper proposed a formulated problem of constructing seminal paper genealogy and also defined its sub-problems of finding seminal papers and identifying their influence relationships and it also proposed a novel algorithm for finding seminal papers in the framework of random walk with restart, and discussed its performance issue. It addressed influence scores among papers and parent–child relationships and proposed a method for genealogy construction by using them. It verifies the effectiveness and efficiency of the proposed methods by conducting extensive experiments with a real-world academic literature [1].

Text Mining Based on Domain Ontology paper the three methods are used for text mining using ontology. The first method is text data mining or text knowledge discovery its main purpose is to extract the interesting, important patterns and knowledge from the unstructured text documents. In second method text clustering mainly uses the method of concept mapping and the method of semantic distance. The method of concept mapping firstly mapped the words to the concepts based on domain ontology, and then put up text mining based on these concepts, such as Concept Selection and Aggregation. In third method Text classification mainly used the method of class definition. This method drew some standardized concepts from the vocabulary, defined the hierarchy of the concepts as the ontology to build the ontology, and made the domain ontology as the class of classification. This process included establishing feature vectors for each category and calculating the comparability of feature vectors of text document and categories to determine the specific classification [2].

Large Scale Text Classification using Map Reduce and Naive Bayes Algorithm for Domain Specified Ontology Building. Internet provides a lot of data. Many data type in Internet is unstructured, such as image, video, text, and sound. Text classification can be used to organize large unstructured data on Internet into structured data in specified domain. This provides a robust system and good accuracy for classifying document into specified domain as pre-processing phase for domain specified ontology building. In this paper two methods are used. First method is pre-processing data. In this method content extraction, tokenization, stop word removal, pre-processing phase is used. Second method presents about text classification. Text classification phase is divided into two phases. First phase is learning process to build some knowledge and second phase is document classification. Feature extraction is needed to help classifier to learn or classify document [3].

The Extension of Domain Ontology Based on Text Clustering. Text clustering is used to deal with a numerous unstructured text data sources. After clustering model converges, the data of multiple fields is obtained, and the text data of the domain we want is quickly obtained. The advantage is to remove dirty data from the data source, reducing the workload for the next natural language processing, and improving the ontology construction efficiency; secondly, the combination of text clustering and natural language processing combine the advantages of linguistic rules and statistical methods. This article wants to achieve an expected result: through the combination of text clustering and natural language processing, we propose this method to achieve higher precision and recall ratio after consistency detection [4].

An Ontology-Based Text-Mining Method to develop intelligent information system using cluster based approach. This paper has present a clustering on ontology based text mining for grouping paper or proposals and assigning that grouped proposal to reviewers systematically. It facilitates text-mining and some text extraction techniques to cluster approach based on their similarities and then to assign them to reviewer and obtain text summarization [5].

Tagging of Name Records for Genealogical Data Browsing. In this paper genealogical information taken from historical archives as easy to search as websites on the internet. It also makes finding information on certain people mentioned in the documents extremely easy and this tool being very useful for many genealogical researchers. It presented a method for annotating the short name records taken from summaries of documents. Also, followed a method developed for name cleaning and standardization, using Hidden Markov Models. Finally, designed a method for incorporating rules governing the labelling of terms which allow a user to be more expressive in annotation [6].

A Semantic Ontology-based Document Organizer to Cluster e-learning Documents. This paper introduced an ontology based approach for e-Learning documents clustering. The proposed method used term re-weighting and a clustering algorithm, namely two-step algorithm on papers set. First generated an ontology based on documents from a famous international e-Learning conference. Then in the clustering procedure, after stop words removal, stemming and merging synonyms and complex words unification, concept weights were calculated by taking into consideration the TF-IDF of the words and the corresponding weights of relations between each word and the other words in ontology. The term weights vectors for these documents were then clustered using the clustering algorithm [7].

A Novel Ontology Relation Graph-based Ontology Module Partition Algorithm In this paper author discussed ontology module partition method based on relation graph. The frame of ontology module partition method is studied first, and then the data structure of ontology relation graph is described. Thirdly, extended broad first traverse of ontology relation graph is proposed. The evaluation of the resulting modules is introduced at last. An ontology relation graph-based ontology module partition method is employed. Three main phases are included in the method. Reading and analysing ontology is achieved in the first phase, during which, subsumption, and equivalence between classes are identified, complex classes, including: intersection classes, union classes, complement classes, enumerated classes, disjoint classes are analysed. Property characteristics and restrictions are also identified respectively. All the identified information is recorded in an ontology relation graph. Secondly, according to the graph achieved in phase I, the algorithm of extended broad first search is done to every named class. Thirdly, evaluating is executed according to the evaluation criteria, through which partitioned ontology modules are evaluated [8].

Graph-based Partitioning of large-scale ontologies proposed method for partitioning large ontologies into smaller modules. When the weighted graph derived from the original large-scale ontology has been constructed, an efficient partitioning algorithm is utilized to divide the monolithic ontology into small modules. This paper takes the problem of ontology modularization as the problem of clustering of ontology entities, and proposes a novel two-phase partitioning approach Modularization of Large-scale Ontology. In the first phase construct the weighted graph of large ontology, and in the second phase cluster the weighted graph into several modules. The consequent experiments have shown that the Modularization of Large-scale Ontology is efficient. As the next step, to matching the modules of large-scale ontologies clustered by the Modularization of Large-scale Ontology to realize the matching large ontologies , which is a divide-and-conquer method [9].

Research of literature Information Retrieval Method based on Ontology is proposed A literature information retrieval model based on ontology. Mapping the index query entry into the ontology and examine they semantic links, Web text is represented as concept vector space model, and document similarity is calculated on the basis of ontology concept similarity. Based on this semantic similarity calculation method, a workable information retrieval method is put forward. This method can detect the similarity of the documents, which contain concepts with similar meaning but different forms, compared with the traditional method of keyword matching retrieval improved semantic retrieval precision [10].

3. PROPOSED WORK

The aim of the project is to propose such system that will work on the seminal paper genealogy by using the concept of ontology. In this, project make use of genealogy and will apply it on the papers to refer the year of publication, topic name and keyword .Ontology is used to understand the concept and relationship observed in seminal paper. An Ontology-based search engine helps identify the most effective and useful result in a seminal paper. The result produced by ontology - based search engines is based on the literal meaning of the word in the given sentence. It doesn't take the keyword in a given sentence; instead it takes the meaning of the keyword delivered. There are a number of techniques followed in using search ontology search engines. By using ontology search engine we find the seminal papers and identifying their year of conference.

3.1 Algorithm

Algorithm (algo) plays a very important role. The algo used for searching the input query and provide the output in less amount of time. There are two algorithms which are very famous. Linear search and Binary search.

3.1.1 Linear Search

A direct search can be a sequential search, which uses a loop to list, starting with the main item. It compares each item to the number of searches yes, and stop at the right amount or up the list was met. If the correct search value is not between the same array, the algorithm it will investigate successfully until the end of compilation. Since the items in the array are stored sequentially searching for an object in its own order makes it easier too efficient. Search may be successful or unsuccessful. The algo does not required the data to be stored in particular order.

3.1.2 Binary Search

The usual time for a good search for organized data is a binary path search. It works according to the following steps.

a. The first region to search is the entire list.
b. Look at the amount of data within the search region.
c. Once you have achieved your goal, stop.
d. If your target is below the average data value, a new search it circulates the lower part of the data.
e. If your target is larger than the average data value, new the search region is the top portion of the data.
f. Continue from step b.

3.2 Methodology

The figure 1 methodology consists of building blocks document annotation, preprocessing, grouping, genealogy and result analysis. The detail explanation is given as follows.

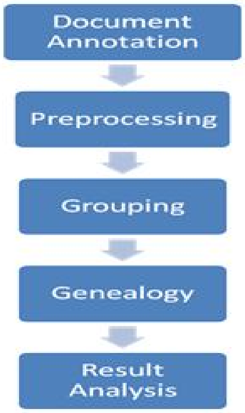

Figure 1: Methodology

3.2.1 Document Annotation

A comment attached to a particular section of a document. Many computer applications unable to enter annotation on text documents, spreadsheets, presentations and other objects. This is a particularly effective way to use computers in a workgroup environment to edit and review work. The creator of a document sends it to reviewers who then mark it up electronically with annotations and return it. The documents creator then reads the annotations and adjusts the document appropriately.

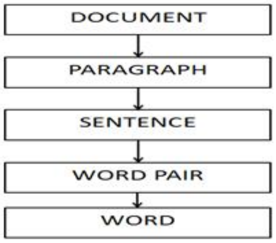

Figure 2: Structure of document annotation.

Figure 2 describe the segregation of document to the paragraph, paragraph to sentence, then to word pair and finally to the word as the document annotation done in the implementation of the paper.

3.2.2 Pre-processing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable formatting. Real -world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to certain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

Data preprocessing is used database-driven applications such as customer relationships management and rule-based applications. The phrase “garbage in, garbage out” is particular applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out -of -range values. e.g.: income -100, impossible data combinations. E.g. Sex: Male, pregnant: Yes, missing values, etc.

3.2.3 Grouping

Grouped data is a statistical term used in data analysis. A raw dataset can be organized by constructing a table showing the frequency distribution of the variable. Such a frequency table is often referred to as grouped data.

3.2.4 Genealogy

Genealogical research is a complex process that uses historical records and sometimes genetic analysis to demonstrate kinship. Readable conclusions are based on the quality of resources, ideally original records, the information within those sources, ideally primary or firsthand information, and evidence that can be drawn, directly or indirectly, from that information. In many instances, genealogists must skillfully assemble indirect or circumstantial evidence and conclusions, together with the documentation that supports them, is then assembled to create a cohesive genealogy or family history.

4. RESULT AND DISCUSSION

The various papers are uploaded based on title and Year wise. The data can be extended depending upon the user input. The database used for this project is SQL server. Figure 3 depicts the webpage of paper uploading section. Figure 4 shows the pie chart with percentage and total number of papers with year. After uploading total number of papers it very easy to analysis the overall scenario.

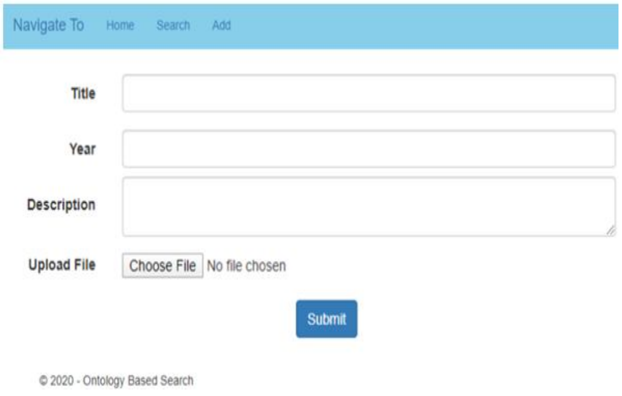

Figure 3: Paper upload section

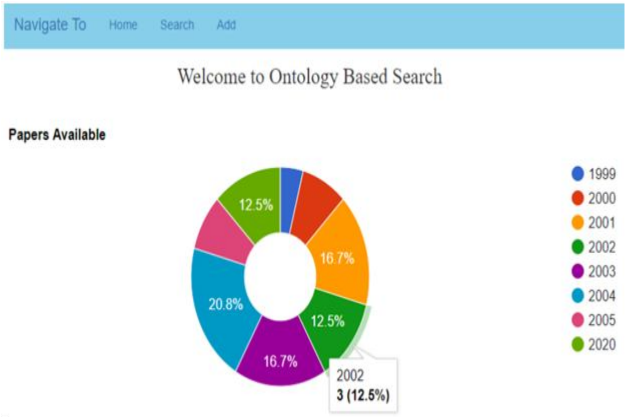

Figure 4: Pie chart for Year based Papers

5. CONCLUSION AND FUTURE SCOPE

The overall conclusion for seminal paper genealogy by using ontology is to save time as well manpower for researchers to find paper using ontology concept. This requires a considerable amount of time for searching for high-quality and relevant papers from conferences, journals, or scholar search engines. For this purpose we have proposed this report which helps basically in searching papers.

In this project, we proposed effective and efficient solutions to the problem of seminal paper genealogy. The contributions of the paper are summarized as follows:

- The problem of seminal paper genealogy and also defined its sub problems of finding seminal papers and identifying their year of conference.
- The effectiveness and efficiency of the proposed methods by conducting extensive experiments with a real-world academic literature.

We believe that our contributions would greatly help researchers who begin with their new topic:

- To capture the overview of the topic quickly only with a small number of seminal papers.
- To understand the research history and trend only with the genealogy.at will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

In future, the real-world situations, topics are influenced one another through cross-disciplinary research. Therefore, it is greatly useful because researchers normally have difficulty in finding seminal papers in other topics that could inspire their current research. Second, we are interest in the problem of survey papers. Sometimes, researches would not want to get survey papers as seminal papers in the area. So, it could be a good candidate for next research to identify the survey paper.

REFERENCES

1. Duck-Ho Bae, Se-Mi Hwang, Sang-Wook Kim, and Christos Faloutsos. **On Constructing Seminal Paper Genealogy**, *IEEE Transactions On Cybernetics,* Vol. 44, no. 1, pp. 54 - 65, January 2014. https://doi.org/10.1109/TCYB.2013.2246565
2. Feng Hu, and Yu-feng Zhang. **Text Mining Based on Domain Ontology**, *2010 International Conference on E-Business and E-Government,* pp.1456-1459, 2010.
3. Joan Santoso, Eko Mulyanto Yuniarno, and Mochamad Hariadi. **Large Scale Text Classification using Map Reduce and Naive Bayes Algorithm for Domain Specified Ontology Building**, *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Vol. 1, no. 1, pp.428-432, 2015. https://doi.org/10.1109/IHMSC.2015.24
4. Fuchao Liu, and Guanyu Li. **The Extension of Domain Ontology Based on Text Clustering**, *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Vol. 01, pp. 301-304, 2018.
5. Ms. Komal Rajput, and Mr. Narendra Kandoi. **An Ontology-Based Text-Mining Method to develop intelligent information system using cluster based approach**, *International Conference on Inventive Systems and Control (ICISC-2017),* 2017.
6. Mike Perrow, and David Barber. **Tagging of Name Records for Genealogical Data Browsing**, *6th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '06),* pp. 316-325, 2006. https://doi.org/10.1145/1141753.1141827
7. Sara Alaee, and Fattaneh Taghiyareh. **A Semantic Ontology-based Document Organizer to Cluster eLearning Documents**, *2016 Second International Conference on Web Research (ICWR),*2016. https://doi.org/10.1109/ICWR.2016.7498438
8. li Xiaohui, and Xie lie. **A novel ontology relation graph-based ontology module partition algorithm**, *2012 IEEE International Conference on Computer Science and Automation Engineering*, pp. 181-184, 2012. https://doi.org/10.1109/ICSESS.2012.6269435
9. Hongke Xia, Xuefeng Zheng, and Xiang Hu. **Graph-based Partitioning of large-scale ontologies**, *2010 2nd International Conference on Industrial and Information Systems,* Vol. 01, pp. 371-375, 2010. https://doi.org/10.1109/INDUSIS.2010.5565834
10. Song Yibing, and Ma Qinglong. **Research of literature Information Retrieval Method based on Ontology**, *2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI),* 2014.