||Query||
= sqrt ((0.49236883) ^2+ (0.49236883) ^2+ (0.49236883) ^2)
=0.85285887
||Document1||
= sqrt ((0.2) ^2+ (0.23521825) ^2+ (0.29542425) ^2)
=0.42732086

Cosine Similarity (Query, Document1)
=0.35974559/0.85285887*0.42732086
=0.98710695
Cosine Similarity (Query, Document2)
=0.79029023
Cosine Similarity (Query, Document3)
=0.577315832

From above computation document 1 is most similar to query then document 2 and document 3 is least similar

6. Future Work

The documents are retrieved but still there can be a problem of ambiguity. Ambiguity arises when the query entered is very short or user is not exactly sure about its requirements. This can we solved using query expansion. Combination of Pseudo Relevance feedback and co-occurring terms will be used to get the term which will be used for expansion and user can be given the option where the selection can be done. This process will give the exact query and help in retrieving more relevant documents. The query weight will be recalculated using the first set pseudo relevant documents. New weights will be used for second retrieval. The top n documents will be selected after second retrieval the co-occurring terms with the query terms will be extracted from the documents and will be added to the initial query for expansion.

7. Conclusion

We have retrieved the documents for the user query and shown the relevant documents. The tf-idf values were used for representing the documents and query in vector form and cosine similarity was computed to retrieve and rank the documents relevant to the query. The terms were stored in a inverted index were the corresponding documents in which the terms appears and their tf-idf values in descending order were also stored A threshold value was decided and documents with tf-idf greater than or equal to that value where only considered for computing cosine similarity. This helped in reducing the time complexity. Domain specific dictionary was built, the problem of ambiguity which arises when Hindi terms are converted in English is handled by considering all the different meaning of term in English for a particular Hindi term (multiple selection technique).Further we will combine the techniques of pseudo relevance feedback and co-occurring term technique where the weights of term in query will be recalculated and then will be used to retrieve new set of documents where the extracted terms can be added to the initial query for query expansion.

References

[1] Pratibha Bajpai, Parul Verma “Cross Language Information Retrieval: In Indian Language Perspective” International Journal of Research in Engineering and Technology (IJRET) Jun-2014.
[2] Benoit Gaillard, Jean-Leon Bouraoui, Emilie Guimier de Neef, Malek Boualem “Query Expansion for Cross Language Information Retrieval Improvement” 2010 IEEE
[3] Vivek Pemawat, Abhinav Saund, Anupam Agrawal “Hindi - English Based Cross Language Information Retrieval System for Allahabad Museum” 2010 International Conference on Signal and Image Processing
[4] Abdelghani Bellaachia and Ghita Amor-Tijani “Enhanced Query Expansion in English-Arabic CLIR” 19th international conference of database and expert system application.
[5] Lam Tung Giang, Vo Trung Hung and Huynh Cong Phap, “Improve Cross Language Information Retrieval with Pseudo-Relevance Feedback” International Journal of Engineering Research & Technology (IJERT), June 2015
[6] Rekha Vaidyanathan ,Sujoy Das and Namita Srivastava “Query Expansion Strategy based on Pseudo Relevance Feedback and Term Weight Scheme for Monolingual Retrieval “International Journal of Computer ,November 2014
[7] Xuwen Wang, Qiang Zhang, Xiaojie Wang and Yueping Sun ”LDA Based Pseudo Relevance Feedback For Cross Language Information Retrieval” Proceedings of IEEE CCIS2012

Author Profile

Aditi Agrawal received Bachelor of Engineering Degree in Computer Science and Engineering from Nagpur University, India in 2013 and is currently pursuing Master of Technology in Computer Science and Engineering from Shri Ramdeobaba College of Engineering and Management, Nagpur

Avinash J. Agrawal received Bachelor of Engineering Degree in Computer Technology from Nagpur University, India and Master of Technology degree in Computer Technology from National Institute of Technology, Raipur, India in 1998 and 2005 respectively. He has Ph.D. in Computer Science and Engineering from Visvesvaraya National Institute of Technology, Nagpur in 2013. His research area is Natural Language Processing and Artificial Intelligence. He is having 18 years of teaching experience. Presently he is Associate Professor in Shri Ramdeobaba College of Engineering and Management, Nagpur. He is the author of more than 50 research papers in International Journal and Conferences.