

A Novel Approach for Information Retrieval Using CCBIR System

Dr. R. Malathi Ravindran

Associate Professor, Computer Applications, NGM College, Pollachi, Coimbatore, Tamil Nadu, , India

ABSTRACT

The tremendous growth of data generated by various resources like digital images, videos and scientific instruments & simulations is very high. The usage of social networks and internet are increasing day-by-day. The effective information retrieval from this huge volume of data is very much obligatory. The existing information retrieval like Boolean, Meta and Probabilistic are not sufficient to manipulate this huge volume of data. In this scenario, the entry of Cloud Technology is very much needed. In this paper a new cloud based information retrieval system is proposed with the inclusion of Vector Space Model and Semi-Supervised Clustering. KEGG Metabolic Relation Network and Dow Jones Industrial Index data set are experienced with the CCBIR system using math work. This system eliminates bottlenecks in information flow, time delay and also it provides a remarkable ability to overcome the traffic congestions.

Keywords : Cloud Computing; Data Set; Information Retrieval; Pathway.

I. INTRODUCTION

Information retrieval is a thrust area. Several methodologies are adopted for retrieving information. Information retrieval is a discipline involved with the organization, storage, retrieval and display of bibliographic information. Information Retrieval systems are designed with the objective of providing, in response to a user query, references to documents which would contain the information desired by the Information retrieval is the activity of user. obtaining information resources relevant to an information need from a collection of information resources. The existing information retrieval like Boolean, Meta and Probabilistic are not sufficient to manipulate this huge volume of data. [2]. A cloud technology is introduced to retrieve the information from the huge volume of data.[4]

II. CLOUD COMPUTING

Cloud Computing is a specialized form of distributed computing that introduces utilization models for remotely provisioning scalable and measured resources. It is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models [5].

A. Essential Characteristics of Cloud Computing

Cloud technology is in the news quite often these days, but it still seems to be mysterious and confusing to the non-technical crowd. Here are the five main characteristics that cloud computing offers businesses today [www.erpbloggers.com].

- On-demand Capabilities
- Broad Network Access
- Resource Pooling
- Rapid Elasticity
- Measured Service

B. Cloud Computing on Information Retrieval

Cloud computing encompasses a pay-per-use paradigm for providing services over the internet in a scalable manner. Supporting data intensive applications is an essential requirement for the clouds. However, dynamic and distributed nature of cloud computing environments makes data management processes very complicated, especially in the case of real-time data processing/database updating [1]. A new cloud based information retrieval system is proposed with the inclusion of Vector Space Model and Semi-Supervised Clustering. It has given a name called CCBIR(Cloud Computing Based Information Retrieval System) [3].

III. DATA SOURCE

A. KEGG Metabolic Relation Network

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances. In this work, KEGG Metabolic pathways dataset is used for analytics. It comes under the systems information. It is a collection of pathway maps integrating many entities including genes, proteins, RNAs, chemical compounds, glycans, and chemical reactions, as well as disease genes and drug targets, which are stored as individual entries in the other databases of KEGG.

This data set contains 24 attributes such as Pathway, Nodes, Edges, Connected, Components, Network Diameter, Network Radius, Shortest Path, Characteristic Path Length, Avg.num.Neighbours, Isolated Nodes, Number of Self Loops, Multi-edge Node Pair, NeighborhoodConnectivity, Outdegree, Stress, Self-Loops, PartnerOfMultiEdgedNodePairs, EdgeCount, BetweennessCentrality, Indegree, Eccentricity, ClosenessCentrality, AverageShortest PathLength and ClusteringCoefficient.

B. Dow Jones Index Data Set

On July 3, 1884, Charles Henry Dow began publishing his Dow Jones Average. After every eight months of its publication, the index was composed of 12 stocks, 10 of which were railroads. The first Dow Jones Industrial Average (DJIA) was published through September 29, 1916. Dow Jones Index dataset contains weekly data for the Dow Jones Industrial Index. It has been used in computational investing research. It contains 16 native attributes such as quarter, stock, date, open, high, low, close, volume, percent_change_price(PCP), percent_ chagne_volume_over_last_week(PCV_OLWeek), previous weeks volume, next_weeks _open(NW_open), next_weeks_close(NW_close), percent_change_next_weeks_price (PC _NW_price, days_to_next_dividend (DTN_dividend) percent_ return_next_dividend(PRN_dividend)

C. Mathworks in CCBIR System

The datasets KEGG Metabolic Relation Network and Dow Jones Index are experienced with Mathworks. Simulink Real Time and Statistics & machine learning are used to simulate with collected dataset. The results have been graphically demonstrated with these tools.

II. EXPERIMENTAL RESULTS

A. Experimental Analysis of KEGG Metabolic Relation Network

KEGG Pathway Metabolic Network dataset is experienced with Mathworks. Ten experimentations were taken for both datasets KEGG Metabolic Relation Network and Dow Jones Index Dataset before and after applying CCBIR system [5] Experimental Analysis of KEGG Metabolic Relation Network using Native mechanisms

The average experimental analysis is taken for the ten experiments using native mechanisms. The graph figure 1 illustrates the various aspects of metabolic network of average experimental analysis



Figure 1. Average Experimental Analysis for KEGG Metabolic Relation Network Using Native Mechanisms

Experimental Analysis of KEGG Metabolic Relation Network using CCBIR system

The average experimental analysis is taken for the ten experiments using CCBIR system. The graph figure 2 illustrates the various aspects of metabolic network of average experimental analysis.



Figure 2. Average Experimental Analysis for KEGG Metabolic Relation Network Using CCBIR System

Comparison of Average Experimental Analysis of KEGG Metabolic Relation Network before and after implementation of CCBIR System

The graph in Figure 3 illustrates the comparison of average experimental analysis of KEGG Metabolic Relation Network before and after the implementation of CCBIR System.[5]



Figure 3. Comparison of Experimental Analysis between CCBIR System and Native Mechanism

The graph shows Comparison of Average Experimental Analysis of KEGG Metabolic Relation Network before and after implementation of CCBIR System. It contains 24 attributes. It clearly depicts about after implementation of CCBIR system, some of the parameter values are reduced such as; closeness centrality, Self-Loops, shortest path, Isolated nodes, Clustering Coefficient, Eccentricity, Stress, Edges, Average Shortest Path, Outdegree, Number of self-loops, Characteristic Path Length, Nodes. Shortest path maintains the highest value, whereas isolated node in existing contains zero but after implemented CCBIR system it contains 1.87231E-05. Similarly Number of Self Loops and Self Loops contain zero value both after and before implementation of CCBIR System.

B. Experimental Analysis of Dow Jones Index Dataset

Dow Jones Index Dataset is experienced with Mathworks. Ten experimentations were taken before and after applying CCBIR system and clearly represented using graphs.





Figure 4 Average Experimental Analysis for Dow Jones Index Data Set using Native Mechanisms [Quarter: 2 -Stock: XOM - Date: 24/06/2011]



Average Experimental Analysis for Dow Jones Index Dataset using CCBIR System

Figure 5. Average Experimental Analysis for Dow Jones Index Data Set using CCBIR System [Quarter: 2 - Stock: XOM - Date: 24/06/2011]

Comparison of Average Experimental Analysis of Dow Jones Index Dataset before and after implementation of CCBIR System

The graph Figure 6 illustrates the comparison of average experimental analysis of Dow Jones Index dataset before and after implementation of CCBIR System.



Figure 6. Comparison of Experimental Analysis between CCBIR System and Native Mechanism

This graph shows the comparison of average experimental analysis of Dow Jones Index dataset before and after implementation of CCBIR System. Brown colour is CCBIR system result whereas Blue colour illustrates native mechanism result. This comparison clearly depicts that the proposed system provides better result when compared with native mechanism.

V. CONCLUSION

The major goal of the research work is to improve the information retrieval. In this work, intelligent technique has been proposed to achieve this goal. There are lots of methods developed by the researchers in information retrieval to improve retrieval performance. These methods include restricting the allowable indexing, training the intermediaries to generate terms from these restricted vocabularies and automatically clustering documents. In this work, CCBIR based system has been introduced to overcome these drawbacks and improves information retrieval.[5] From each datasets KEGG Metabolic Relation Network and Dow Jones Index datasets, 10 experiments were done and then average has been taken from those 10 experiments. Finally, after comparison of those average results, this research work proved that the current system called CCBIR system performs well.[5]

III. REFERENCES

- [1]. Amir H.Basirat and Asad I.Khan (2010), "Evolution of Information Retrieval in Cloud Computing by Redesigning Data Management Architecture from a Scalable Associative Computing Perspective", Springer Link, Neural Information Processing. Models and Applications Lecture Notes in Computer Science Vol. Part II, pp 275-282.
- [2]. Hari Priyanka R. and Malathi Ravindran R. (2016), "A Cloud Computing with Search Engines", International Journal for Scientific Research & Development, Vol.4, Issue 08, 2016, ISSN : 2321-0613.
- [3]. Malathi Ravindran R. and Antony Selvadoss Thanamani (2015), "K-Means Docoment Cluestering using Vector Space Model", Bonfring International Journal of Data Mining, Vol. 5, No.2.
- [4]. Riktesh Srivastava (2012), "Implementation of Information Retrieval (IR) Algorithm for Cloud Computing: A Comparative Study between

with and without MapReduce Mechanism", Journal of Contemporary Issues in Business Research, Vol.1, No.2, pp.42-56.

[5]. Thomas Erl, Zaigham Mahmood, and Ricardo Puttini (2013), "Cloud Computing Conepts, Technology & Architecture", Prentice Hall, ISBN-10: 0-13-338752-6.



Dr. R. Malathi Ravindran Associate Professor and Head, Research Department of Computer Science, NGM College, Pollachi, Coimbatore, India. She has 13 years of teaching

experience. Her area of interest includes data Mining. She has completed her Ph.D in Computer Science in the Research Department of Computer Science, NGM College, Pollachi under Bharathiar University, Coimbatore, India. She has published more than 10 papers in international/national journals and conferences. He has authored two books on recent trends in Information Retrieval.