I.J. Information Engineering and Electronic Business, 2015, 6, 1-7
Published Online November 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2015.06.01

Bi-gram based Query Expansion Technique for Amharic Information Retrieval System

Abey Bruck, Tulu Tilahun
Department of Computer Science and IT, AMiT, Arba Minch University, Arba Minch, 21, Ethiopia
Email: {bruckabey, tuttilacs}@gmail.com

***Abstract*—**Information retrieval system has been using to connect users of the information and information repository corpora. Even though the task of information retrieval systems is to retrieve relevant information, it is very difficult to find a perfect information retrieval system which is capable of retrieving relevant and only relevant documents as per user's query. The aim of this research is to increase precision of an Amharic information retrieval system while preserving the original recall. In order to achieve this bi-gram technique has been adopted for the query expansion. The main reason for performing query expansion is to provide relevant documents as per users' query that can satisfy their information need. Because users are not fully knowledgeable about the information domain area, they mostly formulate weak queries to retrieve documents. Thus, they end up frustrated with the results found from an information retrieval system. Amharic language has many meaning for a single word and also the word can be found in different form. These are some of the challenges that made the information retrieval system performing at very low level. Query expansion methods outperform in differentiating the various meanings of a polysemous term and find synonymous terms for reformulating users' query. Bi-gram technique uses the underling theory of expanding a query; using terms that appear adjacent to a query term frequently. The proposed technique was integrated to an information retrieval system. Then the retrieval system is tested with and without using bi-gram technique query expansion. The test result showed that bi-gram based method outperformed the original query based retrieval, and scored 8% improvement in total F-measure. This is an encouraging result to design an applicable search engine, for Amharic language.

Index Terms—Information retrieval, bi-gram, query expansion, relevant document, relevance feedback.

I. Introduction

As the knowledge of mankind grows, the written materials representing knowledge grows as well. This intern makes it difficult to retrieve a single document among all of those accumulated document collection. Imagine what the difficulty would be, to find a single sentence from a book, than to find it from a paragraph. This is exactly what is observed in the WWW these days. This problem forces users to elicit their information from an ocean of documents. Information retrieval systems are designed, because of the fast growing electronic or digital information worldwide. The recent decade has witnessed, an explosive growth of online information, including Web pages, news articles, email messages, scientific literature, and information about all kinds of products on WWW [1]. As EMC Corporation [2] released a statistical data which states that, in 2009, the amount of digital information grew 62% over 2008, which is 800 billion GBs or 0.8 ZBs. The amount of digital information created in 2010 is recorded to be 0.9 ZB, representing a compound annual growth rate of 57%, according to [2]. This clearly shows that the pace of information growth is at alarming rate. An information retrieval system serves as a bridge between users with information need, and a vast amount of information repository corpora.

In order to achieve the goal of this research; further literature review has been done on researches that have been made on Amharic IR systems and related concepts. Related works regarding the historical, cultural and future directions of the Amharic language and Amharic IR are explored. Information can be found in many forms such as: in audio, video, text, images, text etc. Text information is the area of our focus. In this research query expansion and its values with its critics has been reviewed and studied. These tasks are carried out to understand the problem area and scientifically support the results. By consulting the domain expert in Amharic language and phonetics, characteristics of the language have been studied and data selection is carried out accordingly. The data selected constitutes polysemous and synonym words and word variants or phrase of the Amharic language. Moreover, it is checked that the document corpora comprises the polysemous query terms selected for testing. Finally, a relevance judgment, stating which documents are relevant as per all testing queries are prepared.

II. Literatures and Related Works

One of the challenging areas in information retrieval is, evaluating IR system's performance [4]. Different evaluation mechanisms may be used to assess the level of IR system's success. According to [5], system's performance evaluation can be categorized in to six. The information users need might be factual or opinions.

These are engineering (hardware and software of the system), input (input data), output (output data), use and users, social level [5]. Though, among the listed evaluation mechanisms, the last two are found to be best, unpredictable nature of users' information seeking behavior and their level of cognitive power makes it unreliable [3]. Users do not want to spend much of their time in searching (i.e. query generation, query execution, scanning results of query to select items to read, reading non-relevant items) [3]. In addition, users mostly want to have control over the results, without the knowledge of how the systems reach there. Thus, in order to make IR systems match users' expectation, control should be over the kind of information retrieved, not which terms used to modify the query. Unfortunately it is often quite difficult to build interfaces, which are complex enough to behave in this manner [4]. The size of information on the web is exponentially increasing due to the fast development of front and back end applications. Information in factual form mostly processed by information retrieval system where as information in opinion form can be processed by opinion mining tools [10] [11]. In addition to this, there are systems that have been developed so far in order to enhance quality of the websites and back end applications [12] [13] [14].

Nowadays, the research frontiers in IR domain focuses on, the problem of making existing IR systems retrieve better results. Regarding this problem, different mechanisms have been introduced by many researches, in order to achieve a solution [1]. Some of these proposed techniques, as listed in [6], are: User-oriented search result organization; Incorporating user negative feedback; supporting effective browsing and Effective query reformulation. Among the listed mechanisms, effective query reformulation or query expansion based on semantic thesaurus has been found to be better [6].

The purpose of query expansion is basically, assisting users, to reformulate a more specific meaning baring query, so that best results can be obtained. However, findings show that, there are various reasons for bad users' query formulation [1]. The first one is ambiguity; since the user is only aware of a single meaning of the query term. The other is vocabulary mismatch; when the user is not aware of the right terminology, of his/her search domain. The last reason is lack of discrimination; since a user may not know any alternative term for his query reformulation, when he/she cannot get the desired information on the first attempt [1].

Till now, there are basically two approaches to analyze a corpus for query expansion. These are global and local automatic analysis [3]. Global automatic analysis looks in to the whole document corpora, to come up with expanding terms, while local analysis methods utilize documents that initially appear to match the query [7]. Therefore, global methods are neither interested in the query nor in the results returned from it prior to query expansion. Though unlike the local analysis, global automatic analysis involves rigorous and time taking task [3]. In this research local automatic analysis is favored than global, because it is less computational work load and results can be promising for augmented precision [4]. Local automatic analysis can also be referred as relevance feedback, because it considers documents retrieved using users' original query as relevant.

There are two relevance feedback mechanisms [7]: users' relevance feedback and pseudo relevance feedback methods. The former works by involving users to select documents relevant to their query, while the later assumes all of the initially retrieved K documents as relevant.

The main objective of this research is, to design bi-gram based query expansion as well as to integrate it to the existing Amharic IR system so that better combination of more relevant documents are retrieved as per the information need of the users.

III. Query Operation

As it is well known, users formulate their query without detailed knowledge of the document corpus and the retrieval environment. Thus, most of the time they are not satisfied with the results found at their first attempt. Therefore, there is a need for a better query that better express their information need, so that better set of relevant documents can be retrieved.

Improving the original query, hoping to retrieve more relevant documents is called query reformulation. This can be done either by expanding the original query with new terms or reweighting the original users' query terms [4]. Expanding a query means, adding terms which share similar meaning with the query terms. On the other hand reweighting involves, attaching various weights to query terms, so that documents carrying a query term with the biggest weight can be favored. In the latter case, a term with the highest weight is considered as a query meaning baring term. In this research, methods involving query expansion are investigated.

Relevance feedback is the most popular strategy used to select expanding terms for query reformulation purpose [4]. A process involving, the use of relevance feedback and query reformulation is called query operation. The two basic approaches to provide relevance feedback are user's relevance feedback and pseudo-relevance feedback [4].

User relevance feedback is the process of involving users in the retrieving process. A system that uses this approach needs judgment from users, so that a query with better meaning expression power can be formulated.

User relevance feedback involves a series of steps. First the user issues a query on which the system returns an initial set of documents. Among the retrieved documents, the user marks some of them as relevant. The system then computes a better representation of the query, based on users' relevancy judgment. Finally, the system uses the reformulated query to retrieve the revised set of documents, Manning et al. [7].

The objective for query expansion is, coming up with more relevant documents in retrieved document set, than the original set of documents retrieved using the original users query. If the original set of relevant documents in

the whole of the corpus is known, one can calculate the query vector that can fit the whole of the relevant documents in the corpus. Rocchio's algorithm [4] successfully distinguishes the relevant documents from the non-relevant.

Though, user's relevance feedback tends to reformulate users' queries as per their judgment, it can also be boring and time taking for them, which in turn degrades their interest of further searching. As a solution to this problem, pseudo-relevance feedback method was introduced, which doesn't need users' involvement. An IR system which uses pseudo-relevance feedback method automatically generates expanding terms. There exist two approaches that utilize pseudo-relevance feedback method called, local automatic analysis and global automatic analysis [4]. For relevance feedback, the former utilizes documents that are initially retrieved using users original query and the latter analyzes the whole document corpus [4]. In both cases, the process is completely automatic, such that users have no clue, whether their first query have been reformulated or not.

IV. AMBIGUITIES IN AMHARIC WRITING SYSTEM

Amharic writing system adopted all the symbols in Ge'ez and added 8 other symbols and the other 44 symbols. The result is that there is a considerable systemic redundancy of several consonant sounds which lacks in the phonology of Ge'ez [8]. Ambiguities in Amharic writing system arise mainly due to symbol redundancy [3]. Thus, 4 distinct sets of 7 can represent the sound /h/ + vowel: ("ሀ", "ሐ", "ኃ", "ኸ"), 2 sets represent /s/: ("ሰ", "ሠ") and 2 /s'/ ("ጸ", "ፀ") [8]. A similar problem is observed in usage of some letters interchangeably, such as "ቆ" vs "ቋ", "ኰ" vs "ኮ", and "ጎ" vs "ጐ" [3]. In addition to the symbolic redundancy of characters, Amharic writing system suffers slightly from visual similarity or different character, such as ፐ and ፕ, ፖ and ፓ, ዪ and ዩ, ጕ and ጐ [3].

Because of the different kinds of ambiguities in Amharic writing system, designing an IR system based on Amharic text is challenging. The class of symbols with the same sound falls into two [3]. The first class consists of symbols in the 1st order and the 4th order in a base symbol having the same sound during reading; for example, ሀ and ሃ, ሐ and ሓ, ኀ and ኃ, አ and ኣ, and ዐ and ዓ. The other class is a group of different alphabets that share the same sound, which are ሀ, ሐ, and ኀ, ሰ and ሠ, አ and ዓ, and ጸ and ፀ. This makes words with the same meaning to have different spelling structure. For instance, the same word "tsehay" ፀሐይ, can be written differently as ጸሐይ, ፀኃይ, ፀሀይ, ጸሃይ, etc [3]. Using these symbols interchangeably in words doesn't make reading, or forwarding ideas difficult for human beings. But unlike humans it is difficult for systems to consider them having the same meaning. Because IR systems only match the symbols in words to check weather a word from a document has the same meaning as in the query (i.e. if the words are a match then they are the same and have same meaning) encountering the different interchangeable symbols in words forces it to consider them as different (i.e. the system considers ፀሀይ and ጸሃይ, as different words with different meaning). The other challenge Amharic IR systems face is the combination of two words. There is no convention as to which words should be combined as one word or separately during writing. For example, the word "megneta bet" which means "bed room" can possibly be written as መኝታቤት (without space) and መኝታ ቤት (with space) and also the word "bete mekides" which means "temple" can be written as "ቤተመቅደስ" (without space) and "ቤተ መቅደስ" (with space) which makes it difficult for the IR system to differentiate between them [3]. These ambiguities degrade the performance of IR system.

V. BI-GRAM QUERY EXPANSION

This research is a perpetuation of the information retrieval system used by [3]. This information retrieval system is the base for the current research as it was for [3]. This information retrieval system doesn't use techniques like expanding and reformulating the original query. It only retrieves documents that have one or more query terms from the inverted index. Improving the results of this original Amharic IR system is one of the challenges of this research.

Indexing terms appear just side by side to the query term, as expansion terms. For example in a text "$A\ B\ C$" where $A\ B$ and C are terms let us say B is the query term, then A and C are taken as expansion terms according to this technique. The name bi-gram is given to this method because it only considers a query term and its immediate successor or predecessor [9]. This method is used for various other purposes mentioned on chapter three and on literatures such as [9]. In these different purposes, bi-gram is designed to consider only successors of a term. It means, for hand-writing recognition it is designed to predict what would the unread term be after a given term, or in spelling error detection it is designed to predict what would be the word misspelled coming after a given term [9]. There are also other relatives of the bi-gram method according to [9], given their names by the number of terms they consider. If it analyzes three terms it is given a name tri-gram or if it analyzes five terms penta-gram would be its name [9].

The idea behind the Bi-gram model is guessing a word coming after $N-1$ words. Guessing the next word or word prediction is an essential subtask of speech recognition, hand-write recognition, augmentative communication for the disabled, and spelling error detection [9]. N-gram uses probability coefficient to predict which word is appropriate to come after a given $N-1$ word sequence. The generic equation for probability of finding wn word after a sequence of ${w_1}^{n-1}$ words is given in equation 1 as given in [9]:

$$p(w_n/w_1^{n-1}) \qquad (1)$$

But it is not easy to compute the probability finding a word after a long sequence of preceding words [9].

Therefore, an approximation is devised which works by finding the probability of a word given a single previous word. This is called the Bi-gram model, where in the above equation, n becomes 2. Therefore, the revised equation for the Bi-gram model looks like the equation given on equation 2.

$$p(w_2/w_1^1) \quad (2)$$

This assumption, the probability of a word depends only on the previous word is called a Markov assumption [9]. The bi-gram method adopts the Markov Assumption and formulates a group of expansion terms, which has a high probability to be found next to the query term in the document corpus. Therefore the total expansion terms are extracted by returning all the terms from left and right of the query term, in the relevant retrieved documents or from the pseudo feedback.

It is unusual to find two query terms apart with a single term in the middle, which that term in the middle is the only term which can be considered as a common expansion term. Therefore top expansion terms for each query term in equal amount are taken as a total expansion terms for the whole query. Therefore, if k amount of terms are needed for expansion and n amount of query terms are given, k/n amount of terms are taken from each query term expansion terms. Given $q1$, $q2 \ldots qn$ query terms and documents $d1$, $d2 \ldots dm$ as pseudo-relevance feedback, Fig. 1 shows how Bi-gram based expansion terms selection works.

```
accept (q1…qn, D1…Dn)
for i = 0 to n do
   for j = 0 to m do
      top = Select(qi , Dj) using p(w2/w1^1)
final = Add(top)
return final
```

Fig.1. Pseudo code for Bi-gram based expansion terms selection

The function (qi,) in figure 1 selects top k/n terms to be added in the final total expansion terms. Expansion terms are selected by checking if they exist to the left or right of a query term among the top documents from the pseudo-relevance feedback. Top is a variable which holds the selected k/n terms for each query term. Then this top k expansion terms are returned for retrieving the re-ranked document collection using the Bi-gram model.

Finally equal amount of k/n expansion terms are selected independently from each query term's expansion term collection and the reformulation process ends.

Bi-gram is implemented in this research with slightly different approach; by letting it to consider a query terms immediate predecessor also. From its implementation, an assumption arises which states that a query terms immediate predecessor and successor can be a good expansion term. The overall bi-gram based query expansion architecture is shown in Fig. 2.

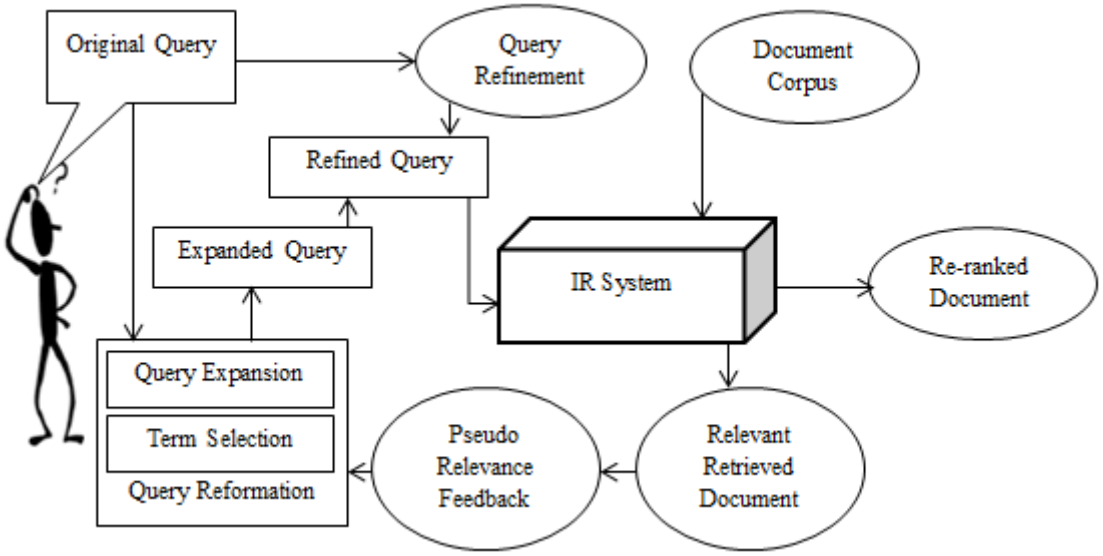

Fig.2. Bi-gram based query expansion architecture

VI. Experimentation

Any research is conducted to find some solution. The found result need to be evaluated to make sure that, prototype systems are performing well effectively and efficiently.

Efficiency is the amount of time and space a system uses to perform its task. On the other hand, effectiveness refers to the extent a system fulfills its objective. In the case of IR systems the amount of relevant documents retrieved determine its effectiveness. The effectiveness measurement technique used for IR systems is known as, retrieval performance evaluation. There are various retrieval effectiveness evaluation techniques. In this research recall, precision, and F-measure are used for the performance evaluation.

Recall and precision are the most common retrieval evaluation measurements used this days. Given a set of retrieved documents, a system can be evaluated based on systems centered evaluation approach as follows. Let the number of relevant documents as per the relevance judgment and as per the IR system is $|R|$ and $|A|$ respectively. The intersection of A and R, is a set of documents retrieved and relevant, let the number of these documents be $|Ra|$. Having these set of information, recall and precision can be calculated as follows.

Recall is the fraction of relevant documents retrieved by the system among the whole of the relevant documents, and is given by equation 3 as stated in [4]:

$$Recall = \frac{|Ra|}{|R|} \quad (3)$$

Precision is the fraction of relevant documents retrieved by the system among the whole of the documents retrieved, and is given by equation 4 as stated in [4]:

$$Precision = \frac{|Ra|}{|A|} \quad (4)$$

Though, recall and precision are both evaluation methods of a system, they measure two different aspects of the system and thus they are inversely proportional. If recall of a system is enhanced then the precision is minimized. That is because; there exists many non-relevant documents among the retrieved once, in an attempt to include many of the relevant documents. On the other hand, if precision of a system is improved then the recall is minimized. Because there are little amount of relevant retrieved documents among the whole relevant documents found in the corpus. An ideal system that scores both its recall and precision 100% retrieves all the relevant and only the relevant documents, which is difficult to achieve in reality.

F-measure is a harmonic mean evaluation measurement, which combines both recall and precision and is given by equation 5 [4].

$$F\,j = \frac{2*Re*Pr}{Re+Pr} \quad (5)$$

Where (j) and (j) are recall and precision at the j^{th} document respectively. The harmonic mean (j) assumes the value between [0, 1]. It is 0 when no relevant document is retrieved and 1 when all the first ranked documents are relevant. Further, the harmonic mean F assumes a high value only when both recall and precision are high.

This study attempts to design a generic relevance feedback to control polysemous words that affects the performance of a system. The relevance feedback is integrated to the Amharic IR system developed by [3]. A prototype has been built using Java NetBeans.

The testing is carried out on the Amharic Bible Old Testament taken as a document corpus. It contains 21,000 steamed terms and 930 documents. Since a systems centered testing procedure is carried out, test reference collection is prepared for the test queries. Nine queries with polysemous terms are formulated for testing.

The query refinement process is implemented to make the pseudo-relevance feedback as good as possible. Users are not satisfied in separating their query terms using ORs and ANDs and thus an automatic way of clustering query terms is necessary. Query terms separated in these logical operators contribute to broaden the information domain which the user intended. This process delivers a query with terms separated in logical operators.

The implementation of this process first should have a way to select query terms in a manner that any of the terms are not combined again (i.e. (q1 AND q2) and (q2 AND q1) must not be present in the refined query at the same time. This is because computational time and operational cost of the system can be wasted if the same kind of aspect is formulated twice. Fig.3 shows the way java written program handles separating two query terms without repeating the combination.

```
for(int i=0; i < queryTerms.length; i++)
    for(int j=i; j < queryTerms.length; j++)
        if(i != j) //checking out the two query terms are not the same
  String[] booleanResult = ResultReturner.foundDocsReturner(queryTerms[i],
  queryTerms[j]);
//Boolean result is documents in which query terms i and j exist.
```

Fig.3. Separating query terms and combining them back in pairs

The other task is to decide whether the combined query terms connect by AND or by OR. If the two query terms have same meaning then they connect by the logical operator OR, because one of the terms can represent the meanings of both query terms. And if the two query terms have different meaning they connect using AND, because neither of them can represent the meaning of both query terms. Fig. 4 shows how a java written program handles the metric clustering algorithm to calculate similarity between pairs of query terms.


```
for(int x=0; x<booleanResult.length; x++)
    int[] term1Position = TermPositionReturner(booleanResult[x],queryTerms[i]);
    int[] term2Position = TermPositionReturner(booleanResult[x],queryTerms[j]);
    for(int z=0; z<term1Position.length; z++)
       for(int y=0; y<term2Position.length; y++)
           if(term1Position[z] > term2Position[y])
               similarity = similarity + (term1Position[z] - term2Position[y]);
           else
               double devidend = ((term2Position[y] - term1Position[z])/5);
               similarity = similarity + (1/devidend);//metric similarity calculation
               similarity = similarity/(term1Position.length * term2Position.length);
return similarity;
```

Fig.4. Metric clustering similarity calculation and normalization

Fig. 5 shows how a java written program handles the decision that two query terms should be connected using the logical operator OR or AND.

```
if (similarity > 0.05)
    newQuery = queryTerms[i] + " OR " + queryTerms[j]
else
    newConstructedQuery = queryTerms[i] + " AND " +
queryTerms[j]
return newConstructedQuery;
```

Fig.5. Combining query terms with logical operator OR or AND.

The bi-gram based method outperformed the original query based retrieval, and scored 8% improvement in total F-measure, Fig. 7. And the result of bi-gram based query expansion technique is shown Fig. 6.

Query	REL	Using bi-gram method				
		RET	RETREL	R	P	F
የጌታዬ ባርያ	100	2	2	0.02	1.0	0.03
እጥፍ ሰጠው	26	2	2	2	1.0	0.14
ቀና እና ቅን ስራዎች	140	683	129	0.92	0.18	0.31
ሰው የተባለ አለቀ	10	32	7	0.7	0.21	0.32
ዘይት ቀባው	69	123	46	0.66	0.37	0.47
መጥፎ ስራውን ገሰጸ	18	5	5	0.27	1.0	0.5
አመድ እና ማቅ ለብሰ	35	61	28	0.78	0.45	0.58
አመታት ተቀመጠ	326	258	219	0.67	0.84	0.75
ሲጨልም አንቀላፉ	10	14	10	1	0.71	0.83
			Avg	0.67	0.57	0.61

Fig.6. Systems performance using bi-gram method

	Average Recall	Average Precision	Average F-measure
Original query	**0.83**	**0.39**	**0.53**
The Bi-gram method	**0.67**	**0.57**	**0.61**

Fig.7. Comparing IR System designed without and with bi-gram query expansion

VII. Conclusions

This paper studied the bi-gram technique for retrieving relevant document from large corpus. The bi-gram technique which uses the pseudo relevance feedback for expansion, selects those terms which appear to the right or left of a query term. Query expansion is vital in order to meet the users' information need. For expanding terms not to be polysemous, the original users query terms were taken as a guarantee to seek the intended meaning and vice versa. But results clearly showed that, both original users query terms and expanding terms can appear in different contexts, conveying completely different meanings. Therefore, there is a need for IR systems to adopt an approach to check weather expanding terms are polysemous or not. We will carry out more extensive techniques to improve the performance of Amharic Information Retrieval system.

References

[1] Wang, X. (2009), "Improving web search for difficult queries", Unpublished paper available at University of Illinois at Urbana-Champaign.

[2] Photoxels (2010), "Amount of digital information created in 2010 reach 1.2 Zettabytes", Canada, EMC Corporation.

[3] Alemayehu (2002), "Application of query expansion for Amharic information retrieval system", MSc Thesis, Addis Ababa University Ethiopia.

[4] Baeza-Yates R. and Ribeiro-Neto B. (1999), "Modern information retrieval" , 2nded, Addison-Wesley-Longman Publishers, England.

[5] Saracevic, T (1995), "Evaluation of evaluation in information retrieval", proceedings of the 18th annual international ACM SIGIR Conference on Research and development in information retrieval special issue of SIGIR Forum, pp. 138-146.

[6] Greenberg, J. (2001), "Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology", Journal of the American Society for information science and Technology, Vol. 52, No. 6 pp. 487-98.

[7] Manning, C. Raghavan, P. Schutze and H. (2009), "An introduction to information retrieval", Cambridge university press, England.

[8] Bloor, T. (1995), "The Ethiopic Writing System: A Profile", Journal of the Simplified Spelling Society, Vol.19, No. 2, pp. 30-36.

[9] Jurafsky, D and Martin J.H (2000), "Speech and Language Processing" 2nd ed, John Benjamins Publishing Company, Amsterdam.

[10] Tilahun, T. (2014), "Linguistic Localization of Opinion Mining from Amharic Blogs" International Journal of Information Technology & Computer Sciences Perspectives, 3(1), 890.

[11] Tilahun, T. and Sharma, D. (2015), "Design and Development of E-Governance Model for Service Quality Enhancement." Journal of Data Analysis and Information Processing, 3, 55-62. doi: 10.4236/jdaip.2015.33007.

[12] Balakumaran P.J, Vignesh Ramamoorthy. H,"Evolving An E-Governance System for Local Self-Government Institutions for Transparency and Accountability", IJIEEB, vol.5, no.6, pp.40-46, 2013. DOI: 10.5815/ijieeb.2013.06.05.

[13] Zakaria Itahriouan, Noura Aknin, Anouar Abtoy, Kamal Eddine El Kadiri,"Harnessing Social Networks Resources to Bring Social Interactions into Web-based IDEs", IJIEEB, vol.7, no.4, pp.24-30, 2015. DOI: 10.5815/ijieeb.2015.04.04.

[14] Shivani K. Purohit, Ashish K. Sharma,"Database Design for Data Mining Driven Forecasting Software Tool for Quality Function Deployment", IJIEEB, vol.7, no.4, pp.39-50, 2015. DOI: 10.5815/ijieeb.2015.04.06.

Authors' Profiles

Abey Bruck is a lecturer at the Department of Computer Science and IT, AMiT, Arba Minch University, Ethiopia. He completed his BSc degree in Computer Science from Arba Minch University and his MSc degree in Information Science from Addis Ababa University. His research interests include information retrieval, artificial intelligence, artificial neural network and algorithm analysis.

Tulu Tilahun is a lecturer at the Department of Computer Science and IT, AMiT, Arba Minch University, Ethiopia. He completed his BSc degree in Computer Science from Mekelle University and his MSc degree in Computer Science from Addis Ababa University, Ethiopia. His research interests include data mining, big data analysis, cloud computing, green computing, data modeling and software engineering.