International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471
Vol. 3 Issue 4, April-2014, pp: (4-12), Impact Factor: 1.147, Available online at: www.erpublications.com

Performance evaluation of ontology based Information Retrieval

V S Dhaka[1], Aditi Sharma[2]
[12]Jaipur National University, Jaipur, India

Abstract: Main objective of this paper is to achieve personalization of information retrieval activities. Personalised retrieval effectiveness is defined in terms of retrieving relevant documents and not retrieving non-relevant documents. Two traditional factors of measuring effectiveness are Recall and Precision are focused upon. Based on these two factors the effective performance of Ontology based retrieval is evaluated.

Keywords: Recall ratio, Ontology, deteriorated performance.

1. Information Retrieval Techniques

The IR systems rank documents based on estimation of the usefulness of a document for a user query. The technique that has been shown to be effective in improving document ranking is query modification via relevance feedback. A state-of-the-art ranking system uses an effective weighting scheme in combination with a good query expansion technique.

2. Relevance Feedback

In the early years of IR study, researchers realized that it was quite hard for users to formulate effective search requests. In 1965,Rocchio proposed using relevance feedback for query modification or reformulation. Relevance feedback is motivated by the fact that it is easy for users to judge documents as relevant or non-relevant for their query [2][3]. Relevance feedback in information retrieval in order to personalize information. Using such relevant judgments, a system can then automatically generate a better query for further searching. The basic idea consists of choosing important terms attached to certain previously retrieved items that have been identified as relevant by the user, and of enhancing the importance of these terms in the new query formulation. In general, the user is asked to judge the relevance of the top few documents retrieved by the system. Based on these judgments, the system modifies the query and issues the new query for finding more relevant documents from the collection. The classic experiments on relevance feedback were conducted and the results demonstrated that relevance feedback improves retrieval performance.

3. IR Measures

A number of studies have been conducted to measure the performance of the system. Some criteria of evaluation have been proposed by several researchers in the area of the evaluation of information retrieval systems. These criteria include coverage of the system, form of presentation of the search output, user effort, the response time of the system and recall & precision.

4. Recall & Precision

Recall indicates the ability of a system to present all relevant documents. In fact, it may not be possible to retrieve all the relevant documents from a large collection. A system may be able to retrieve only a proportion of the total relevant documents. Thus, the performance of a system is often measured by recall ratio, which denotes the percentage of relevant items retrieved in a given circumstances. It is measured in terms of Recall ratio.

$$\text{Recall} = \left\{ \frac{\text{Total number of relevant items in the collection}}{\text{Number of relevant items retrieved}} \right\} \times 100$$

Precision shows the ability of a system that present only relevant documents. It states the ability not to retrieve non-relevant documents. This factor demonstrates that how the system is able to keep away the unwanted documents in a given circumstances.

$$\text{Precision} = \left\{ \frac{\textit{Total number of items retrieved}}{\textit{Number of relevant items retrieved}} \right\} \times 100$$

5. Experimental Evaluation

Data Sets are used for evaluation of the proposed method. Two datasets **Generated Data Set** and **FIRE 2010 Data Set**, are used for evaluation of the proposed method.

5.1 Generated Data Set

The first data set is manually generated [see Appendix A] based on the Web that Google has indexed. We generated dataset by web interactions of 15 users, who used the Google search engine for 60 days, an average of three query topics per day from a collection of 60 query topics. The query topics contain an average query length 2.2. The queries utilized in our experiments were deliberately designed to be short after eliminating stop words to reflect the general trends in user search queries. The set of pre-defined query topics is accumulated from a variety of users with similar as well as non-similar backgrounds. Although query topic was formed physically however users were cautiously inquired from dissimilar background and having dissimilar context. In these experiments, users were asked to offer the relevance feedback without much interfering them. All the pertinent documents were processed and user profiles were produced.

5.2 FIRE 2010 Data Set

The second dataset used for evaluation of the proposed approach is FIRE 2010 dataset. In FIRE 2010 data set consists of a collection of 50 Query topics with description and narration (Appendix A). In this evaluation process, 15 users were asked to interact with search engine by undertaking our system .Since second data set has predefined context of query topics, so it is considered that all users had same context with each query topic. Some users posed few overlap query topics also and provided relevance feedback. These data sets are used throughout this thesis for all approaches.

5.3 Results and Analysis

The results obtained from search engine are compared with results obtained from the proposed method. The average precision and average recall measures are utilized to estimate the reclamation correctness performance of the proposed method. The meanings of these measures presume that, for a specified query, there are so many documents that is relevant and a set of documents that is not relevant. Mean Average Precision (MAP), Mean Average Recall (MAR) for each user in the data set are computed as different precision and recall values are obtained for the same query posted by different user context.

6. Generated Data Set Results

The Generated Data Set Results discussed as: Mean Average Precision results and Mean Average Recall results.

6.1 Mean Average Precision Results

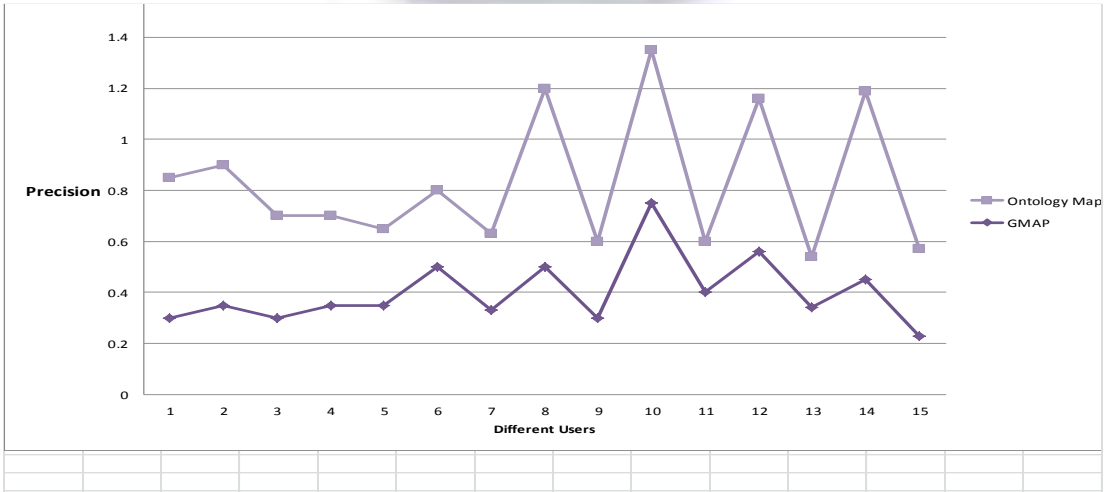

Figure 1: Mean Average Precision for Google and Ontology

6.2 Analysis of Generated data set Results

The users are classified on the basis of their performance. The Mean Average Precision performance comparison between the Google and the Ontology is calculated as

$$Performance = \frac{\sum_{i=1}^{15}(OMAP_i - GMAP_i)}{\sum_{i=1}^{15} GMAP_i} \times 100$$

Equation - 1

Where GMAP = Mean Average Precision for Google results; OMAP = Mean Average Precision for Ontology results.

Comparison	No. of user	Performance
OMAP>GMAP	7	73.72%
OMAP=GMAP	2	-
OMAP<GMAP	6	-28.84%

Table 1: Mean Average Precision comparison between Google and User Profile + Ontology

In our proposed approach using Generated Data set, 46% users have retrieved approximately 73.72% improved precision and the 13% users found no change in the precision compared to the search engine Google. While rest 40% users found the precision deteriorated by approximately 28%.

6.3 Mean Average Recall Results

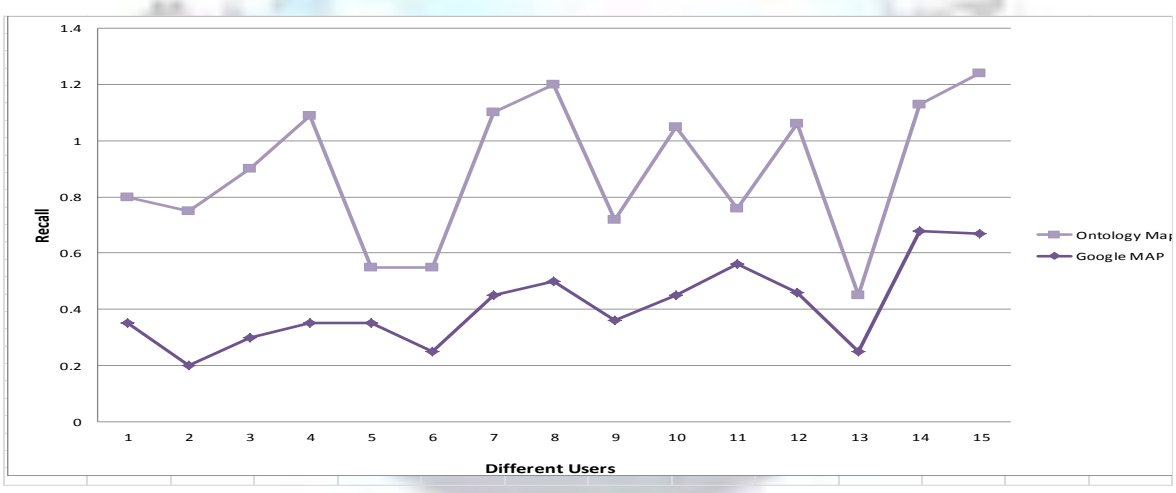

Figure 2: Mean Average Recall for Google and Ontology

6.4 Analysis of Generated data set Results

The Mean Average Recall performance comparison between the Google and the Ontology is calculated as Where

$$Performance = \frac{\sum_{i=1}^{15}(OMAR_i - GMAP_i)}{\sum_{i=1}^{15} GMAP_i} \times 100$$

GMAR= Mean Average Recall for Google results; OMAR=Mean Average Recall for Ontology results

Comparison	No. of user	Performace
OMAR>GMAR	9	56.79%
OMAR=GMAR	1	-
OMAR<GMAR	5	-29.96%

Table 2: Mean Average Recall comparison between Google and User Profile + Ontology

In our proposed approach using Generated Data set, 60% users have retrieved approximately 56.79% improved recall and the 7% users found no change in the recall compared to the search engine Google. While the rest 33% users found the recall deteriorated by approximately 29.96%.

7. Fire Data Set Results

The FIRE Data Set Results discussed as: MAP results and MAR results.

7.1 Mean Average Precision Results

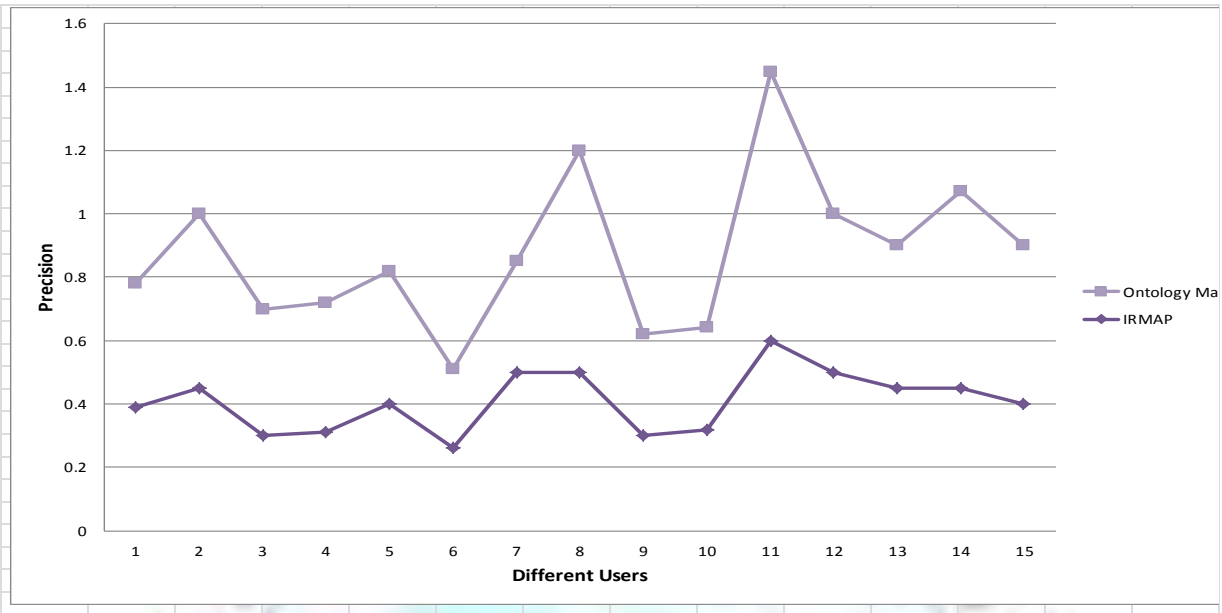

Figure 3: MAP for IR System and Ontology

7.2 Analysis of FIRE Data Set Results

The users are classified on the basis of their performance. The MAP performance comparison between the IR System and Ontology is calculated as

Where IRMAP = Mean average precision for IR system results; OMAP = Mean average precision for Ontology results

$$Performance = \frac{\sum_{i=1}^{15}(OMAP_i - IRMAP_i)}{\sum_{i=1}^{15} IRMAP_i} \times 100$$

Comparison	No. of user	Performace
OMAP>IRMAP	9	33.42%
OMAP=IRMAP	4	-
OMAP<IRMAP	2	-21.05%

Table 3: MAR comparison between Google and User Profile + Ontology

7.3 Mean Average Recall Results

Using FIRE Data set with Terrier (http://www.terrier.org/) as an underlying search engine to evaluate the proposed approach, 60% users have retrieved approximately 33.42% improved precision and the 26% users found no change in the recall compared to the search engine Terrier. While the rest 14% users found the recall deteriorated by approximately 21%.

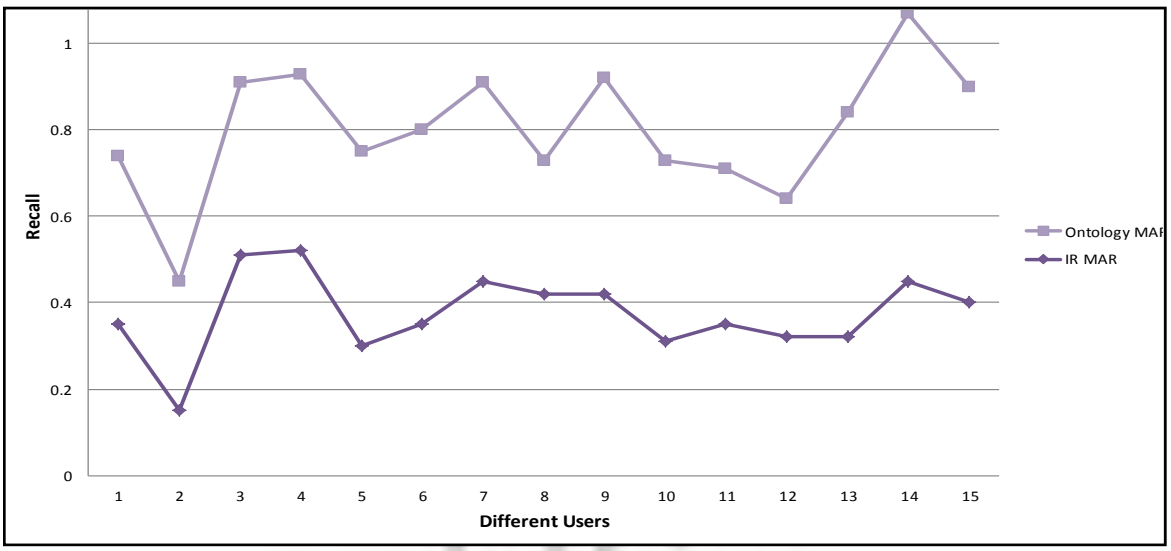

Figure 4: MAR for IR System and Ontology

7.4 Analysis of FIRE Data Set Results

MAR Results

The MAR performance comparison between the Google and the Ontology is calculated as

$$Performance = \frac{\sum_{i=1}^{15}(OMAP_i - IRMAR_i)}{\sum_{i=1}^{15} IRMAR_i} \times 100$$

Where IRMAR = Mean average recall for IR system results; OMAR = Mean average recall for Ontology results

Comparison	No. of users	Performance
OMAR>IRMAP	11	29.09%
OMAR=IRMAP	1	-
OMAR<IRMAP	3	-22.75%

Table 4: MAR comparison between IR System and User Profile + Ontology

FIRE Data set with Terrier (http://www.terrier.org/) as an underlying search engine to evaluate the proposed approach,73% users have retrieved approximately 29% improved recall and the 7% users found no change in the recall compared to the search engine Terrier. While the rest 20% users found the recall deteriorated by approximately 22.75%.

8. Comparison of Approaches

In table (a) comparison (%) of proposed approach with Google in table (b)

(a) Generated Data Set (proposed approach)		(b) Fire Data Set (Google)	
MAP	35.96	MAP	19.16
MAR	62.44	MAR	23.02
AVG Precision	10.34	AVG Precision	8.14
AVG Recall	4.18	AVG Recall	5.92

Table 5:(a) comparison(%)of proposed approach with Google in table (b)

9. Summary and Conclusions

The observations with the generated data set agree with FIRE data verifying our generated data set. The improved results are due to extracting close context of the user at that particular time and searching by expanding query for retrieving the documents buried in morass. A careful analysis of our results shows that users who posed Queries in a less popular context than well liked context got better performance. For example - Virus in context of micro-organism (health infection) is not as popular as computer associated context. In this scenario, the proposed approach performed better than the search engine. The possible reason of the deteriorated performance may be non-identification of the desired context in the WordNet ontology. WordNet vocabulary is limited and it does not cover special domain vocabulary and may resulting into the degradation of the performance of the proposed approach.

Results show the precision and recall of this Ontological User Profile approach is better over Google on an average.

References

[1]. Bhogal J, Macfarlane A and Smith P, "A review of ontology based query expansion", Science Direct. Information Processing and Management, vol. 43, 2007, pp. 866-886.
[2]. Buckley C and Voorhees EM, "Evaluating evaluation measure stability". 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2000, pp.33–40.
[3]. Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma," A survey of content-based image retrieval with high-level semantics", Pattern Recognition, Volume 40, Issue 1, January 2007, Pages 262-282
[4]. Todorovic, S.,'' Texel-based texture segmentation/"Computer Vision, 2009 IEEE 12th International Conference on Sept. 29 2009-Oct. 2 2009, pp. 841 - 848J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
[5]. Gowsikhaa.D Abirami.S and Baskaran.R," Construction of Image Ontology using Low-Level features for Image Retrieval" 2012 International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, INDIA.
[6]. Tajman sandhu and parminder singh ,"domain specific CBIR for highly textued images " Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 2, April 2013.
[7]. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. Scientific American, 284(5):34–43, 2001.
[8]. P. A. Bernstein and E. Rahm. Data warehouse scenarios for model management. In ER 2000, pages 1–15, 2000.
[9]. D. Calvanese and G. De Giacomo. Data integration: logic-based perspective. AI Magazine, 26(1):59–70, 2005.
[10]. D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. Knowledge representation approach to information integration. In Proc. of AAAI Workshop on AI and Information Integration, pages 58–65. AAAI Press/The MIT Press, 1998.
[11]. C. Collet, M. N. Huhns, and W.-M. Shen. Resource integration using a large knowledge base in carnot. IEEE Computer, 24(12):55–62, 1991.
[12]. R. Dhamankar, Y. Lee, A. Doan, A. Y. Halevy, and P. Domingos. imap: Discovering complex mappings between database schemas. In Proceedings of SIGMOD Conference 2004, pages 383–394, 2004.
[13]. A. Doan and A. Y. Halevy. Semantic-integration research in the database community: a brief survey. AI Magazine, 26(1):83–94, 2005.
[14]. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map between ontologies on the semantic web. In Proceedings of the World-Wide Web Conference, 2002.
[15]. D. Dou, D. V. McDermott, and P. Qi. Ontology Translation on the Semantic Web. Journal of Data Semantics, 2:35–57, 2005.
[16]. E. Dragut and R. Lawrence. Composing mappings between schemas using a reference ontology. In Proceedings of International Conference on Ontologies, Databases and Application of SEmantics (ODBASE), 2004.
[17]. L. M. Haas, M. A. Hernandez, H. Ho, L. Popa, and M. Roth. Clio Grows Up: From Research Prototype to Industrial Tool. In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, pages 805–810, 2005.
[18]. A. Halevy, Z. Ives, D. Suciu, and I. Tatarinov. Schema mediation in peer data management systems. In Proc. Of ICDE, 2003.
[19]. Y. Halevy, N. Ashish, D. Bitton, M. J. Carey, D. Draper, J. Pollock, A. Rosenthal, and V. Sikka. Enterprise Information Integration: Successes, Challenges and Controversies. In Proceedings of SIGMOD, pages 778–787, 2005.
[20]. M. Lenzerini. Data integration: A theoretical perspective. In PODS 2002, pages 233–246, 2002.
[21]. J. Madhavan, P. A. Bernstein, P. Domingos, and A. Halevy. Representing and Reasoning about Mappings between Domain Models. In Proc. AAAI 2002, 2002.
[22]. A. Maedche, B. Motik, N. Silva, and R. Volz. MAFRA – A Mapping Framework for Distributed Ontologies. In Proceedings of EKAW 2002, 2002.
[23]. D. McDermott. The Planning Domain Definition Language Manual. Technical Report 1165, Yale Computer Science, 1998. (CVC Report 98-003).
[24]. D. McDermott and D. Dou. Representing Disjunction and Quantifiers in RDF. In Proceedings of International Semantic Web Conference 2002, 2002.
[25]. R. Pottinger and A. Levy. A scalable algorithm for answering queries using views. In Proceedings of the 26th VLDB Conference, pages 484–495, 2000.
[26]. E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. VLDB J., 10(4):334–350, 2001.
[27]. S. Russell and P. Norvig. Artificial Intelligence: A Modern Approach. Prentice-Hall, Inc, 1995.
[28]. A. P. Sheth and J. A. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys, 22(3):183–236, 1990.
[29]. L. Stojanovic, N. Stojanovic, and R. Volz. Migrating data-intensive web sites into the semantic web. In Proceedings of the 2002 ACM Symposium on Applied Computing, pages 1100–1107, 2002.

Appendix–A

FIRE 2010 Data Set

Query Index Query Topic

1	Clashes between the Gurjars and Meenas
2	Attacks by Hezbollahguerrillas
3	Conflict between Advani and Singh alover the Ram Mandirissue
4	Building roads between China and Mount Everest
5	Babri Masjid demolition case started against Advani
6	Problems related to the immunization programme against Japanese Encephalitisinn In
7	Proposed bus service between Srinagar and Muzaffarabad
8	Election campaign of Laloo Prasad Yadav and Ram Vilas Paswan
9	Brinda Karat's allegations against Swami Ramdev
10	Abu Salem, accused in the Mumbai Bomb Blast case ,in jail custody
11	Privatization of the Mumbai and Delhi airports
12	Discussions between Manmohan Singh and Pervez Musharraf regarding the position
13	Popular protests against the arrest of the accused in the Shankar Raman murder case
14	Involvement of Congress ministers in the oil-for-food scam
15	Indian representatives visit Bangladesh
16	Allegations of financial corruption against Pratibha Patil
17	Activities of the Tamil Tigers of SriLanka
18	Taking bribes for arising questions in parliament
19	Indian Navy accused of leaking classified information
20	Racismro won the Big Brother show
21	Pramod Mahajan 'skiller
22	Quarrel between the Ambani brothers regarding ownership of the Reliance Group
23	India dismisses China's claimson Arunachal Pradesh
24	Laloo Prasad Yadav and the fodder scam
25	Monica Bedi and the passport forgery case
26	Drug party at Pramod Mahajan's bungalow
27	Pakistani cricketers involved in a doping scandal
28	Bilateral problems surrounding the Baglihar hydro-electric power project
29	Jaya Bachchan sacked from Rajya Sabha membership
30	Taj heritage corridor scandal
31	Banon Taslima Nasreen's novel"Shame"
32	Furore over there lease of a CD containing anti- Muslim sentiments in UttarPradesh
33	Greater Nagaland
34	New political party formed by Raj Thackeray
35	Sino-Indian relations and border trade
36	Dance bars banned in Mumbai
37	Links between Gutkha manufacturers and the underworld
38	Political clashes in Bangladesh
39	Investigation of the arms scandal in the Defence Ministry
40	Serial blasts in Varanasi
41	Encounter specialist Daya Nayak
42	Controversy over land at Kalinganagar
43	Terrorist strike at Ayodhya
44	Taj Mahal controversy
45	Sex CD scandal involving Anara Gupta
46	Blastson Samjhauta Express
47	Sanjay Dutt's surrender
48	Death of Yasser Arafat
49	Sale of illegal drugs in various Indian states
50	Attack on the Lal Masjid

Appendix–B

Generated Data Set

Query	Query Term Index
1	Mail Server Login
2	Bus
3	Logic Gates
4	Tug Of War
5	Operating System
6	Information Retrieval Techniques
7	Networking
8	Broadband Connection
9	Trojan Horse
10	Antivirus Works
11	Mobile Software
12	Java
13	Company Software Requirements
14	Morning Light
15	Media Files Converter
16	Blue Ray
17	Sql Server
18	Crow Bar
19	Red Cross
20	Plain Text
21	Access File
22	Rolling Thunder Comes Back India
23	Search Engine Seminars
24	Types Linux
25	Windows Types
26	Natural Processing Language
27	Data Processing
28	Congestion Control Mechanisms
29	Parity Check
30	Mail Service Providers
31	Programming Language Contracts
32	Host Server
33	Rack Server
34	Loader Linker
35	Control Panel
36	Data Mining
37	Data Warehouse

38 Image Processing Techniques
39 Data Structure
40 Internet Connection
41 Password Encryption
42 Remote Computer
43 Logical Error Rectification
44 Prefix Infix Conversion
45 Independence Day
46 Bag and Baggage
47 Virus
48 Apple
49 Smart Card
50 Pen Drive
51 Internet Protocol
52 Over the Top
53 Huns Yellow Pages
54 Brooks Brothers Coupons
55 Save the Children
56 Implementation of Security System
57 Steepest Hill Climbing
58 Research Methods and Methodology
59 Ethical Issues in Advertising
60 Electronic Circuits Design