

Comparative Study of MBEWCM and Google

Neha Gupta

Research Scholar: Faculty of Engg & Tech,
Manav Rachna International University
Faridabad, Haryana, India
neha.fbc@mriu.edu.in

Abstract—MBEWCM is a mobile based web content miner for kids, which help the kids to search the content of the web in a secure and reliable manner. The paper discusses various issues that have been considered to evaluate the performance of MBEWCM comparing it with Google. Google is a worldwide known web search Engine and every user is using Google to search the content. So we have compared the results of the Google with MBEWCM to verify the performance of the interface. The comparison has been done on the basis of various web content filtering factors.

Keywords—Google, MBEWCM, Queries, Educational Quality

I. INTRODUCTION

The comparative analysis of MBEWCM with Google (the most popular search engine of the web) was done to assess the relative value, strength and weakness of the MBEWCM. MBEWCM is a mobile based educational web content miner that extracts the age appropriate educational content from the web. The performances of the two systems were assessed in terms of Coverage, Precision, Educational Quality, Security, Recall, Presentation format and Accessibility of results.

The purpose of the comparative analysis was to compare the performance of MBEWCM with Google in answering the queries of the sort likely to be asked by the students/Kids i.e. educational queries.

II. STEPS FOLLOWED

A. Identifying Test Queries:

Test queries have been divided into two groups:

- Educational Queries
- Non-Educational Queries

For Educational queries, five different subject areas were chosen to evaluate the performance of both the systems. These were Science, Social Studies, Math's, English, and Computers. These 5 subject areas are further divided into two groups:

- General educational queries (Math's, English)
- Educational Queries which may extract inappropriate content (Science, Social Studies, Computers)

Specific queries suitable for these subject domains were then chosen from holiday assignments, syllabus, Projects of the kids. These were the “real” student queries.

The tabular representation of Test queries are given below:

Search Group	Search Category	Subject Domain	Evaluators
Educational Queries	General Educational Queries (Search Group 1)	English (Set 1)	Parents/ Teachers
		Math's (Set 2)	Kids
	Educational Queries that may extract inappropriate content (Search Group 2)	Science (Set 3)	Parents
		SST (Set 4)	Kids
		Computers (Set 5)	Teachers
Non-Educational Queries	Entertainment	Songs, Games, Videos	Parents
	Social Networking	Facebook, Twitter	Parents
	General Awareness	Fashion trends, Roadies, bikes, cars etc	Parents

The detail related to the test queries have been provided in Appendix I.

B. Assessment factors:

Comparative analysis of both the systems was done with both the quantitative and qualitative aspects. Quantitative results allowed an assessment of recall, precision and similar factors, while qualitative results allowed inclusion of ideas of quality, reliability and ease of use.

Both the systems were evaluated on the basis of the following factors:

1. Relevance: in terms of Precision and recall
 - Precision: No. of relevant links out of first 10 links on the first page.
 - Recall: No. of irrelevant links on the first page.
2. Educational quality: in terms of Authority, Provenance and Objectivity.
 - Authority: Search results are from an authenticate resource.
 - Provenance: Search results are of educational quality and don't contain any distracted or inappropriate content.
 - Objectivity: Content is related to the topic.
1. Reliability: in terms of Currency, Accuracy and Coverage.
 - Fidelity: Is the information up-to-date.
 - Accuracy: Is the content accurate and as per the age of the child.
 - Coverage: Results are covering the syllabus and are of educational use.
2. Ease of use: Every type of data format is supported and downloadable.

Note: For calculation purpose, we have taken 1 as Yes and 0 as No

C. Methodology Adopted

To avoid possible *investigator* bias in the choices of the queries, the queries are divided broadly into two categories: Educational and Non-educational. Educational queries are selected from five different subject areas of the kids namely English, Math's, Science, SST, Computers and non educational queries are of general nature like Facebook, Songs, Videos download etc. Test queries were then searched on MBEWCM and Google.

D. Result Analysis

10 test queries in each subject domain have been taken to evaluate the performance of both the systems. Search group 1 consists of two sets: Set 1 and Set 2. Search Group 1 is for General Educational Queries. For Search Group 1, both the systems are performing fairly well, although MBEWCM results are superior to Google. Ease of use is same for the systems, reliability, relevance, quality are fairly better in MBEWCM than Google. Charts demonstrated below are showing the results.

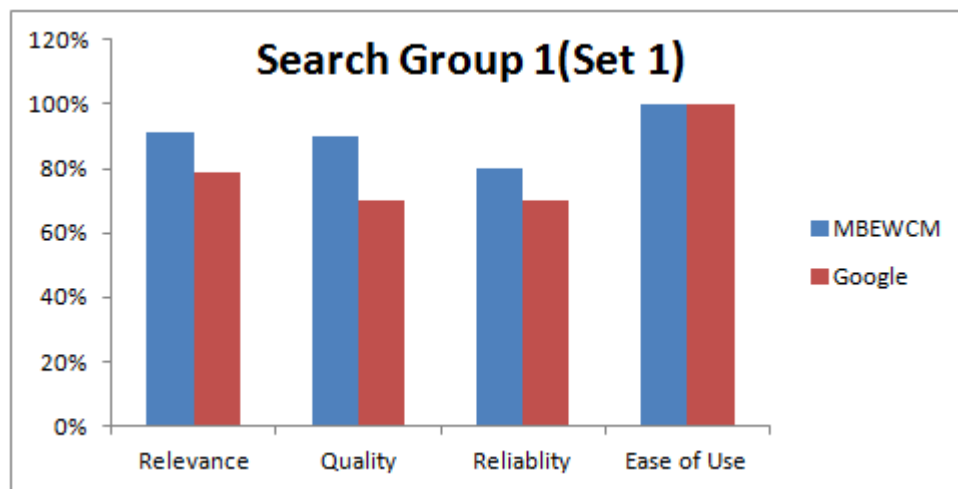


Figure 1: Search Group 1 (General Educational Queries (English))

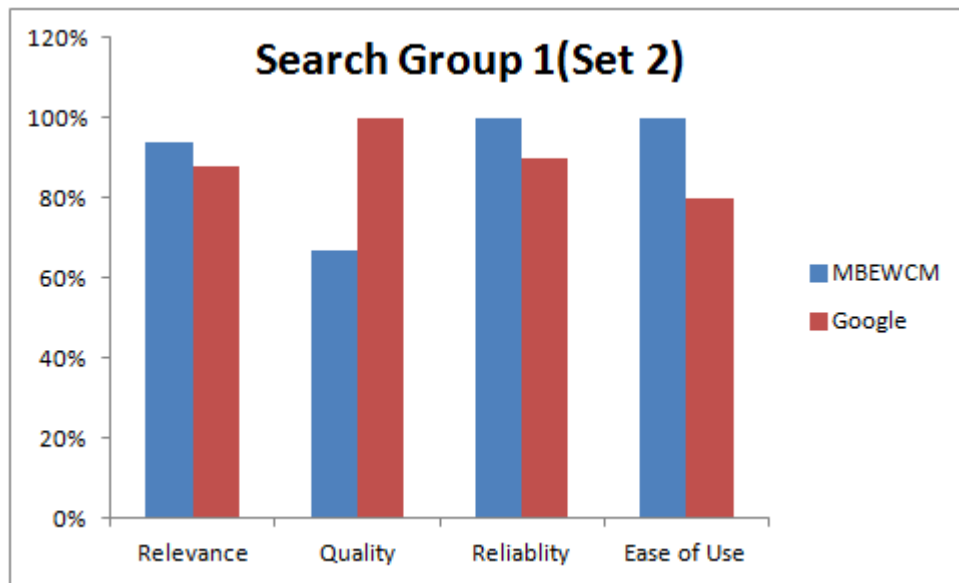


Figure 2: Search Group 1 (General Educational Queries (Math's))

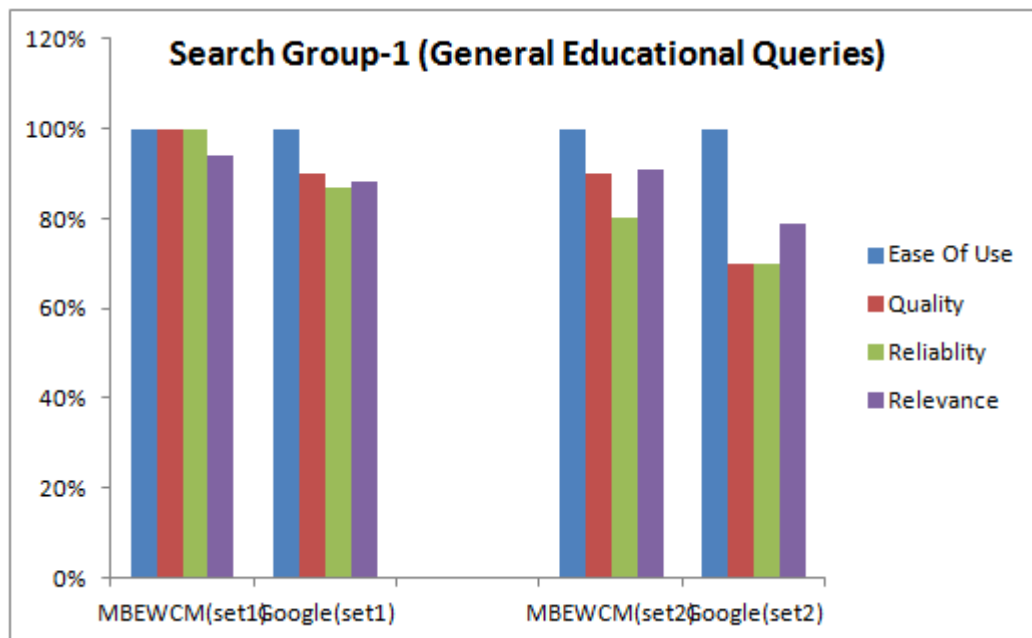


Figure 3: Result Analysis of Search Group 1 (General Educational Queries)

The superiority or the betterment of MBEWCM over Google has been evaluated with the help of Educational Queries that may extract inappropriate content. These types of queries are defined in Search Group 2. Search Group 2 has 3 sets from three different subject domains. (See Appendix II)

For Set 3 having subject domain “Science” MBEWCM results are 35% more relevant than Google. In terms of Quality, 95% documents retrieved by MBEWCM are of good quality in terms of Authority, Provenance and objectivity while Google has retrieved only 55% documents of Good Quality for kids. The discriminating factor is Reliability. MBEWCM results are 95% reliable while Google results are only 25% reliable. Ease of use in both the systems is same.

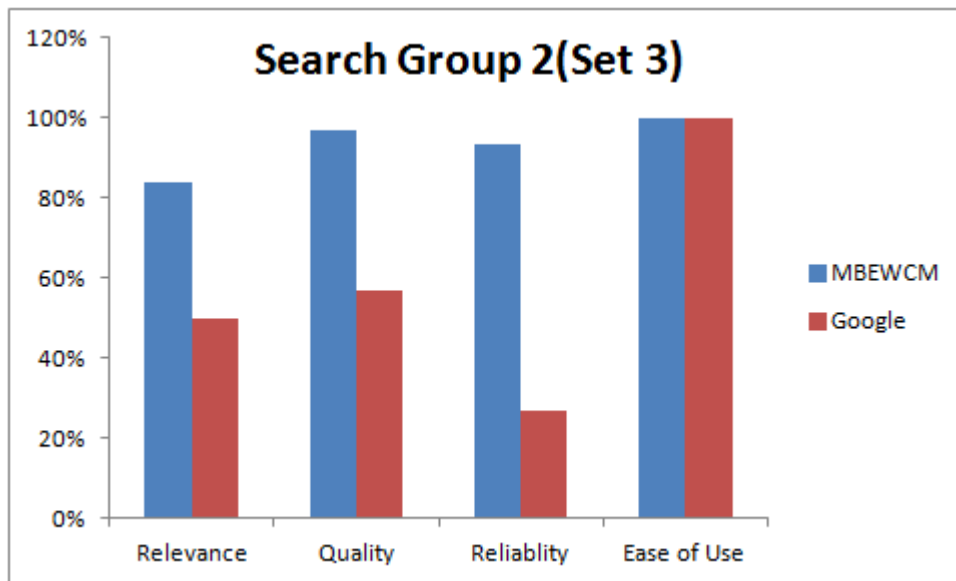


Figure 4: Search Group 2 (Educational queries that may extract inappropriate content (Science))

For Set 4 having subject domain “S.St” MBEWCM results are 25% more relevant than Google. In terms of Quality, 95% documents retrieved by MBEWCM are of good quality in terms of Authority, Provenance and objectivity while Google has retrieved only 45% documents of Good Quality for kids. MBEWCM results are 85% reliable while Google results are only 45% reliable. Ease of use in both the systems is same.

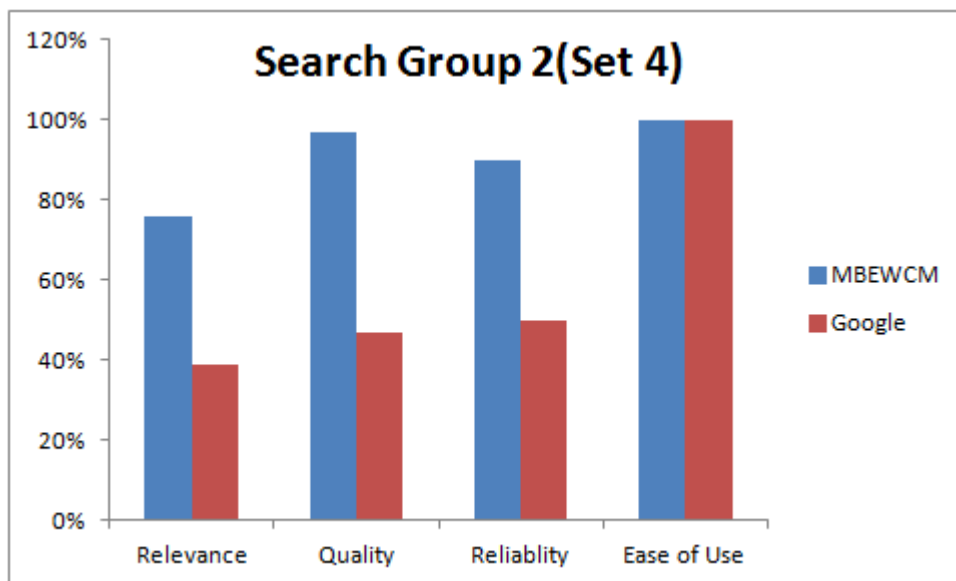


Figure 5: Search Group 2 (Educational queries that may extract inappropriate content (S.St))

For Set 5, Subject domain is “Computers”. The Results given by both the systems are fairly good, but the quality of the documents retrieved by MBEWCM is 20% better than Google. Although Relevance and Reliability is also better in MBEWCM but the difference is not very large. Ease of use is again the same for both the systems.

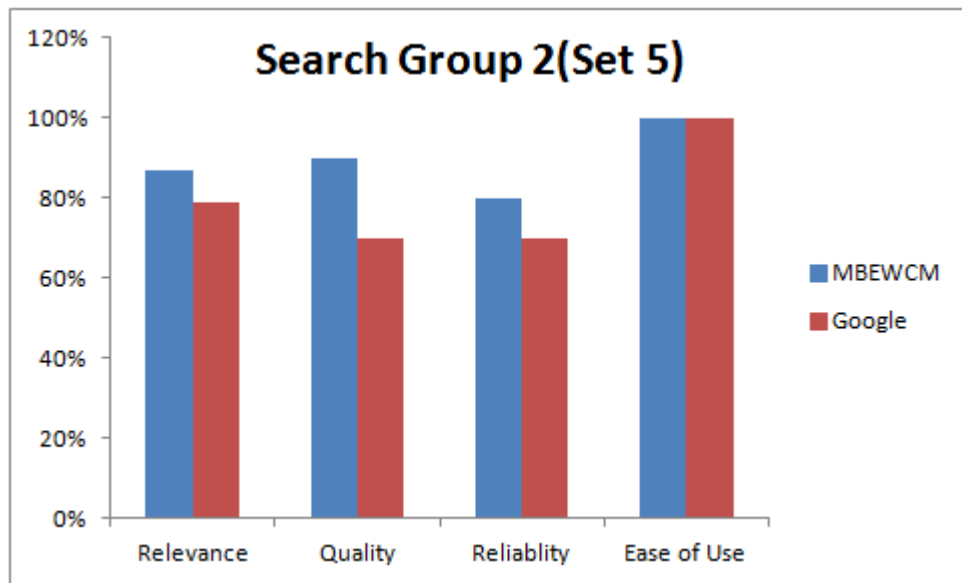


Figure 6: Search Group 2 (Educational queries that may extract inappropriate content (Computers))

The consolidated results of MBEWCM and Google are shown in figure. For Search Group 2 MBEWCM proves to be superior to Google as it is giving better results that are relevant and reliable for both parents and for Kids and are of good Quality that help kids to focus on studies and contain Accurate content.

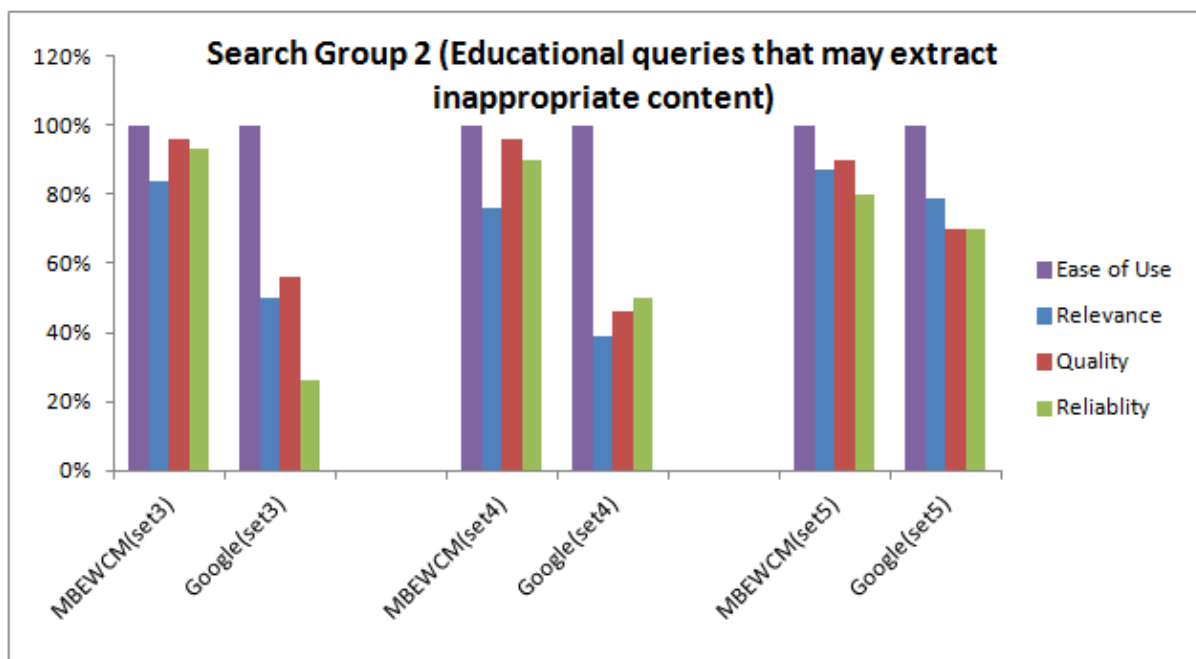


Figure 7: Result Analysis of Search Group 2 (Educational queries that may extract inappropriate content)

III. CONCLUSION

The results of the comparison of Google and MBEWCM show a clear shift of betterment towards MBEWCM. Although for General Educational Queries both MBEWCM and Google are performing well, but the results show a remarkable difference for the Educational queries that may extract the inappropriate content. Ease of use in both the system is same so we can say user is comfortable with both the systems. MBEWCM appears clearly superior in Relevance, Quality and Reliability. The following chart illustrates the same.

The main discriminating factor seems to be Reliability and quality for Search group 2. The results are 40% more reliable and 45% better in quality than Google.

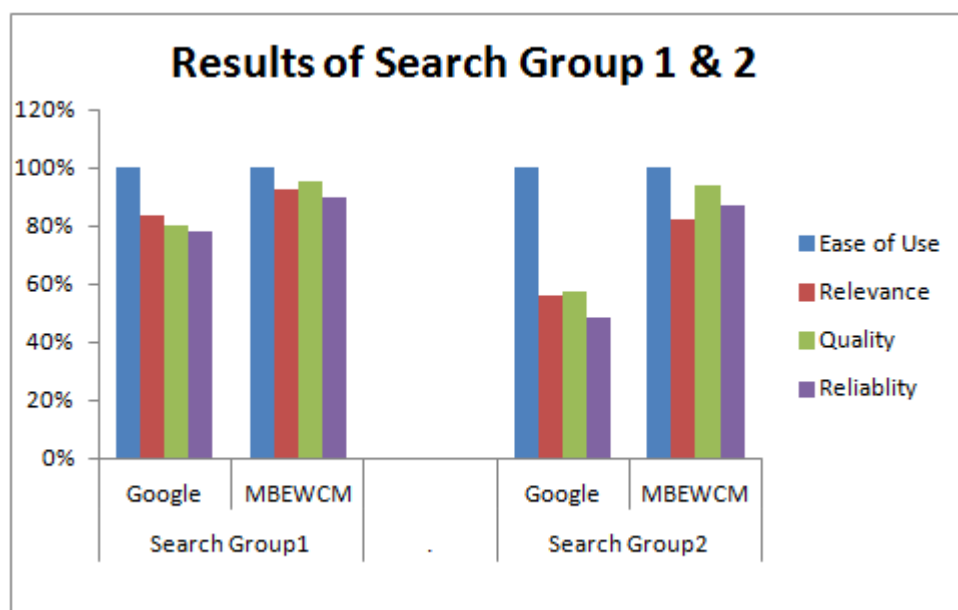


Figure 8: Comparisons of Results of Search Group 1 & 2

In comparing the systems, in terms of advantages and disadvantages, the conclusion can be summarized as follows:

MBEWCM:

- A high proportion of relevant documents retrieved; For educational queries that may extract inappropriate content, remarkable difference can be seen in terms of relevant documents.
- An ability to retrieve a fairly precise set of documents, by prioritizing the documents as per the need of the kid.
- A high proportion of adequate or good quality results;
- No problems with accessibility;
- Retrieved documents have high reliability and are secure for kids;
- For Non educational queries, the interest of the kid is diverted towards the educational aspect of the query.

Google:

- A moderate proportion of relevant documents retrieved;
- An ability to retrieve a fairly precise set of documents; no differentiation for any user.
- Good quality results are comparatively much lower;
- No problems with accessibility;
- Retrieved documents have moderate or low reliability and are not fully secure for kids;
- For Non educational queries, there is no provision to keep the kids secure.

Appendix I

Details of Comparative Analysis of Google and MBEWCM

SEARCH GROUP 1:

Query Set: Set1
 Search Group Name: Educational Queries
 Search Category: General Educational Queries
 Subject Domain: English
 Evaluator: Parents/ Teachers

Test Queries:

1. Homophones
2. Word and their adjectives
3. Punctuations
4. Informal Letters
5. Tenses
6. Prepositions

7. Conjunctions
8. Biosketch
9. Verb Agreements
10. Active and passive voices

Query Set: Set2
Search Group Name: Educational Queries
Search Category: General Educational Queries
Subject Domain: Math's
Evaluator: Parents/ Teachers

Test Queries:

1. Rational Numbers
2. Fraction and Simplification
3. Linear Equation
4. Commercial Mathematics
5. Congruence
6. Convex and Concave Polygons
7. Parallelograms
8. Cyclic Quadrilateral
9. Mid-Point Theorem
10. Empirical Probability

SEARCH GROUP 2:

Query Set: Set3
Search Group Name: Educational Queries
Search Category: Educational queries that may extract inappropriate content.
Subject Domain: Science
Evaluator: Parents/ Teachers

Test Queries:

1. Physical and chemical changes observed in daily life
2. Polymers
3. AIDS
4. Digestion of food in human.
5. Breast Cancer
6. Sex Determination and inherited traits.
7. Reproductive health
8. Human brain and reflex action
9. Fiber to fabric
10. Translucent material.

Query Set: Set4
Search Group Name: Educational Queries
Search Category: Educational queries that may extract inappropriate content.
Subject Domain: S.St
Evaluator: Parents/ Teachers

Test Queries:

1. Disaster Management.
2. Growing up as boys and girls.
3. Food security in India
4. Retreating Monsoon.
5. Literacy and health.
6. Gender Inequality.
7. Microliths.
8. Poverty Line.
9. Diversity and discrimination.
10. Secularism.

Query Set: Set5
Search Group Name: Educational Queries
Search Category: Educational queries that may extract inappropriate content.
Subject Domain: Computers
Evaluator: Parents/ Teachers

Test Queries:

1. GUI Based OS.
2. Computer Virus.
3. Photoshop
4. Cloud Computing.
5. Mobile Communication.
6. Computer Networking.
7. Transmission media.
8. Wi-Fi Technology.
9. Internet and its uses.
10. Software downloads.

SEARCH GROUP 3:

Query Set: Set6
Search Group Name: Non - Educational Queries
Search Category: Entertainment.
Subject Domain: Songs, Games, Videos
Evaluator: Parents/ Teachers

Test Queries:

1. Downloading a Song.
2. Searching a movie/ downloading a movie.
3. Downloading games.
4. Playing online games.
5. Watching videos online.
6. Uploading a video from the mobile.

Query Set: Set7
Search Group Name: Non - Educational Queries
Search Category: Social Networking
Subject Domain: Facebook, Twitter, Second world,
Evaluator: Parents/ Teachers

Test Queries:

1. Chatting
2. Searching a friend
3. Facebook login
4. Twitter Login
5. Exploring second world.
6. How to hack a friend's account
7. Creating online photo albums.
8. MySpace
9. Sharing photographs, videos.
10. E-mailing

Query Set: Set8
Search Group Name: Non - Educational Queries
Search Category: General Awareness.
Subject Domain: Fashion trends, Roadies, bikes, cars etc
Evaluator: Parents/ Teachers

Test Queries:

1. Searching latest bikes, cars
2. Places to hangout
3. Latest mobile phones
4. Fashion trends and accessories
5. MTV Roadies/ Big Boss contestant information, their lifestyle
6. IPL Players
7. How to hack passwords