

Improvement of An Abstractive Summarization Evaluation Tool using Lexical-Semantic Relations and Weighted Syntax Tags in Farsi Language

Ahmad Estiri, Mohsen Kahani, Hadi Ghaemi, Mohsen Abasi

Web Technology Laboratory
Ferdowsi University of Mashhad
Mashhad, Iran

ahmad.estiri@stu.um.ac.ir, kahani@um.ac.ir, hadi.qaemi@stu.um.ac.ir, abasi.mohsen@stu.um.ac.ir

Abstract— in recent years, high increase in the amount of published web elements and the need to store, classify, restore, and process them have intensified the importance of natural language processing and its related tools such as automatic summarizers and machine translators. In this paper, a novel approach for evaluating automatic abstractive summarization system is proposed which can also be used in the other Natural Language Processing and Information Retrieval Applications. By comparing auto-abstracts (abstracts created by machine) with human abstracts (ideal abstracts created by human), the metrics introduced in the proposed tool can automatically measure the quality of auto-abstracts. Evidently, we can't semantically compare texts of abstractive summaries by comparison of just their words' appearance. So it is necessary to use a lexical database such as WordNet. We use FerdowsNet with a proper idea for Farsi language and it notably improves the evaluation results. This tool has been assessed by linguistic experts. This tool contains metric for determining the quality of summaries automatically by comparing them with summaries generated by humans (Ideal summaries). Evidently, we can't semantically compare texts of abstractive summaries by comparison of just their words' appearance and it is necessary to use a lexical database. We use this database with a proper idea together with Farsi parser in order to identify groups forming sentences and the results of evaluation improve significantly.

Keywords-Farsi Natural Language Processing (NLP); Semantics; Evaluation; Automatic Abstractive Summarizer; Sentences groups; Parse tree; parser;

I. INTRODUCTION

Today with the increasing volume of information and published data on the web, access and reading the required information in the shortest possible time are constantly a challenge that researchers face.

Automatic summarization of a document is the generation of, by a machine, a briefer version of document in which features and main points are maintained [1]. Because of the high importance of summarizers and extension in the proposed methods, precise and accurate evaluation of these methods is very important.

There are two approaches in the assessment of summarization systems: human judgment and comparison with

a reference summary. In order to assess a summary produced by a machine (machine-summary) we can compare it with the summaries generated by humans (human-summary). To reduce the impact of personal tastes and opinions in human-summaries, each machine-summary is compared with multiple human-summaries of the same text; and the final score can be the average of all scores or the maximum calculated one.

Before comparing summaries, for standardization purpose, some pre-processing tasks must be done. Naturally the more powerful pre-processing modules, the more reliable results of texts' comparison we have. It should be noted that since Farsi is an unstructured language, it has more complexities and difficulties.

In the comparison process of summarized texts, a stemmer is used for stemming each word in the text and then, in the evaluation process, these stems are what which are being compared with each other. But obviously, for evaluating abstractive summaries, the mere comparison of words' appearance with each other, won't lead to desirable results. Because in abstractive summaries words are changing and summaries may be reproduced. So comparison process is different from when we have selected/chosen summaries. For this reason, using a lexical WordNet is necessary.

In recent years, word networks are used in various applications. These networks are used in knowledge based and semantic intelligent systems generally in the role of ontologies or calculation words. In reality, WordNet is a network including thousands of concepts linked together by semantic relationships. Each concept represents an abstract set of elements constitute a group based on their common properties. Any of these groups is composed of one or more words as they are a specific concept and constituent words of the group can be used instead of each other in a text. A group may be linked to other groups through some semantic relations. Semantic relations between family groups are different in according to the group's type (noun, verb, adjective, or adverb).

It is notable that the number of researches done on implementing Farsi automatic summarization systems is very small. Among automatic summarizers for Farsi language, there is a tool called FarsiSum [2] that is a modified version of

SweSum [3] developed and used for Swedish texts. This system is applied for news summarization and allocates more weight to the first sentence and first paragraph. There is also another way for summarization of Farsi texts which works in single-document and sentences-selection. In selecting output sentences, this method uses hybrid of lexical chain and graph theory methods and final output can be in general or in according to the user's query [4]. Among multi-document summarization techniques, we can point to [5] by which generated outputs are abstractive. In [6] after reviewing challenges in processing of Farsi texts, types of methods in summarization have been reviewed. In [7] a sample tool for Farsi texts summarization by a novel extraction method has been developed. This tool is used for multi document summarization of Farsi texts; its output is by selection and applies a clustering based method. Based on singular value decomposition and hierarchical clustering, Honarpisheh and his colleagues [8] have developed a multi-document multi-lingual text summarizer. For any language, "a word segmentation system" and "a dictionary of words in conjunction with their document frequencies" are just two main resources their summarizer relies on. In the first place, their approach gets a collection of related documents as inputs and transforms them into a matrix; this resulted matrix is the input of singular value decomposition module. Using a binary hierarchical clustering algorithm, it then selects the most important sentences of the most important clusters to create the summary. Poormasoomi and his co-authors in [9] propose a novel way of multi document summarization using a new form of extraction of semantic relations of the text, Latent Semantic Analysis (LSA), and Semantic Role Labeling (SRL) of words.

The next sections in this paper are as follows: In the second section, we have a glance on the previous related works and researches and in third one the suggested approach, for evaluating Farsi abstractive summaries is proposed. Evaluation result is in section four. Finally conclusions and future works are represented.

II. RELATED WORK

In this section, we will introduce some of tools and methods used for evaluation of summarizers in the other languages [10]. Proposes SEE: an evaluation package. In this environment, evaluators can test quality of a summary according to a reference summary. Texts used in evaluation process are preprocessed by splitting into a list of its parts (phrases, sentences, and ...). Another tool used in summaries evaluation is MEADeval. MEADeval is a tool for evaluating summaries extracted in formats DUC and MEAD that do its evaluation by comparing the summary to a reference summary. MEADeval [11], [12] is sufficient for summarizers that generate extracts (summaries created by extraction of important sentences from the text). ISI ROUGE introduced in [13] and [14] as a package of automatic summarizers of evaluations. ISI ROUGE package, later known as ROUGE1, is an attempt to automate evaluation summaries. This tool, that is

implemented for English language, is similar to evaluation tool BLUE [11] that is used in Machine Translation and is based on calling common n-tuples between machine summaries and reference ones. At present ROUGE is the most famous and widely used evaluation tool. However the main problem of it is that the units involved in the evaluation process and their similarity as the similarity of texts, studied are not necessarily meaningful and sometimes meaningless and not very important combinations are evaluated.

In the following some of proposed methods for evaluation of summarizers will be introduced. In the other method, called as pyramid approach, one person sets range of summary content units (SCUs) in reference text and another one or two adopt summaries to find the compliance [15], [16]. One another way used for summarizers' evaluation in year 2006 was application of basic elements (BEs) for comparing summaries [17]. In this regard each reference sentence is divided into a series of minimal semantic units named as basic elements and then compliance between these basic elements in two texts under evaluation, is examined. In [18] and [19] Using search engine and ranking of documents and their summaries based on the specified query and Compliance Evaluation ranking of documents and their summaries, an effort was made to develop fully automated methods for the evaluation summaries.

In [20] a tool for evaluating Farsi summaries, as the first sample, has been proposed that introduces new metrics for evaluating summaries and determining similarity of summarized texts. In [21] following previous work [20], we also tried to improve the procedure of summaries evaluation by using the FerdowsNet lexical database and by considering semantic in words' comparison.

In this paper, to improve the results of the comparison of the summaries in the assessment procedure of summary instruments, firstly, by the parser designed for Persian language, groups forming the sentences of summarized texts identified, then those groups are weighed, and then the similarity between words in the same groups is calculated.

III. THE PROPOSED METHOD

Given that the human summaries are often abstractive, most of automated text summarization systems are moving toward generating abstractive summaries to be consistent with the reality of human summaries. This exposes existence necessity of one standard metric for comparing abstracts in terms of compression rate and amount of retaining information. Besides it, in automated text summarization, there may exists several good summaries for a specific source text so that evaluating these summaries by comparing to a fixed and immutable reference summary, is not satisfactory. Also considering that different compression rates are appropriate for different types of texts, evaluation methods that give possibility of different rates evaluation should be considered. Therefore a comprehensive and standard way for covering more aspects of the summaries evaluation must be applied.

The Works done for Farsi abstractive summaries evaluation are divided into several major phases. At first, all characters of

¹Recall-Oriented Understudy for Gisting Evaluation

the text must be homogenized by replacing with its equivalent standard. In the processing of Farsi calligraphy, considering similarity it has to Arabic calligraphy, there is always a problem in some of the characters including special ones such as “ک”, “ی”, “هزه”, “ء”. The first step should be to solve problems related to these types of characters. Additionally there are another tasks among necessary actions before the evaluation phase: reforming and integration of “half space” (there is a special character called “نیمفاصله” in Farsi which is typed with keys Ctrl+Shift+4 OR Shift + Spacebar for typing words such as “می‌شود”) and space in its different applications and also removing character “-” that is used for stretching sticky characters and similar tasks done for integration of texts. After pre-processing of texts, three tools Sentences Detector, Words Recognizer and Stemmer must be respectively imposed on the prepared texts. Sentences Detector identifies sentence boundaries using the following signs: “.”, “,”, “!”, “?”, “#”, “\$”, “%”, “&”, “’”, “:”, and some Farsi grammatical rules. Words Recognizer uses below symbols to detect each word: empty space, “_”, “_”, “_”, “_”, etc. Stemmer also finds stems of the recognized words of the previous step. Removing stop words list even before the evaluation acts is possible. Stop list contains high frequency and low importance words which must be removed from texts before evaluation. In Farsi language this list should be prepared with great care. Removing the words of this list from Farsi texts improves evaluation results in most cases. Automatic determining the quality of summaries is done by comparing them with the ones generated by humans. The metric used in the proposed method calculate the structural and semantic similarity between the system summary (summary generated by machine) and human summary in number of words, n-tuples and pair words.

In abstractive summarization evaluation, calculating the semantic similarity of words is the most important and key part of evaluation. For calculating the semantic similarity of words, utilization of a database is essential. Following, the methods identify groups forming sentences, lending weight to them, and then metric used to calculate the extent of semantic similarity of texts are described.

A. The architecture of the proposed method

After pre-processing on the summarized texts, parser tool designed for Persian language [21] is used. The Parser, from vocabulary structure, position and word order in sentences, characters or phrases before and after them, and word type, forms the Parsing Tree of sentences such that groups forming sentences of summarized texts detected. Now the extent of similarity between the existing words in similar parts of system summarization and machine summarization using the following metrics calculated and weighed average of the extent of the similarity of sections containing different groups according to their importance will be calculated.

Averaging from the scores of five texts Obtained from the separation of the groups identified by the text parser

Between groups detected by the parser, nominal and verbal groups are always more important than any other groups. The groups are weighted as follows:

$$weight_{Group_type} = \begin{cases} 1 & type = Noun\ or\ Verb \\ 0.9 & type = Prepositional \\ 0.8 & type = Adjective \\ 0.7 & type = Adverb \end{cases}$$

Thus, the average scores given to the five sections resulting from the summarized text separation to its constituent groups calculated according to the following formula:

$$S_{Final} = \frac{S_{Noun} + S_{Verb} + 0.9(S_{Prepositional}) + 0.8(S_{Adjective}) + 0.7(S_{Adverb})}{4.4}$$

Metrics that used to calculate the similarity of each of the text parts, are as below.

B. Metric of word similarities

In this Metric, the extent of similarity between each part of the system summaries with its same part in each of human summaries calculated and divided on the total number of words in that part of the human summaries.

As the simplest metric for measuring closeness of two texts, syntactic and semantic similarity between their words can be noted. In this metric, words correlation between system summary and each human summary, is calculated and then divided by the total number of words in human summaries. The formula is presented in the following:

$$Score = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{Word} Score_{similarity}(Word)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{Word} Count(Word)} \quad (1)$$

In this equation $Score_{similarity}(Word)$ is maximum similarity between words in summaries generated by machine and by human. As well, $Count(Word)$ expression represents the number of words existed in reference summaries.

C. Metric of n-tuples similarities

This measurement considers semantic similarity of n-tuples in system summary and a set of human summaries. In this metric, which is shown in formula (2), n-tuples similarity between system summary and human summaries is divided by the total number of n-tuples existed in human ones.

$$Score = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n} Score_{similarity}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n} Count(gram_n)} \quad (2)$$

In this equation n in the expression $gram_n$ is length of the n-tuple and $Score_{similarity}(gram_n)$ is the maximum similarity between n-tuples of system generated summary and reference summaries. By the help of studies conducted, if evaluated n-tuples are constructed from only the words in a sentence, not whole text, accuracy of evaluation will be improved. As well, instead of considering the similarity of just n-tuples existed in two texts, if all existed i-tuples ($i < n$) are formed and the extent of their similarity is assessed, the evaluation will be performed with more and more higher accuracy. It is worth noting that more reference summaries in calculation of this

metric, more n-tuples in denominator of above formula, so more accuracy of evaluation we will have.

D. Evaluation metric of word pairs with free distance in text

Each word pair (respectively) in a sentence is called Skip-bigram. This metric measures semantic similarity of Skip-bigrams existed in system summary and reference ones. For example suppose the following sentences:

- A. پلیس آن دزد را کشت. (The police killed that thief.)
- B. پلیس آن دزد را می‌کشد. (The police kills that thief.)
- C. دزد پلیس را می‌کشد. (The thief kills the police.)

Each sentence has $C(5,2) = 10$ Skip-bigrams. As a sample for sentence A. we have:

A = ("آن دزد", "پلیس کشت", "پلیس را", "پلیس دزد", "پلیس آن", "را کشت", "دزد کشت", "دزد را", "آن کشت", "آن را")

This metric (see formula (3)) is calculated by computation of semantic similarity between word pairs in system summary and reference ones.

$$\text{Score} = \frac{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{Skip-bigram}} \text{Score}_{\text{similarity}}(\text{Skip-bigram})}{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{\text{Skip-bigram}} \text{Count}(\text{Skip-bigram})} \quad (3)$$

In this equation $\text{Score}_{\text{similarity}}(\text{Skip-bigram})$ is maximum computed similarity between Skip-bigrams of the system summary and reference summaries. Also $\text{Count}(\text{Skip-bigram})$ shows number of Skip-bigrams existed in reference summaries.

In this equation $\text{Score}_{\text{similarity}}(\text{Skip-bigram})$ is maximum computed similarity between Skip-bigrams of the system summary and reference summaries. Also $\text{Count}(\text{Skip-bigram})$ shows number of Skip-bigrams existed in reference summaries.

E. How to calculate similarity of two words

In all the above metrics, the last step is comparing stems of two Farsi words with each other in order to calculate amount of their similarity. As it was already mentioned, this step is the most important and key part of the evaluation process. For calculating similarity stems of two Farsi words, we directly correlate them with each other. We do it by using FerdowsNet. If they are completely similar, the result will be 1, else it will be between 0 and 1. As we noted above, FerdowsNet is a network containing thousands of concepts, called Synsets, which are linked together by semantic relations. Among these relations in FerdowsNet, we can name the following ones (most of considered links are about Nouns and there are another relation types about Adjectives, Verbs, and etc.):

- Synonym: e.g. correlate and compare
- Antonym: e.g. good and bad

Hypernym: if X is a more specific type of Y, then Y is Hypernym of X. e.g. person is a hypernym of rancher

Hyponym: if Y is Hypernym of X, then X is Hyponym of Y. e.g. Niagara Falls is a hyponym for the concept of waterfall.

Holonym: if X is a part of Y, then Y is a Holonym of X. e.g. body is holonym of brain.

Meronym: if X is a part of Y, then X is Meronym of Y. e.g. heart is meronym of body.

Similarity amount of two words is computed by considering their positions in words network and the relationship between them using Lin metric [23] (see formula 4). The gained value is always between 0 and 1 and this formula is used for calculating similarity amount of two words in above metrics.

$$\text{Sim}_L(c_1, c_2) = \frac{2 \log p(\text{lso}(c_1, c_2))}{\log p(c_1) + \log p(c_2)} \quad (4)$$

$\text{LSO}(C_1, C_2)$ Points to the lowest super-ordinate (most specific common subsumer) of two concepts of C_1 and C_2 .

In this Metric, in order to enhance the quality of relationships between concepts, the $p(c)$ function introduced which indicates the possibility of hit with a sample from c concept or the concept itself in a large text configuration.

1- EVALUATION

For a proper and thorough assessment of the proposed method for the Persian Summarization Evaluation Tool, a set of standard and proper data is needed. Thus, the Pasokh [24] corpus specification of single-document summarization was used to evaluate the proposed method. We can see the corpus specification in the following table. (Table 1)

TABLE I. THE PASOKH CORPUS SPECIFICATION OF SINGLE-DOCUMENT SUMMARIZATION

Title	Number
Number of the corpus documents	100
Number of Topic	6
Number of News agencies	7
Number of extractive summarization for each document	5
Number of abstractive summarization for each document	5
Percent of Summarization	30

Now, the implemented instrument will give a score to machine summary with comparing the machine summary and five human summaries or reference from the above corpus and considering the mentioned metric in the proposed part. Considered metric are such that the score of instrument will always between one and zero.

After that, ten other people are used to evaluate the tool. So that each person in each subject scores to the extent of overlap of machine summary and also the reference summaries used in tool. This human scoring process for system summaries simplified by designing software for it.

After that, we average the scores given to the machine summaries comparing five reference summaries. Finally, by taking the average of the scores given by each of the ten people involved in the evaluation process, the accuracy will increase.

It is obvious that an appropriate assessment tool must give high scores and low scores to the good and bad summaries respectively. The results of each of the proposed metric comes below.

Then for each summary of corps, computed Pearson product-moment correlation coefficient between the average of the scores given by tool and the average of the scores given by human.

In the first approach, only used the WordNet and parser used to identifying the groups of sentences and then weighting the identified groups that not used.

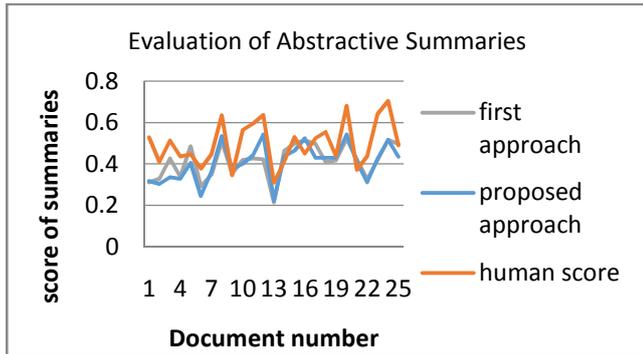


Fig. 1. Evaluation Metric of word similarities

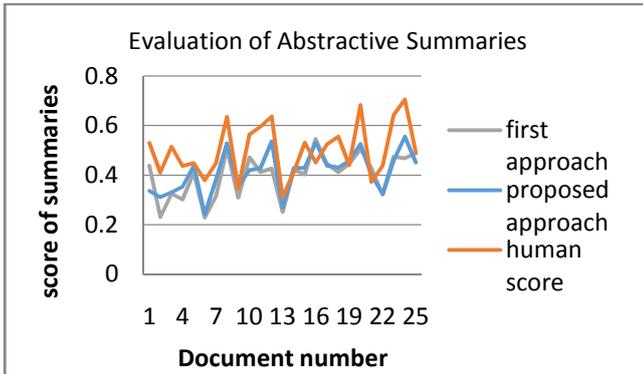


Fig. 2. Evaluation Metric of n-tuples similarities

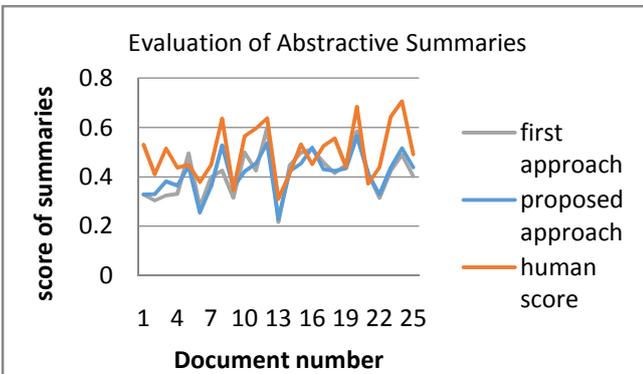


Fig. 3. Evaluation Metric of word pairs with free distance in text

TABLE II. RESULT OF APPROACHES

Metric	First approach	Proposed approach
Metric of word similarities	0.5963	0.6976
Metric of n-tuples similarities	0.6347	0.7164
metric of word pairs with free distance in text	0.6510	0.7425

F. Conclusion and future works

Nowadays due to the sheer volume of contents and lack of time, and also the ineffectiveness of existing summarizers, existence of a powerful system to summarize large volumes of books, articles and news on the Web is so important. As we see, despite this reality that the summarization field has been raised since year 1960, there are still many cons. in this area and much work must to be done to achieve the ideal situation in this field. The issue of summarization and its evaluation in Farsi language is far more than other languages. Existing linguistic complexities as well as the absence of tools with precision sufficient for working in Farsi are such researching problems in this language.

However, the proposed approach for evaluating Farsi summarizers, which is applicable in evaluation of both abstractive and extractive summaries, with the help of the prepared words network, named FerdowsNet introduced above, is usable with an acceptable precision. This tool has been not only implemented to evaluate the summarizers, but it also could be used to assess the similarity of two texts. Among applications of evaluating the similarity of two texts, evaluating machine translators or automatic correction of descriptive answers to examinations or similar situations can be named.

Despite of its appropriate performance and precision, the designed tool considers sense just in level of word comparisons and is not able to identify and recognize the semantic in a combination of words or sentences. If we could recognize subject and concept of any numbers of words, a nominal group or a combination from several sentences, we would be able extract semantic in a higher level and improve results of evaluating the similarity of two texts.

Moreover, taking the place of words in the groups forming summarized texts by weighing them may improve the results.

ACKNOWLEDGEMENT

In this part, we should thank the efforts of Excellency Dr.Jahangiri and staffs of Web Technology Laboratory of Ferdowsi University of Mashhad, specially Mr.Pourmasoumi, Mr.Asgarian and dear Mr.Saedi.

REFERENCES

- [1] I. Mani and M. T. Maybury, "Advances in automatic text summarization," vol. 293: MIT Press, 1999.
- [2] N. Mazdak, "FarsiSum-a Persian text summarizer," Master thesis, Department of Linguistics, Stockholm University,(PDF), 2004.
- [3] H. Dalianis, "SweSum-Atext summarizer for Swedish No.," TRITA-NA-P0015). Stockholm, Sweden: NADA, KTH2000.

- [4] Z. Karimi and M. Shams Fard, "automatic summarization system of Farsi texts," (in Farsi) in International Conference of Computer Society of Iran Tehran, 2007.
- [5] A. S. Shahabi, "summarization of Farsi language texts," (in Farsi) in Conference on Cognitive Sciences, Tehran, p. 56, 2003.
- [6] M. Moshki, "review of Farsi unstructured texts summarization approaches," (in Farsi) Ms, Computer department of engineering faculty, university of Science and Technology 2008.
- [7] M. Moshki and M. Analuyi, "Multi Document Farsi News Summarization for generating background-news extract," (in Farsi) presented at the Iranian Conference on Electrical Engineering (ICEE), Tehran, 2009.
- [8] M. A. Honarpisheh, G. Ghassem-Sani, and G. Mirroshandel, "A Multi-Document Multi-Lingual Automatic Summarization System," pp. 733-738, 2007.
- [9] A. Poormasoomi, M. Kahani, S. V. Yazdi, and H. Kamyar, "Context-Based Persian Multi-document Summarization (Global View)," Asian Language Processing (IALP), pp. 145-149, 2011.
- [10] M. Hassel, "Evaluation of Automatic Text Summarization: A Practical Implementation," Licentiate thesis, Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden, 2004.
- [11] C. Y. Lin and E. Hovy, "Manual and automatic evaluation of summaries," Proceedings of the ACL-02 Workshop on Automatic Summarization, p.45-51, 2002.
- [12] D. Radev, T. Allison, S. Blair-Goldensohn, J. Blitzer, A. Celebi, S. Dimitrov, E. Drabek, A. Hakim, W. Lam, and D. Liu, "MEAD-a platform for multidocument multilingual text summarization," in Conference on Language Resources and Evaluation (LREC), Lisbon, Portugal, 2004.
- [13] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries," In Proceedings of the Workshop on Text Summarization Branches Out, post-conference workshop of ACL 2004, Barcelona, Spain, 2004.
- [14] C. Y. Lin, "Looking for a few good metrics: Automatic summarization evaluation—how many samples are enough," in Proceedings of the NTCIR Workshop 4, pp. 1-10, 2004.
- [15] A. Nenkova and R. Passonneau, "Evaluating content selection in summarization: The pyramid method," In Proceedings of HLT/NAACL, 2004.
- [16] H. Saggion, "Multilingual multidocument summarization tools and evaluation," In: Proceedings of LREC 2006, Genoa, Italy. ELDA, Paris, France, 2006.
- [17] E. Hovy, C. Y. Lin, L. Zhou, and J. Fukumoto, "Automated summarization evaluation with basic elements," In: Proceedings of the 5th international conference on language resources and evaluation (LREC), pp. 899-902, 2006.
- [18] A. Louis and A. Nenkova, "Automatically evaluating content selection in summarization without human models," Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1, pp. 306-314, 2009.
- [19] D. R. Radev, S. Teufel, H. Saggion, W. Lam, J. Blitzer, H. Qi, A. Celebi, D. Liu, and E. Drabek, "Evaluation challenges in large-scale document summarization," Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, Sapporo, Japan pp. 375-382, 2003.
- [20] A. Estiry, M. Kahani, Z. Keyvanloo Shahrestanaki, and A. Poormasoomi, "presenting an automated evaluation tool of machine Farsi summarizers," (in Farsi) presented at the The 4th Conference on Information and Knowledge Technology (IKT2012), 2012.
- [21] A. Estiri, M. kahani, A. Pourmasomi and M. abbasi , "An Abstractive Summarization Evaluation Tool For Persian Text using word-net," (in Farsi) The First International Conference on Persian Language Processing, 2012.
- [22] A. Estiri, M. kahani, R. saeedi and E.Asgarian, "Designing Persian Language Parser Tool," (in Farsi) The First International Conference on Persian Language Processing, 2012.
- [23] D. Lin, "An information-theoretic definition of similarity," Proceedings of the Fifteenth International Conference on Machine Learning, pp. 296-304, 1998.
- [24] B. Behmadi Moghaddas, M. Kahani, S. A. Toosi, A. Pourmasoumi, A. Estiri, "Pasokh: A standard corpus for the evaluation of Persian text summarizers," Computer and Knowledge Engineering (ICCKE), 3th International eConference Mashhad Iran, pp. 471-475, 2013.