

Afan Oromo News Text Summarizer

Girma Debele Dinegde
Pohang University of Science and Technology
Department of Computer Science and Engineering
Pohang, Korea

Martha Yifiru Tachbelie
School of Information Science
Addis Ababa University
Addis Ababa, Ethiopia

ABSTRACT

Information overload is a global problem that requires solution. Automatic text summarizer, a computer program that summarizes a text, is one of the natural language processing technologies that have got researchers focus to help information users. In this study, three methods have been used for the development of Afan Oromo news text summarizers and the resulting three summarizers have been evaluated both objectively and subjectively. These are : S1 that uses term frequency and position methods without Afan Oromo stemmer and language specific lexicons (synonyms and abbreviations); S2 is a summarizer with combination of term frequency and position methods with Afan Oromo stemmer and language specific lexicons and S3 is with improved position method and term frequency as well as the stemmer and language specific lexicons .

The result of objective evaluation shows that S3 outperformed the two summarizers (S1 and S2) by 47% and 34 %. Moreover, the subjective evaluation result also shows that S3 better than the other summarizers (S1 and S2) with informativeness, linguistic quality, and coherence and structure.

Keywords

Document summarization, extractive summary, single document summarization, News text summarization, Afan Oromo

1. INTRODUCTION

These days, the amount of documents in paper and electronic format is growing dramatically. As a result, users (readers) are facing information overload problem. In almost all languages in the world, texts in any domain are written in detail and readers are forced to see unwanted detail without being interested in it.

Afan Oromo text readers are not exceptional to suffer from this problem. There are many domain areas that produce large content of textual information which needs summarization to save the time of readers. Some of the textual information are large volumes of legal judgments which is very essential if they are used by the experts (for timely justice) and by law students for their study, newspaper texts and online news articles produced by media agencies, criminal investigation document produced by polices at different level, reports from government offices, etc.

Currently, newspapers and other news releases in the language reach the readers from many sources. There are a number of media agencies and presses releasing news in electronic and non-digital format.

With the absence of automatic text summarization services that can potentially reduce the readers' browsing and reading time, it can be said that readers have been and being spending more time than they should browsing over the content that they are not interested in.

Currently, a few - researches in automatic text summarization have been commenced for Ethiopian languages, particularly for Amharic text in different domains by adopting different techniques. The present work is a contribution towards developing natural language processing applications for Ethiopian Languages. Specifically it increases the scope of the text summarization research by investigating its application for Afan Oromo language. The techniques used in this study is term frequency and sentence position methods with language specific lexicons (synonyms and abbreviations) to assign weights to the sentences to be extracted for the summary.

2. OVERVIEW OF AFAN OROMO

Afan Oromo is one of the major African languages that is widely spoken and used in most parts of Ethiopia and some parts of other neighbor countries like Kenya and Somalia [1] and [3]. It is used by Oromo people, who are the largest ethnic group in Ethiopia, which amounts to 34.5% of the total population. Besides first language speakers, a number of members of other ethnicities who are in contact with the Oromos speak it as a second language, for example, the Omotic-speaking Bambassi and the Nilo-Saharan-speaking Kwama in northwestern Oromia [13]. Currently, Afan Oromo is an official language of Oromia regional state (which is the largest Regional State among the current Federal States in Ethiopia). Being the official language, it has been used as medium of instruction for primary and junior secondary schools of the region. Moreover, the language is offered as a subject from grade one to grade twelve throughout the schools of the region. Few literature works, a number of newspapers, magazines, educational resources, official credentials and religious documents are published and available in the language.

3. RELATED WORKS

The research work on text summarization can be traced back to 1950's when the first extractive system developed by [11]. He proposed that words appearing many times in a text furnish good idea about the content of the document though there are words that appear very frequently but not content bearing. As a result, he tried to cut off these words by determining a fixed threshold. The idea of Luhn was acknowledged and used in many automatic information processing systems. It is a domain specific single document summarization system that summarizes technical articles. The system uses features like term filtering and word frequency (low-frequency terms are removed). Sentences are weighted by the significant terms they contained and sentence

segmentation and extraction is performed. [2] Expanded the work of Luhn. He carefully outlined the human extracting principles and noticed that the location of a sentence in a text gives some clue about the importance of the sentence. Thus, he suggested word frequency, cue phrases, title and heading words and sentence location as an extraction feature. Like the work of Luhn, Edmundson's system is a single document and domain specific (that deals with technical articles). Moreover, the output of the system is an extract summary.

GreekSum[4] is an automatic text summarizer for the Greek language. It is built based on the algorithms developed and used for the SweSum [5], text summarizer for Swedish. According to Pachantouris [4] several changes needed to be made to support the differences of the Greek language from the Swedish. A version of SweSum which is language independent called Generic (without Greek keyword dictionary) and the customized version of the summarizer for Greek language called GreekSum were compared. Subjective evaluation was carried out where they found that using the Greek keyword dictionary in GreekSum brought performance improvement (16 percent improvement compared to the system that did not use a dictionary).

FarsiSum [6] is an attempt to create an automatic text summarization system for Persian language. It is a web-based text summarizer for Persian based upon SweSum. It summarizes Persian newspaper text/HTML in Unicode format. FarsiSum uses the same structure used by SweSum [5], with the exception of the **lexicons**, but some modifications have been made in SweSum in order to support Persian texts in Unicode format. It uses a very simple stop-list in order to filter and identify the important keywords in the text.

Currently, a few - researches in automatic text summarization have been commenced for Ethiopian languages, particularly for Amharic text in different domains by adopting different techniques.

The present work is a contribution towards developing natural language processing applications for Ethiopian Languages. Specifically it increases the scope of the text summarization research by investigating its application for Afan Oromo language. The techniques used in this study is term frequency and sentence position methods with language specific lexicons (synonyms and abbreviations) to assign weights to the sentences to be extracted for the summary.

4. METHODOLOGY

Primarily, literatures related to automatic text summarization have been reviewed. As the study is conducted on Afan Oromo news text summarization, the nature of the language and the structure of the documents to be summarized for testing were investigated. To carry out this task, books, journal articles, and relevant websites are consulted.

4.1 Corpus Preparation

A corpus to evaluate the summarizer (Afan Oromo news articles) was selected and prepared as there is no previous research and corpora in Afan Oromo for evaluating summarizers. The prepared corpus consists of 8 news items from different sources. While selecting from news archives, longer articles (at least one page or more than 200 words) are considered due to the fact that as the text itself gets shorter summarizing it becomes unnecessary.

The average length of news items, in the corpus, is approximately 277 words or 11 sentences as shown in Table 1.

Table 1: Prepared corpuses for the study

Text ID	News size in words	News size in sentences
Test 1	250	11
Test 2	403	14
Test 3	250	9
Test 4	231	10
Test 5	290	13
Test 6	295	14
Test 7	232	11
Test 8	269	13
Average	277.5	11.875

4.2 Reference Summary Preparation

For the purpose of manual summary generation, the corpus was provided to four experts together with a guideline prepared by the researchers.. The experts ranked the sentences based on their ability of providing salient information. For a sentence, an average rank was calculated as the sum of its four ranks divided by four. The sentences have then been ordered according to their average rank. Finally, reference summaries were produced from the top ranking sentences at the compression rate of 10 %, 20%, 30 % and 40% of the original text's length.

4.3 Summarization technique

Most research on summary generation techniques still relies on extraction of important sentences from the original document to form a summary [7]. There are several ways in which one can characterize different approaches to text summarization.

The technique proposed for this study is extraction technique for single news text. Using extraction technique most important sentences from the document are extracted and displayed to the reader. To create a summary by this technique there is no need of rewriting the document by making linguistics analysis. To extract important sentence from a text to be summarized, sentence can be weighted based on cue phrases it contains, location of the sentence, sentence containing most frequent words in the document. Then sentences with the highest weight obtained by efficient combination of extraction features will be selected and a summary is written.

4.4 Evaluation technique

Reference (gold standard) summary has been prepared and used to evaluate the performance of the system output (system summary). The evaluation process was conducted using an intrinsic method. It comprised of both subjective (qualitative) and objective (quantitative) evaluation methods. For both measures the four experts are involved.

Subjective evaluation was used to measure the linguistic quality, informativeness and coherence of the automatically generated summaries. The linguistic quality is basically aimed to measure the readability and fluency of the summary. Subjective evaluation techniques has been adapted by Greek text summarizer [9]. On the other hand, objective evaluation was basically used to measure the summarizer's performance in identification and extraction of salient sentences. This performance is measured by the standard recall and precision measures. Given an input text, human's (reference) summary and summarizer's extract, it measures how close the extracts are to the reference summary.

5. IMPLEMENTATION AND EVALUATION

The customized summarizer is named Open Oromo Text Summarizer (OOTS), the version based upon OTS which summarizes Afan Oromo news texts. The basic principles of OOTS are the same with OTS, but adjustments have been made in order to support Afan Oromo language. Every modification considering the specific rule of the language has been done by creating XML file: oro.xml by modifying the English dictionary: en.xml.

The English mode of XML file has been modified and configured the rules of Afan Oromo lexicons. The adjustments made to the original OTS, to support Afan Oromo news text summarization, are changing the rule of stemming as well as compiling and integrating stop word list, synonyms and abbreviations. In general, for this study the most of the work done in adjusting the OTS code so that it can make use of the Afan Oromo lexicon and actually work for the Afan Oromo language.

5.1 Language resources required for the OOTS

To customize OTS so as to support Afan Oromo text summarization, some language resources were required: Afan Oromo stop-word list, Afan Oromo abbreviation list and list of synonyms as well as the rules for stemming. We found all the components required by the original OTS system for supporting Afan Oromo language even if all resources are not complete.

Afan Oromo stop-word list

Stop-word list are a list of words that should not be stemmed by the stemmer as they are non-content bearing words. Commonly, stop-word list consists of prepositions, conjunctions, articles and particles. The stop-word list compiled by [10] has been used. Besides, stop-words found in the book entitled: “A Grammatical sketch of written Oromo” by [11] has been added to enhance term frequency method as Debel’s [10] stop word list is not complete. The total number of stop-words reached 124 that are still incomplete. Randomly selected sample stop-words are shown in Table 2.

Table 2: Sample Afan Oromo Stop-word

Word	Meaning
Ammo	<i>however, but</i>
Garuu	<i>But</i>
Bira	<i>beside, at, near of</i>
Ala	<i>outside, out</i>
Akka	<i>such as, like, according to</i>

Afan Oromo abbreviations

The aim of tokenization is to split the text into sentences, a seemingly trivial task, but which can be complicated by the fact that punctuation marks also serve other purposes, for example, they are used in abbreviations. A language-dependent list of abbreviations is therefore used to prevent false detection of sentence boundaries. Common abbreviations available in different literature (grade 9 to 12 Afan Oromo student text books) has been compiled. Some samples of abbreviations with full meaning are shown in table 3.

Table 3: Sample Afan Oromo abbreviations

Abbreviations	Full meaning
k.k.f	Kan kana fakkaatan
w.k.f	Waan kana fakkaatan
Fkn.	Fakkeenyaaf
Hub.	Hubachiisaa

Afan Oromo synonyms

Even if term frequency method is very important to text summarization, it alone is not enough to produce a good quality summary [2]. It has been criticized for the reason that there may be more than one word to express the same thing which is termed as synonyms. With synonyms one concept can be expressed by different words. For example waangoo ‘fox’ and jeedala ‘fox’ refer to same kind of animal.

A list of available Afan Oromo synonyms has been prepared from Afan Oromo dictionary entitled, “Galmee jechoota Afaan Oromoo” are configured to oro.xml file to enhance the term frequency based method. Table 4 below contains some of Afan Oromo synonyms.

Table 4: Sample synonyms words

Term	Synonymy	Meaning
Tolchuu	Gochuu	Make
Dhibamuu	Dhukkubsachuu	Sick
Qooduu	Hiruu	Share
Jijjiiruu	Diddiiruu	Change
Herreguu	Yaaduu	Think

Afan Oromo Stemmer

In our work, we have used lightweight stemmer rules for Afan Oromo that strips the suffixes using a predefined suffix list using the algorithm developed by [10]. This system takes as input a word and removes its suffixes according to a rule based algorithm. The algorithm follows the known Porter algorithm for the English language and it is developed according to the grammatical rules of the Afan Oromo. According to [10] an evaluation of the system showed the algorithm’s accuracy is 96 percent. Therefore, for our system, lists of affixes integrated to: oro.xml file to apply the rule of stemming to our OOTS similar to the Porter’s stemmer used by OTS.

5.2 Summarization process and techniques used

The adopted summarization method is sentence extraction based. It has three major steps: (i) preprocessing, (ii) sentence ranking and (iii) summary generation.

Preprocessing

As is in other ATS systems, preprocessing step includes tokenizing, stop-word removal, stemming and parsing (breaking the input document in to a collection of sentences).

Furthermore, using stemmer, a word is split into its stem and affix. Affixes striped can be replaced by another affix or replaced by white space as per the rule it matches with. The design of a stemmer is language specific, and requires some significant linguistic expertise in the language. A typical simple stemmer algorithm involves removing suffixes using a list of frequent suffixes, while a more complex one would use

morphological knowledge to derive a stem from the words. Since Afan Oromo is a highly inflectional language, stemming is necessary while computing frequency of a term.

Sentence Ranking

After an input document is formatted and stemmed, the document is broken into a collection of sentences and the sentences are ranked based on two important features: term frequency (TF) and sentence position.

TF is frequency of keyword appearance in an article. This method is the earliest known method to be used for automatic text summarization since research began in this area. It is based on the idea that the most relevant sentences are those containing the largest number of the most frequent words in the document (stop-words excluded) [11].

With the *tf* (term frequency) method, the importance value (score) of a sentence *s*(IVs) is given by:

$$IVs = \sum tf$$

Where, IV is Importance Value based on term frequency

tf, is Term frequency

On the other hand, positional value (score) of a sentence *s* is computed in such a way that the first sentence of a document gets the highest score and the last sentence gets the lowest score in news domain as the original OTS uses constant multiplicative factor of term frequency score calculated. The positional value for the sentence *s* is computed using the following formula by combining two parameters for sentence ranking. Therefore, the total importance value (score) of a given sentence *s* (TIVs)

$$TIVs = IVs * c$$

Where, *c* is constant multiplicative factor. The value of *c* is 2 for first statement of first paragraph, 1.6 for first sentences of all other paragraphs. All other sentences are weighed only by their term frequency score.

TIVs, is total score of importance value of a sentence based on term frequency and position value.

Summary Generation

A summary is produced after ranking the sentences based on their scores and selecting N-top ranked sentences, where the value of N is set by the user. To increase the readability of the summary, the sentences in the summary are reordered based on their appearances in the original text; for instance, the sentence which occurs first in the original text will appear first in the summary.

6. EXPERIMENTS AND EVALUATION

For each news item, three experiments have been conducted with different methods (S1, S2 and S3). All the summaries extracted by each experiment at a given extraction rate are compared against one reference summary. The three experiments with different methods are:

S1: using original English language mode of the summarizer with term frequency and sentence position methods but, without Afan Oromo stemmer and other language specific lexicons (synonyms, stop-word list and abbreviations). In this case, the original English version summarizer has been used to summarize news items in Afan Oromo given the English language version summarizer, as the rules for stemming, synonyms and abbreviations do not match for Afan Oromo; it can be considered as a summarizer for Afan Oromo without

stemmer and other language specific lexicons (synonyms, stop-word list and abbreviations).

S2: using term frequency and position method with Afan Oromo stemmer and other language specific lexicons. In this case, the plan is to test the performance of the summarizer with the term frequency and position methods. To enhance the effect of term frequency method on selecting informative sentences for extraction and consequently improve the performance of the summarizer, rule of stemming and language specific lexicons of Afan Oromo has been included. This mode of the summarizer directly access Afan Oromo language specific dictionaries developed with file name "oro.xml" (consisting of Afan Oromo stemming rule, synonyms, stop-word list, and abbreviations).

S3: using combinations of term frequency with improved position method with stemmer and other lexicons (synonyms and abbreviations). In this case, the plan is to test the performance the summarizer by assigning weight for sentences according to their importance. As it has been discussed in Section 4.3.2, sentence position method is the constant multiplicative factor of the term frequency score to grade (rank) sentences. In order to produce good summary of news items the position method should be given emphasis based on the inverted pyramid structure of the news text. Hence, the position method has been improved by doubling the value of the constant multiplicative factor so as to increase its effect on the total score of a sentence which is given by the following formula:

$$TIVs = \sum tf * c$$

Where TIV, is total importance value of sentence *s*

tf, term frequency of content bearing terms

c, constant multiplicative factor . In this case, the value of *c* is 4 for the first sentence of first paragraph and 3.2 for every first sentence of other paragraphs instead of the commonly used values (2 for the first sentence of the first paragraph and 1.6 for the first sentences of other paragraphs). Therefore, for each news article discussed under Section 4.1, three summaries (using the three methods: S1, S2 and S3) at a given extraction rate has been generated by the summarizers. The following section discusses the performances of the summarizers.

6.1 Evaluation and Discussion of Results

One of the usual and challenging tasks to be carried out in any research is evaluation and discussion of the result. For this study, the summarizers are evaluated using objective and subjective methods. Both subjective and objective evaluation methods used are intrinsic to the summary.

6.1.1 Subjective evaluation

Subjective evaluation technique has been adapted which was used by GreekSum [9], a text summarizer for Greek language in order to evaluate the summaries generated by our customized summarizers (OOTS). The eight Afan Oromo news items, of various contents, used in the objective evaluation have also been used for this evaluation. In our evaluation process three different system summaries (generated by the three different methods discussed in Section 7 have been evaluated by experts. The three system summaries created with different methods are compared according to the following three check points.

- In which summary the most important information is kept?

The experts (evaluators) check the summary's informativeness. Specifically, they check whether the summary includes best sentences that contain the most important information about the topic and satisfies information need of readers or not.

- Out of a scale from 1-5, where 5 is the best, what score would you assign to each summary?

This checks the linguistic quality. It includes the assessment of grammar, non-redundancy and referential clarity.

- Which summary is more coherent?

The evaluators check whether the summary has smooth transition of sentences. While reading the sentences in their rank order, it should not just be a heap of related information, but also should build a coherent body of information.

Results of subjective evaluation and discussion

This section presents results of the subjective evaluation based on the three points and interpretation of the results.

The informativeness of the summary created by one of the three methods for each test item is scaled to 100 out of the expected total 4 votes by the four evaluators. For instance, if S1 is selected by two of the evaluators, the percentage of the informativeness of the summary is measured as $2/4 = 0.5$ i.e. 50%. Table 5 depicts the percentages and average performance of the three methods as it is judged by evaluators.

Table 5: Information preserved analysis result

Text ID	System Summary with different methods		
	S1	S2	S3
Test 1	25%	25%	75%
Test 2	25%	25%	100%
Test 3	25%	50%	25%
Test 4	50%	50%	50%
Test 5	50%	50%	50%
Test 6	25%	25%	75%
Test 7	50%	25%	50%
Test 8	25 %	50%	75%
Average	34.37 %	37.5 %	62.5 %

6.1.2 Objective evaluation

Objective evaluation is one of the evaluation methods employed for this study to measure effectiveness of the summarizers. The objective evaluation assumes one and only one best (reference) summary and compares the system summary against the reference summary. It is intended to measure the system's summary approximation to the reference summary on the basis of standard recall (R), precision (P) and F-measure (F).

The standard recall and precision measures are calculated as follows and f-measure is calculated based on the values of precision and recall:

$$Recall(R) = correct / (correct + missed)$$

$$Precision(P) = correct / (correct + wrong)$$

$$F - measure(F) = 2 * R * P / (R + P)$$

Where:

-Correct = the number of sentences in both the summarizer's summary and the reference summary,

-Wrong = the number of sentences in the summarizer's summary but not in the reference summary,

-Missed = the number of sentences in the reference summary but not in the summarizer's summary.

Since OTS does not have an evaluation tool, a tool used to compute the standard recall, precision and f-measures has been developed and integrated with it

Summaries are required to be generated at four compression rates 10%, 20%, 30% and 40%. From a total of 8 news articles prepared for experimentation, a pair of news items randomly selected to be input to the system for a given extraction rate. For instance, the first two news items extracted at compression rate of 10% and the second two news items at 20% compression rate, etc. the system summaries of all news items are compared against one best reference summary created by expert summarizers.

Results of objective evaluation and discussion

The results of the three summary generation methods (S1, S2 and S3) have been compared against the reference summary and the standard precision and recall as well as F-measure has been computed.

For the sake of discussing evaluation result, F-measure has been used as it is the weighted average of precision and recall. Table 6 summarizes the evaluation results. As can be seen from the table, f-measure is undefined for the first two test sets: Test1 and Test2 for which the summaries are generated using S1 and S2 at a compression rate of 10%. This is because the reference summary and system summary have no common sentences.

Surprisingly, even if the average performance by principle increases with an increase in extraction rate, it did not hold for this experimentation result as it can be observed, better or similar result can be attained at 20% than 30% or 40% extraction rates. This can be due to the limitation of reference summary created by human experts which is highly subjective as there is no correct 'reference summary'.

Table 6: Objective summary evaluation result

Text ID	Compression rate in %	S1			S2			S3		
		P	R	F	P	R	F	P	R	F
Test 1	10	0%	0%	-	0%	0%	-	100%	100%	100%
Test 2	10	0%	0%	-	0%	0%	-	100%	100%	100%
Test 3	20	50%	50%	50%	100%	100%	100%	100%	100%	100%
Test 4	20	50%	50%	50%	50%	50%	50%	50%	50%	50%
Test 5	30	25%	25%	25%	50%	50%	50%	50%	50%	50%
Test 6	30	25%	25%	25%	50%	50%	50%	75%	75%	75%
Test 7	40	50%	50%	50%	50%	50%	50%	75%	75%	75%
Test 8	40	75%	75%	75%	75%	75%	75%	100%	100%	100%
Average score		34%	34%	34%	47%	47%	47%	81%	81%	81%

Comparison of objective and subjective evaluation results

Both subjective and objective evaluation has been conducted for a summary extracted by three different methods. The results of both evaluations support each other; in all cases the summary using S3 outperformed the other methods.

For the three methods (S1, S2 and S3), the average informativeness of the summary is (34.37%, 37.5% and 62.5%); the average language quality is (59.37%, 60% and 65%) and the average coherence and structure is (21.87%, 28.12% and 75%) respectively. On the other hand, the objective evaluation result shows that the average f-measure score for (S1, S2 and S3) is (34%, 47% and 81%).

7. CONCLUSIONS AND FUTURE WORKS

This study dealt with the development and evaluation of the first automatic text summarizer for Afan Oromo news text, named OOTS (Open Oromo Text Summarizer) which is based upon the Open Text Summarizer (OTS). OTS has been customized by modifying the code so that it can support Afan Oromo language. Both subjective and objective evaluations were carried out where we found that using the combination of improved position and term frequency methods showed a promising result. Based on the findings and knowledge acquired from literature the following recommendations are forwarded.

-Being the first tool for this language, further task is required to make it freely accessibly on the Internet for everyone to use.

-A balanced well prepared corpus is essential for further evaluation of the performances of the summarizers. One can work in the development of gold standard summaries.

-Complete stop-word list, synonyms, and abbreviations are very useful to enhance term frequency based method. The development of such resources is important.

-Besides this implementation, more languages can be included in the OTS system especially Latin-based Ethiopian languages can use the OOTS system as a framework to develop a summarizer. The algorithm to be used is basically similar and can be easily adjusted to serve the needs of different languages.

-Like other extraction based summarizers, the result of this study's summarizers lack coherence. More advanced method that would help avoid such problem is abstract summarization. Therefore, future works can concentrate towards the development of abstract summarization.

8. ACKNOWLEDGEMENTS

We are indebted to the people who helped us during this study especially in evaluating the summarizers.

9. REFERENCES

- [1] Abera N. (1988), "Long vowels in Afan Oromo: A generic approach", Master's thesis, School of graduate studies, Addis Ababa University, Ethiopia.
- [2] Edmondson H. P.(1969), "New Methods in Automatic Extracting," Computing, vol. 16, no. 2, pp. 264-285.
- [3] Grage G. & Kumsa T.(1982), "Oromo dictionary", African studies center. Michigan State University.
- [4] Pachantouris G. and DalianisH.(2005), "GreekSum A Greek Text Summarizer," Word Journal of the International Linguistic Association, pp. 1-45.
- [5] Dalianis, H., M. Hassel, J. Wedekind, D. Haltrup, K. de Smedt and T.C. Lech. (2003), "Automatic text summarization for the Scandinavian languages." In Holmboe, H.ed.) Nordisk Sprogteknologi .
- [6] Hassel M.(1999) "FarsiSum - A Persian text summarizer," Cognitive Science, pp. 2-4.
- [7] Kaili M. & Pilleriin M. (2005), "ESTSUM - Estonian newspaper texts summarizer", Proceedings of The Second Baltic Conference on Human Language Technologies Pp. 311-316
- [8] Rotem N.(2001), " Open Text Summarizer" . Available at: <http://libots.sourceforge.net/>
- [9] Pachantouris G. and DalianisH.(2005), "GreekSum A Greek Text Summarizer," Word Journal of the International Linguistic Association, pp. 1-45.
- [10] Debela T.(2010), "Designing a Stemmer for Afan Oromo Text: A hybrid approach", Master's thesis, School of graduate studies, Addis Ababa University, Ethiopia.
- [11] LuhnH. P. (1958), "The Automatic Creation of Literature Abstracts", pp. 159-165.