ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

An Improved Combined Content-similarity Approach for Optimizing Web Query Disambiguation

ShahidKamal[1] RolianaIbrahim[1*] Imran Ghani[1]

ABSTRACT

The web search engines are exposed to the issue of uncertainty because of ambiguous queries, being input for retrieving the accurate results. Ambiguous queries constitute a significant fraction of such instances and pose real challenges to web search engines. Moreover, web search has created an interest for the researchers to deal with search by considering context in terms of location perspective. Our proposed disambiguation approach is designed to improve user experience by using context in terms of location relevance with the document relevance. The aim is that providing the user a comprehensive location perspective of a topic is informative than retrieving a result that only contains temporal or context information. The capacity to use this information in a location manner can be, from a user perspective, potentially useful for several tasks, including user query understanding or clustering based on location. In order to carry out the approach, we developed a Java based prototype to derive the contextual information from the web results based on the queries from the well-known datasets. Among those results, queries are further classified in order to perform search in a broad way. After the result provision to users and the selection made by them, feedback is recorded implicitly to improve the web search based on contextual information. The experiment results demonstrate the outstanding performance of our approach in terms of precision 75%, accuracy 73%; recall 81% and f-measure 78% when compared with generic temporal evaluation approach and furthermore achieved precision 86%, accuracy 71%; recall 67% and f-measure 75% when compared with web document clustering approach.

☞ keyword : Content similarity, query disambiguation, web search, location, temporal information

1. INTRODUCTION

In recent times, web search optimization has become a very active research area among professionals from both the industry and the academia involved in information retrieval and web search [1].This is because it is most likely that Internet and search engines have become essential gears in our daily life. Regardless of marvelous improvements being made to optimize the web search over the last decade, still much has to be done to deal with the ever-increasing size of the web and needs of the users.

A collection of search results that correspond to a search query are retrieved by conventional search engines. Among these search results, some may lead a user to those internet resources that are different to his/her interests, even though having similarity with the search query. Often this situation arises when search queries are related to more than one topic, some or all of which being of little or no interest to the user, in which case the search results are produced that are descriptive of each of the different topics[2].The search result acquirement process (See Figure 1) begins with a query defined by a user. Based on this query, a document search is conducted in different data sources for example Yahoo, Google etc. in general, between 50-200 results from traditional search engines are collected containing a minimum URL, a snippet and a title and returned back to the user[3].

In pursuit of web search optimization, the Temporal Information Retrieval (T-IR) has gained greater attention in recent years. However, majority of these solutions either focus on development of suitable tools or perform behavioral analysis based on log data. Significant numbers of user search queries have strong temporal components or characteristics. These are the queries whose underlying intent may be to obtain newest information, past or anticipated events and largely depend on time [4].

Context is an important source of information in computing environments. The term context is defined by the authors of [5] as “any information that can be used to characterize the situation of an entity”.

[1] Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 skudai, Johor Malaysia
* Corresponding author (roliana@utm.my)
[Received 4September 2015, Reviewed 12 September 2015, Accepted 17November 2015]

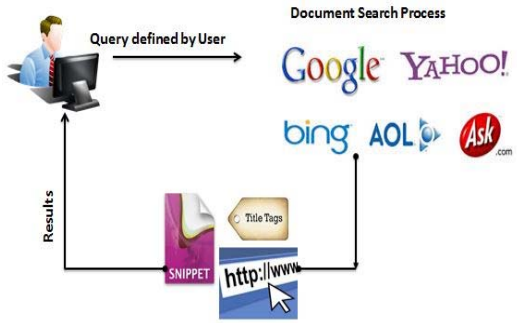

Figure 1. Conventional Search Process

An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves. Based on the information being defined as context, the query disambiguation can be greatly improved by applying contextual information[1].

In essence, we focus on disambiguating a text query with respect to its contextual as well as temporal purpose and propose a combined approach that takes into account the contextual and temporal information exist in the search results. Furthermore, we have introduced a different concept of utilizing implicit user feedback based on the selection frequency for the purpose of refining the search results according to the user requirements. In essence, our method is a three stage process, taking the identified ambiguous queries from the well-known datasets[6, 7] to make them clear in concept with respect to defined parameters i.e. location and year in this case. Based on those queries, the search results are refined according to the user requirements having connectivity with the text query. Furtherance in, the user selects the results and on the basis of their selection, the search process is refined to get more accurate results accordingly.

In this paper, acombined methodology has been created containing on relevant (location) and temporal (year) data to be transformed for improved web information retrieval. So as to assess the methodology, two current methodologies[1, 8] have been chosen for the benchmarking. An algorithm has additionally been proposed for the better hunt as indicated by user needs. The methodology is focused around the algorithm and the results are defined in such a path, to the point that it can understand the logical and additionally transient data accordingly of user queries. The remainder of this paper is structured as follows. In section 2 we present the related work. In section 3 we introduce our methodology. Result and discussion are given in section 4. Finally, we conclude this paper in section 5 with some final remarks and future research directions.

2. RELATED WORK

Disambiguating the search intent and improving the accuracy of resulting information is a hot research area where numerous contributions have been made to address these issues. It is widely believed that some queries submitted to search engines are ambiguous by nature e.g. World Cup, Cultural Show etc. Different studies have investigated the inherent ambiguity issues of search queries in various ways. The work presented in [9, 10] describes what ambiguous queries are. For instance referring to word “java” in search query, it is not clear whether the user intent is the application or the Indonesian city.In information retrieval, removal of ambiguity is of paramount importance both from the user and system perspective [11]. In this section we briefly describe the various contributions that have been made for query disambiguation and search optimization.

Ferragina and Scaiella [12] addressed the problem of query disambiguation by employing a voting system that resolved all ambiguous terms simultaneously. Their system makes use of various characteristics associated with different fragments of the input strings and completely overlooks the temporal and contextual features.Alonso et al. [13] first introduced the temporal clustering on the basis of topics and time. Ricardo, at al. [14]highlighted the disambiguation of text queries with respect to temporal feature. They proposed a two-stage process where relevant temporal expressions are extracted from results and then grouped into same clusters with respect to common year. Their approach was based on the idea of finding one non trivial term in text and focused on temporal clustering. Their work solely relies on temporal features thereby compromising the accuracy of the results.

In a new way, Anastasiu, D.C, et al [1] investigated the problem of query disambiguation by making use of keywords search and contextual information. First the articles were

retrieved on the basis of both combined fragments of query as well as contextual terms. Next they retrieved the articles based on only query terms and finally similarities were computed. Eventually, the commonly retrieved results were presented to users for their selection.Loia et al. [15] has exploited the user behavior and preferences that are extracted during user's navigation to present the personalized clusters.Yu and Jeon[16] have proposed a context-aware recommender system. In order to provide preferable items to the user, the system uses the user's history and current context to filter the content-based information.

Richardson et al. [17]analyzed the user feedback as clicks on ads and used a method to predict user actions. The user feedback for web query disambiguation has not been taken into account previously up to extent of our knowledge. Nevertheless [18]suggested users a solution so they can directly provide feedback and can utilize this feedback for extracting the information, however, it lacks in usage of user feedback properly in order to improve the search results. Most of the work has been done in query disambiguation as mentioned above, but the problem has still not been resolved completely. In search results, certainly there exist errors because user feedback has not been given weight. In this paper, we employ user feedback for search results refinement at continuous level that can reduce the disambiguation constantly.

3. PROPOSED METHODOLOGY

Research methodology is an academia established regularity framework for the collection and evaluation of existing knowledge for the purpose of arriving at, and validating the new knowledge. Research methodology not only frames the study but also identify the tools, strategies, ad criteria for the success of the research. In order to carry out this research, the methodology is based on a hybrid approach, i.e. a combination of exploratory research and experimentation. The research work focuses on the web queries ambiguity identification and disambiguation approaches and their correctness for web information retrieval. The Figure 2 below presents the hybrid procedure for the web query disambiguation that has been proposed in this paper.

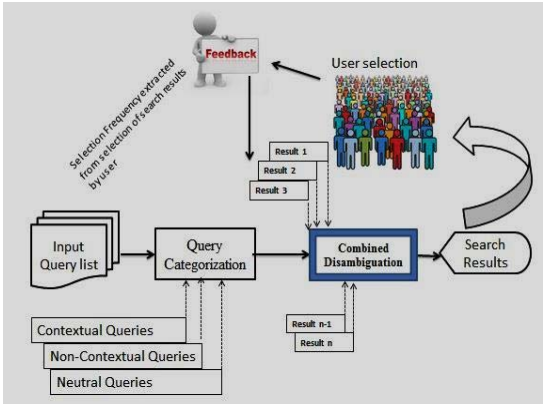

Figure 2. Proposed Research Methodology

The following sub-sections describes the details of the each step in order understand the functioning of each component in the proposed methodology.

3.1 QUERY INPUT

In this component of the proposed approach, we deal with the input queries where the availability of contextual information is helpful in refining the search results. We denote the input query by q and consider a result set n of 10 entries against each query of two different datasets namely; GISQC_DS andAMBIENT datasetsshown in Figure.3 and Figure.4 independently. We use Google Web Services for information retrieval in accordance with the input queries.

```
Search Word: harry potter
----------------------------------------------------
Harry Potter - Wikipedia, the free encyclopedia   --
Harry Potter (film series) - Wikipedia, the free enc
Harry Potter and the Sorcerer's Stone (2001) - IMDb
Warner Bros. Studio Tour London - The Making of Harr
Harry Potter | Scholastic.com   ----   http://harryp
Pottermore: A unique online Harry Potter experience
Harry Potter - Harry Potter Wiki   ----   http://har
Harry Potter Wiki   ----   http://harrypotter.wikia.
J.K. Rowling   ----   http://www.jkrowling.com/---Ur
----------------------------------------------------
Total Country Counter: [0]
Total City Counter: [1]
Total Year Counter:  [1]
```

Figure 3. Search results against query "harry potter"

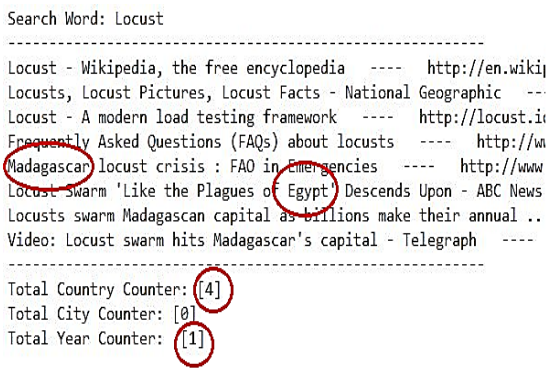

Figure 4. Search results against query "Locust"

3.2 QUERY CATEGORIZATION

After taking the query from user, next we categorize it in accordance with its concept ambiguity. Similar approach has been adopted in [19-21] where three different types of concept queries are defined: ambiguous, broad and clear. In our approach, we further classify the ambiguous queries on the basis of both contextual and temporal information into contextual (location based), non-contextual (year based), and ambiguous given in Table 1 below.

Table 1. Different types of queries categorization and distribution percentage values

Total Number of Queries (450) [GISQC-DS]		
Query Type	**No: of Queries**	**Query Percentage (%)**
Ambiguous	220	49
Clear	176	39
Broad	54	12
Input: Ambiguous Queries (220)		
Contextual	62	28
Non-contextual	51	23
Ambiguous	107	49
Improvement Input: Queries (450)		
Ambiguous	107	24
Clear	289	64
Broad	54	12
Input: Ambiguous Queries (44) [Ambient Dataset]		
Contextual	23	52
Non-Contextual	13	30
Ambiguous	8	18

3.3 SEARCH RESULT SELECTION AND USER FEEDBACK

In this step, the obtained results are clarified based on the nature of the information found in query i.e. temporal as well as contextual comprising of location in form of country name or city name. After organizing the results, the similarities are computed in order to extract the most desired information from the search records. Then the results will be presented to the users for selection according to their preference and in the way of their selection made, the implicit feedback based on selection will be recorded to be utilized further in refinement. The Figure 5 below presents a snapshot of the results being displayed and user feedback recorded in form of selection frequency.

4. RESULTS AND DISCUSSION

The primary goal of our query disambiguation approach is to improve the accuracy of the search results according to user requirements. For the evaluation of an information retrieval mechanism, the most commonly used measures are precision, accuracy, recall, and F1-measure. In informal retrieval context, precision (Eq.1) represent the percentage relevance of the retrieved documents with the user's information needs. Accuracy (Eq.2) is the degree of closeness of retrieved documents to the actual intent of user. Similarly, recall (Eq.3) is the fraction of documents that are relevant to the query that are successfully retrieved. Likewise, F1-measure (Eq.4) is the weighted harmonic mean of precision and recall.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1-\text{measure} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Where True Positive (TP) is the number of locations /years are correctly identified as relevant, True Negative (TN) is the number of locations /years correctly identified as irrelevant or incorrect, False Positive (FP) is the number of locations /years wrongly identified as irrelevant and False Negative (FN) is the number of locations /years wrongly identified as relevant.

Query	Results	IF	O-1	O-2	O-3	O-4	O-5	O-6	O-7	O-8	O-9	O-10	Maximum Priority
ford	Ford ? New Cars, Trucks, SUVs, Hybrids & Crossovers | Ford Vehicles	0	0	1	2	3	0	4	5	6	7	0	7
	Ford Malaysia | New Cars, SUVs & Commercial vehicles. ---- http:/	0	1	0	0	2	0	0	3	4	5	0	5
	Ford Malaysia | All New Ecosport - The new urban SUV from Ford -	0	1	0	2	0	0	3	4	0	5	0	5
	Ford Motor Company - Wikipedia, the free encyclopedia ---- http:	0	0	1	2	0	3	4	0	0	0	5	5
	Ford Motor Company - Dearborn, Michigan - Corporate Office, Car ..	0	0	1	2	0	0	3	4	5	6	7	7
	Ford Malaysia | Facebook ---- https://www.facebook.com/FordM	0	1	2	3	0	0	4	0	0	0	5	5
	Ford Fiesta in Malaysia | Facebook ---- https://www.facebook.cor	0	0	1	0	0	2	3	0	4	5	0	5
	Ford - YouTube ---- http://www.youtube.com/user/Ford---United	0	1	2	0	0	0	3	0	0	4	0	4
	Ford Cars Malaysia Reviews, Prices and Ratings | LIVE LIFE DRIVE --	0	0	1	2	3	0	4	0	5	0	0	5

Figure 5. Results selection frequency in result of different responses

On comparison of precision parameter under evaluation for the proposed approach and existing General Temporal Evaluation (GTE) approach used by [22], the same dataset set is analyzed and results are plotted on the basis of precision, accuracy, recall and f-measure. The Table 2 presents the results being derived after execution.

Table 2. Results obtained using GISQC data set

Approach	Query Count	Performance Parameters			
		Precision	Accuracy	Recall	F-measure
GTE	220	0.64	0.85	0.72	0.78
Our approach	@ 61	0.70	0.65	0.79	0.74
	@ 122	0.73	0.71	0.80	0.76
	@ 183	0.73	0.74	0.82	0.77
	@ 220	0.83	0.82	0.85	0.84
	Overall (@220)	0.75	0.73	0.81	0.78
Improvement		0.11	-0.12	0.09	0

Although our approach lacks in accuracy in compraison with GTE but however the overall precision and recall gain is 11% and 9% respectively. Though it seems marginal improvement but the improvement is progressively increasing over the number of queries being increased. This trend leed to the conclusion that proposed approach will produce more precised results when number of queries will be higher. The Figure 6 shows the performance graph of our approach when it is compared with the GTE at different number of queries. The significant difference between our approacha GTE is that we have applied the progressive query evaluation techniue to correctly observe the performance while GTE has taken all the queries at once for the purpose.

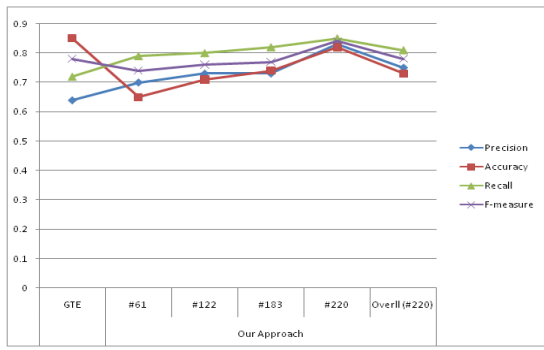

Figure 6. Comparative results with GTE

In the above Figure 6, we have used legend comprising four IR measures namely; precision as in blue line, accuracy as in red line, recall as in green line and f-measure as in purple colored line. The Figure 6is based on the data presented in the Table 2earlier. In comparison with the precision obtained through GTE, our approach has achieved relative performance while in different segments of the queries. Henceforth at overall our approach attained the performance 0.75 and outperformed the GTE by achieving 0.11 more percentage shown in the Figure 6. But in terms of accuracy, our approach did not meet the values obtained through using GTE, however the significant result i.e. 0.82 is obtained when the number of queries are increased to 220. This shows that in future the increasing number of ambiguous queries will result better while utilizing our approach. Again in terms of recall, our approach produced promising results i.e. 0.81 and hence achieved an overall conquest of 0.09 over GTE. Unfortunately, in terms of f-measure the results obtained for both approaches are same i.e. 0.78 and hence remains equal in this measure. But from the obtained

results, we can perceive that the increased number of queries will boost our approach to get better results as in this case, we get 0.84 while the maximum number of queries i.e. 220.

In the similar way, on comparison of precision parameter under evaluation for the proposed approach and existing approach WDC-CSK used by [23]the same dataset set is analyzed and results are plotted on the basis of precision, accuracy, recall and f-measure. The Table 3 presents the results being derived after execution.

Although our approach lacks in accuracy in comparison with WDC-CSK but however the overall recall and f-measure gain is 8% and 13% respectively. Though it seems marginal improvement but the improvement is progressively increasing over the number of queries being increased. This trend lead to the conclusion that proposed approach will produce more precise results when number of queries will be higher. The Figure 7 shows the performance graph of our approach when it is compared with the WDC-CSK at different number of queries.

Table 3.Results obtained by using AMBIENT data set

Approach	Query Count	Performance Parameters			
		Precision	Accuracy	Recall	F-measure
WDC-CSK	44	0.86	0.82	0.59	0.62
Our approach	@ 21	0.87	0.69	0.67	0.75
	@ 44	0.85	0.72	0.67	0.75
	Overall (@44)	0.86	0.71	0.67	0.75
Improvement		0	-0.11	0.08	0.13

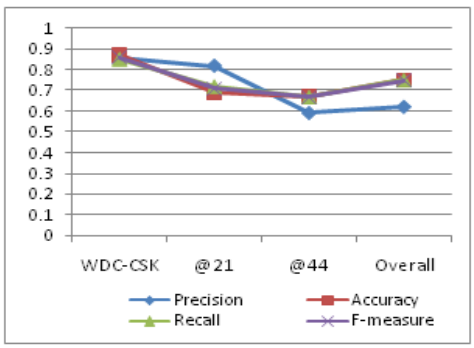

Figure 7. Comparative results with WDC-CSK

The above Figure 7 shows the IR common measures namely precision, accuracy, recall and f-measure with the blue, red, green and purple colored line respectively. The Figure 7 is based on the data presented in the Table 3 earlier. Although our approach produced unattractive result in terms of precision when compared to WDC-CSK, but rather it shows the better value when there is small set of queries to be disambiguated. Hence our approach can perform well in the case of small set of ambiguous queries to be processed for the purpose of disambiguation in retrieving the relevant information. In terms of recall, our approach has achieved better performance by achieving 0.08 percentage more than the compared one. Furtherance in, our approach produced 0.13 percentage more by having the value 0.75 when compared to WDC-CSK that attained the value as 0.62. The significant difference between our approach and WDC-CSK is that we have applied the progressive query evaluation technique to correctly observe the performance while WDC-CSK has taken all the queries at once for the purpose.

Furtherance in, the objectives of this study was to improve the accuracy of the search results in user perspectives. So we also concentrate on user selection results that are recorded based on his/her selection previously depicted in Figure 4 earlier. Initially the first column “initial frequency” is set to 0, and further in response selecting any of the retrieved result, the frequency has been updated. For example the first result has been selected by the user in 2, 3, 4 and 6,7,8,9 iterations respectively, so as having the user frequency selection value as 7 in the row against the selections. Moreover, the 5th result has also been selected in 2, 3 and 6, 7, 8,9,10 iterations as well. We can perceive from the Figure 4 that 1st and 5th result of the query has been selected more by the users and hence got the user selection frequency value as 7. So in concluding remarks, it can be inferred that the when-ever the query “ford” will be given to the systems, the system will give the maximum priority to the 1st and 5th results i.e. it will give priority to the results that are associated with the 1st and 5th web snippets to displayed in response of user query.

The performace evaluation was accomplished in terms of precison and authority[24] in order to evaluate the performance of our adaptive disambiguation approach using the user feedback. Authority measures the ability of the algorithm to produce pages that are most likely to be selected by the users;

while precision measures the degree to which the algorithm produces an accurate results. The performance of our approach is presented in Figure 8 with different methods GTE , Constraint-based Precision (CbP)[25] and Iterative Algorithm (IA)[4]. Although CbP has closer relation with our approach but generally our performance has outperformed all the previous approaches in terms of precision 0.93 and authority 0.90.

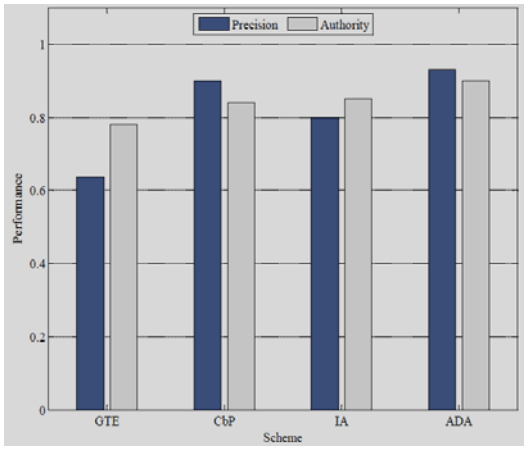

Figure 8. Comparative Results Analysis

From the above Figure 8, we have noticed the improvement of our proposed appraoch, when compared to other approaches i.e. GTE , CbP and IA respectively. The GTE reperesents 0.638 as precision and 0.78 as authority . the CbP has produced the values for the precision and authority as 0.9 and 0.84 respectively. Furthermore, the Iterative Algorithm (IA) has prodcued the results as 0.8 precisiona and 0.85 authority seperaratly. Overall, our approach, has generated the values as 0.93 in terms of precision and 0.90 in terms of authority respectively. Although our approach is in close relation with CbP in terms of precison but in other parameter it has outperformed it. In the conclusion of this experiment results, our approach has shown better results than all the three approaches in both the parameteres specified and hence it has created a posoition in the research community to be further explore and utilized for the future studies.

In the next phase, we have compared our approach in terms of data size being generated in reponse of user selection. The relative performance of our approach is depicted in the Figure 9 below.

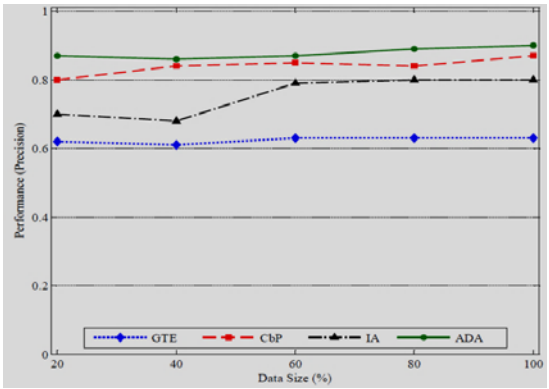

Figure 9. Comparison with other approaches

we found that the performance of GTE is constant when the selection data is increasing gradually represenetd as blue colored line. The CbP, that has been shown in red colored line caused variations in the values i.e. from 0.80 to 0.88 with different data size i.e. number of queries. Furthermore, the IA being displayed in black colored line, produced the results as 0.65 to 0.80, but it has been noticed when the number of queries has increased from 60 , the value become constant and produced no change in the result.When our approach ADA based on user selection frequency is introduced, the performance is improved being presenetd as in green colored line. It has produced results more than 0.80 in all the cases with having diiferent set of queries i.e. 20 to 100. The more selection data is generated, the higher is the performance of search results based on user selection. In the conclusion of these experiment results, our approach has shown better results than all the three approaches in both the parameteres specified and hence it has created a posoition in the research community to be further explore and utilized for the future studies.

5. CONCLUSION

Query disambiguation for the web information retrieval is a challenging task due to ever increasing size of web information. In the perspective of web search queries, a combination of context and temporal information may help in producing high quality search results with reduced search efforts and processing time. Through this combination, an integration of information extraction mechanisms for web information

retrieval and query disambiguation with regard to better and efficient search can be perceived. This paper has addressed the problem of web query disambiguation for the effective web information retrieval. To address these issues, this paper has contributed a combined query disambiguation approach to apply contextual information in combination with temporal information. Furthermore, this paper exploits the user feedback to obtain better and refined search results according to the user requirements. The contributions of this paper have demonstrated that using contextual and temporal information together, improved performance in terms of precision, accuracy, recall and f-measure is achieved. In the future directions for the researchers, we need to look for the control over web snippets, contextual similarity in robust applications, optimal web information retrieval based on different aspects of user feedback and more interestingly the contextual relevant image retrieval.

Acknowledgement

We are thankful to Universiti Teknologi Malaysia and Ministry of Higher Education (MOHE) Malaysia (Vot No: 4F315) and Research University Grant Scheme (Vot No.Q.J130000.2528.05H84) in provision of facilities and support in order to conduct this research study. Moreover we extend our thanks to Higher Education Commission (HEC) of Pakistan and the Gomal University D.I.Khan,Pakistan.

References

[1] Anastasiu, D.C., et al., "A novel two-box search paradigm for query disambiguation," World Wide Web, 16(1), pp.1-29, 2013. http://dx.doi.org/10.1007/s11280-011-0154-0,

[2] Chowdhury, A.R. and G.S. Pass, Query disambiguation, Google Patents, 2014. http://www.google.com/patents/US7428530

[3] Carpineto, C., et al., "A survey of web clustering engines," ACM Computing Surveys (CSUR), 41(3), pp.17, 2009. http://dx.doi.org/10.1145/1541880.1541884

[4] Joho, H., A. Jatowt, and B. Roi. "A survey of temporal web search experience," in Proceedings of the 22nd international conference on World Wide Web companion, International World Wide Web Conferences Steering Committee, 2013.

[5] Dey, A.K., "Understanding and Using Context," Personal and Ubiquitous Computing, 5(1), pp. 4-7, 2001. http://dx.doi.org/10.1007/s007790170019

[6] Campos, R., Google Insights for Search Query Classification dataset (GISQC_DS), 2011. http://www.ccc.ipt.pt/~ricardo/datasets/GISQC_DS.html

[7] Carpineto C. and R. G., Ambient dataset, 2008. http://credo.fub.it/ambient/

[8] Campos, R.N.T., "Disambiguating implicit temporal queries for temporal information retrieval applications," Universidade do Porto, 2013. https://dias.users.greyc.fr/publications/PhdThesisRCampos.pdf

[9] Brin, S. and L. Page, "The anatomy of a large-scale hypertextual Web search engine," Computer networks and ISDN systems, 30(1),pp. 107-117, 1998. http://dx.doi.org/10.1016/S0169-7552(98)00110-X

[10] Kleinberg, J.M., "Authoritative sources in a hyperlinked environment,"Journal of the ACM (JACM), 46(5),pp. 604-632, 1999. http://dx.doi.org/10.1145/324133.324140

[11] Boston, C., et al., "Wikimantic: Toward effective disambiguation and expansion of queries," Data & Knowledge Engineering, 90, pp. 22-37, 2014. http://dx.doi.org/10.1016/j.datak.2013.07.004

[12] Ferragina, P. and U. Scaiella. "Tagme: on-the-fly annotation of short text fragments (by wikipedia entities),"in Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010. http://dx.doi.org/10.1145/1871437.1871689

[13] Alonso, O., M. Gertz, and R. Baeza-Yates. "Clustering and exploring search results using timeline constructions,"in Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009. http://dx.doi.org/10.1145/1645953.1645968

[14] Campos, R., et al. "Disambiguating Implicit Temporal Queries by Clustering Top Relevant Dates in Web Snippets,"in Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE/WIC/ACM International Conferences on 2012. Macau: IEEE, 2012. http://dx.doi.org/10.1109/WI-IAT.2012.158

[15] Loia, V., et al., "Interactive knowledge management for agent-assisted web navigation," International Journal of Intelligent Systems,22(10),pp. 1101-1122, 2007. http://dx.doi.org/10.1002/int.20239

[16] Yu, J. and M. Jeon,"A context-aware intelligent recommender system in ubiquitous environment," in 10th IASTED international conference on artificial intelligence and applications,pp.229-234, 2010. http://www.actapress.com/Abstract.aspx?paperId=37710

[17] Richardson, M., E. Dominowska, and R. Ragno,"Predicting clicks: estimating the click-through rate for new ads," in Proceedings of the 16th international conference on World Wide Web, ACM, 2007. http://dx.doi.org/10.1145/1242572.1242643

[18] Chai, X., et al.,"Efficiently incorporating user feedback into information extraction and integration programs," in Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, ACM, 2009. http://dx.doi.org/10.1145/1559845.1559857

[19] Song, R., et al.,"Identification of Ambiguous Queries in Web Search," Information Processing & Management,45(2), pp. 216-229, 2006. http://dx.doi.org/10.1016/j.ipm.2008.09.005

[20] Song, R., et al.,"Identifying ambiguous queries in web search," in Proceedings of the 16th international conference on World Wide Web, ACM, 2007. http://dx.doi.org/10.1145/1242572.1242749

[21] Campos, R., A. Jorge, and G. Dias,"Using web snippets and query-logs to measure implicit temporal intents in queries," in SIGIR 2011 Workshop on Query Representation and Understanding, University of Massachusetts Amherst, 2011. http://www.researchgate.net/publication/265284219_Using_Web_Snippets_and_Web_Query-logs_to_Measure_Implicit_Temporal_Intents_in_Queries

[22] Campos, R., et al., "GTE-Cluster: A temporal search interface for implicit temporal queries, in Advances in Information Retrieval," Springer International Publishing: Switzerland,pp. 775-779, 2014. http://dx.doi.org/10.1007/978-3-319-06028-6_94

[23] Cobos, C., et al., "Clustering of web search results based on the cuckoo search algorithm and Balanced Bayesian Information Criterion," Information Sciences, 281,pp. 248-264, 2014.http://dx.doi.org/10.1016/j.ins.2014.05.047

[24] Xue, G.-R., et al.,"Optimizing web search using web click-through data," in Proceedings of the thirteenth ACM international conference on Information and knowledge management, ACM, 2004. http://dx.doi.org/10.1145/1031171.1031192

[25] Li, Y., et al., "Name disambiguation in scientific cooperation network by exploiting user feedback,"Artificial Intelligence Review, 41(4),pp. 563-578, 2014. http://dx.doi.org/10.1007/s10462-012-9323-5

◐ 저 자 소 개 ◑

Shahid Kamalis a PhD student in the Software Engineering Research Group (SERG) at Faculty of Computing, UniversitiTeknologi Malaysia. He received his BSc Computer Science from Gomal University, MSc Computer Science from ICIT, Gomal University D.I.Khan Pakistan. His research includes information systems, data mining, web search and its related issues and information integration. Besides, he is faculty member of the ICIT Gomal University D.I.Khan, Pakistan since 2007.

Roliana Ibrahim is a senior lecturer at the Faculty of Computer Science and Information Systems, UniversitiTeknologi Malaysia. She has been servicing UTM for more than 10 years after a few years' experience working as a system developer in the industry. She received her BSc (Hons) Computer Studies from Liverpool John Moores University, MSc Computer Science from UniversitiTeknologi Malaysia and PhD in the field of Systems Engineering from Loughborough University. Her current research project relates to the development of ontology based data warehousing and data mining model for oral cancer research data repository and improvement of data mining techniques for risk and survival analysis of cancer patients.

Imran Ghani is a Senior Lecturer at Faculty of Computing, UniversitiTeknologi Malaysia (UTM) Johor Campus. He received his Master of Information Technology Degree from UAAR (Pakistan), M.Sc. Computer Science from UTM (Malaysia) and Ph.D. from Kookmin University (South Korea). He is the member of Software Engineering Research Group (SERG). He teaches Software Architecture and Design, Software Engineering, Secure Software Development and (Software) Application Development. Dr. Imran Ghani has 50+ publications in Journals, Proceedings, and Book chapters.