British Journal of Mathematics & Computer Science
19(6): 1-15, 2016; Article no.BJMCS.27549
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Achieving More Coherent Summaries in Automatic Text Summarization; an Ontology-based Approach

Majid Ramezani[1*] and Mohammad-Reza Feizi-Derakhshi[2]

[1]Department of Computer Engineering, University College of Nabi Akram, Tabriz, Iran.
[2]Department of Computer Engineering, University of Tabriz, Tabriz, Iran.

Authors' contributions

This work was carried out in collaboration between both authors. Author MR designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author MRFD managed the analyses of the study. Finally both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2016/27549
Editor(s):
(1) Alexander A. Katz, Department of Mathematics & Computer Science, St. John's University, Staten Island, NY 10301, USA.
(2) Qiankun Song, Department of Mathematics, Chongqing Jiaotong University, China.
(3) Doina Bein, Applied Research Laboratory, The Pennsylvania State University, USA.
(4) Tian-Xiao He, Department of Mathematics and Computer Science, Illinois Wesleyan University, USA.
Reviewers:
(1) R. Nedunchelian, Sri Venkateswara College of Engineering, Anna University, India.
(2) S. Sridhar, R V College of Engineering, Bangalore, India.
(3) R. Jayashree, PES University, Bangalore, India.
(4) Arti Arya, Visvesvarya Technologivcal University, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/16891

Received: 7th June 2016
Accepted: 2nd November 2016
Published: 11th November 2016

Original Research Article

Abstract

Since the growth of information available on the internet has grown out of hand, automation of text summarization process has become more important. Summaries that are in the form of a condensed version of original text, containing its important information, and considered as a good alternative to reading the original text. One of the main requirements of the machine produced texts, is their coherence and the semantic relation between their sentences. Reading a non-coherent summary, in spite of that does not help the readers become aware of the information in the original text, it also creates confusions in their mind. The main purpose of this paper is how to achieve more coherent summaries in automatic text summarization process. For this purpose there has been a system designed, that using the concepts of ontology, automatically summarizes Persian documents, and tries to produce more coherent summaries. In this light, a technique has been devised that based on it, presence of a sentence in the summary, increases the probability of choosing its adjacent sentences. The FarsNet ontology is the basis for

**Corresponding author: E-mail: sir.ramezani@gmail.com;*

ontology-based calculations in this paper. The results show that the suggested approach succeeds in producing coherent summaries.

Keywords: Automatic text summarization; ontology; coherency.

1 Introduction

Text summarization is a reductive transformation of source text to summary text through content selection and/or generalization on what is important in the source [1]. Nowadays with the exponential growth of the information available through internet, in other words, the information overload [2], the volume of the information is much more than the need of people. Certainly, studying this massive volume of information would be tedious and time consuming. Meanwhile, without studying a document, one cannot understand its contents and decide about its relevance with the topic. Because of this, having an alternative approach that can help reader achieve the main points without reading the whole text, so he can decide whether or not to read the whole text, would be really beneficial.

Different categorizations, based on different points of view, have been suggested for automatic text summarization systems ([3-6] for more study). Based on their output, summarization systems are categorized as either '*extractive'* or '*abstractive'*. In extractive systems the output contains the most important parts (sentences, paragraphs, etc.) of the original text, without making any changes to them [4]; while in abstractive systems, the summarization is done through understanding the original text, then retelling it using fewer words [7]. Also based on their output, text summarization systems are categorized as '*generic'* or '*query-based'*. Output of the generic summarization systems contains the most important points of the given document, while in query-based systems, the summary only contains the concepts that are closely related to the query [8-9].

Perhaps the rules for correct composition of paragraphs that in 19th century Alexander Bain have suggested in his English Composition and Rhetoric (1866), is the beginning of taking coherence in text seriously [10]. This concept, which garnered much more attention later, has two '*local*' and '*global*' kinds [11]. The local coherence is semantic relationship between each pair of consecutive sentences in text (sentence to sentence transactions); while the global coherence is the semantic relation that exists between the all sentences. Of course the local coherence is a necessary prerequisite for global coherence, to the extent that McKoon and Ratsliff say psycholinguisticaly it is the most important source for inference-making during the study/listening [12]. Hence, 'local coherence' (hereafter 'coherence') has been subject to more attention in computational linguistics.

Lack of coherence in text or speech will most likely prevent the message from being delivered to the reader or listener [10]. That is why coherence of the output text is one of the key aspects of any text producing systems [13].

Distribution of entities in locally coherent texts, displays certain regularities [13]. Therefore ontology as a knowledgebase of entities and their semantic relations, can be a suitable basis for achieving this regularity and consequently, coherence. "Ontology" is a Greek word consisting of "*onto*" meaning "beings" and "*logos*" which usually is interpreted as "science", therefore it can be said that ontology is science or study of beings [14]. In other words ontology is a systematic account of existence [15] that is used for modeling the beings in real world and the relations between them.

Usually, based on the nature of query-based text summarization (only sentences closely related to the query are put in the summary), their output have more coherence than summaries produced using generic text summarization systems (which put sentences related to all the topics made in the text, in the summary). In other words, in generic summarization, because of the variety of topics there is less semantic relation between sentences. The goal of this paper is to devise and implementation of an extractive and generic text summarization system for Persian documents that produces more coherent summaries. Therefore, without any queries, system will omit less important topics and produce a summary consisting of the most relevant sentences to the main topic discussed in the original document, which has a high level of semantic relation between its sentences. In many of the applications, without any queries, there is need to get a summary of the most important/central topic of the document, in a way that sentences have semantic relation. For this purpose a system will be designed that ontology is the basis for forming the summaries (ontology-based automatic text summarization), also the concepts extracted from it will be used for producing more coherent summaries. In this regard, the second part is dedicated to studying the related works done in this field, and in the third part the suggested automatic text summarization system and its approach toward achieving coherent summaries is discussed. Then in the fourth part, discussing the evaluation methods of automatic text summarization systems, the results will be evaluated, and finally in the fifth part, conclusion of this research will be presented.

2 Related Works

Numerous works about coherence has been done in the field of computational linguistics [16-20]. Our focus would be on the summarization systems that produce coherent summaries. First automatic text summarization system was built in 1958 [21], which produced summary using a set of statistical features. Coherence was not of much importance in the initial works done in this field. During the years, there has been so much work done in the automatic summarization field, that we can say this branch of Artificial Intelligence has reached maturity [22]. Nevertheless, attention to coherence in the texts produced using automatic summarization systems is relatively new. In the following, we will name and discuss some of the studies done in this field.

In [23], McKeown and Radev studied the ways for producing coherent summaries from many news documents about a common event. They used a system for interpreting original documents and to achieve a semantic representation of them.

Lin & Hovy presented the NeATS summarization system [24] which is a multi-document summarization system (in which the summary contains important portions of more than one documents as input) for extracting relevant portions of a set of documents, and presenting them as a coherent summary. Also QueSTS is an extractive multi-document summarization system that produces summaries based on the queries made by user [25]. In this system an intermediate representation of contextual relationship among sentences in a graph topology from all the input documents is produced. Existing sub-graphs are ranked using scoring models, and most relevant of them to the query is selected to be in the summary. This system uses a sentence ordering strategy to improve coherence in the results. Also in [26], for addressing the incoherence issue in extractive multi-document text summarization, based on a schema-based summarization approach, a schema for modeling discourse structure that is usually used by humans for summarizing, has been suggested. In this query-based approach, each sentence represents part of a schema in the discourse structure. G-FLOW is another extractive multi-document summarization system that creates an intermediate graph-based representation of the input text and by using it, estimates the discourse relations between sentences and produces a coherent summary [27].

Miller in [28] presents an extractive summarization system based on the Latent Semantic Analysis (LSA) technique for producing coherent summaries. LSA is an unsupervised technique that its results are cognitively like human performance [29]. In this system an initial extract is produced that is merely based on topic sentences, then the holes between the sentences that do not have semantic relation is filled with creating relations between them. Also In [30] an unsupervised probabilistic approach for modeling latent concepts in texts and correlation between these texts for producing topically coherent summaries has been suggested.

In [31] using Hidden Markov Models (HMMs) an approach for accurately detecting and extracting coherent relevant passages has been suggested. In this research HMMs has been introduced as the best approach for extracting passages. Also in [32] a summarization system consisting of three parts for generating coherent summaries with textual aspects has been suggested; first part is detecting sentence-level textual aspects, second part is coherence modeling using HMMs, and finally the third part is selection and arranging sentences according to textual aspects for achieving coherent summaries.

Yoo, Hu &Song in [33] suggest a coherent graph-based semantic clustering and summarization approach for biomedical literature. In this research, Text Semantic Interaction Network (TSIN) is built, and for finding the core semantic relations, an ontology-enriched graphical representation of the documents is used, then the summarization and clustering is done.

In [34], Parsumist is introduced as an automatic summarization system for Persian documents that relies on lexical chains and semantic features for summarizing texts. This system that is both single-document and multi-document, utilizes synonym sets to increase the coherence of results. Also in [35] with reliance on linguistics properties of text and semantic chains among them, a system has been suggested for producing coherent summaries of Persian documents.

Zhang in his PhD thesis [36] stressed upon the importance of coherence in automatically produced summaries and presented three different approaches for improving coherence in automatic summarization; shallow content-driven coherence in which words, phrases, sentences, discourse units and their literal relation are taken into consideration for increasing coherence; deep content-driven coherence which is determined based on the news aspect and speech acts; and finally, cognitive model-driven coherence which determined using cognitive models.

In [37], increasing coherence in single document summarization is treated as an optimization problem for choosing the best sentences of input based on the given objective functions. The researchers also have been used sentence compression techniques to reduce the length of sentences and increase their informativeness.

Discourse has generally been defined as anything beyond sentences [38]. It can be said that when use a language for any purpose, we create a discourse. By creating a discourse, speaker/writer wants to convey their thoughts, ideas and feelings to other people. As Linguistic Society of America defines it, *discourse analysis* is defined as the analysis of language 'beyond the sentence'. Therefore it can be deduced that discourse analysis means analyzing text to understand the goals of language user, for creating a coherent and unified part of the language. Also coherence relations are fundamental issue in the study of discourse [39-41]. As it can be seen in recent researches, despite the different approaches, all of them intersect at trying to produce coherent texts, and consequently achieving a stronger discourse structure. In this paper we are going to use ontology knowledgebase to understand the semantic hierarchies formed in mind of the writer, and by using that summarize the documents in a more coherent fashion.

3 Automatic Text Summarization System

The summarization system designed in this research is extractive and generic and does the summarization for Persian documents. As mentioned earlier, ontology is the basis for producing summaries in this system. Also ontology will be used for achieving more coherent summaries. For this purpose FarsNet [42] ontology, which is a Persian version of WordNet [43], will be used. This summarization system, in a graph-based work space, selects most important and coherent sentences as the text units to put in the summary. The architecture of this system consists of two *preprocessing* and *processing* phases, which after introducing FarsNet, we will detail them out.

3.1 FarsNet

FarsNet is a version of WordNet ontology [43] as a large electronic lexical database of English, which is designed by Natural Language Processing lab of Shahid Beheshti University, for Persian language. WordNet is designed by Princeton University and contains names, adjectives, verbs and adverbs which have been categorized in sets of cognitive synonyms (*synset*). There are many versions of WordNet designed for different languages such as German, French, Spanish, Dutch, and Italian. First version of Persian WordNet [44] contains lexical, syntactic and semantic knowledge for more than 15,000 Persian words and phrases, which are placed in 1,000 synsets of nouns, adjectives and verbs. Synsets (synonym sets) are sets of words that are cognitively synonymous, and they have formed the most fundamental relation between words in this lexical database. The second version of this lexical database [42] contains more than 30,000 entries in categories of noun, verb, adverb and adjective, and also more than 20,000 synsets. The relationships between senses and synsets in FarsNet are like the relationships existing in WordNet; relationships such as synonymy, hypernymy, hyponymy, meronymy and antonymy. Of course there is a notable difference between the size of FarsNet and the size of WordNet, which contains 155,287 lexical entries and 117,659 synsets.

3.2 System architecture

Architecture of the system suggested in this paper has two main phases for automatic summarization: *preprocessing* phase, in which all the activities needed for extracting information and detecting the semantic relationships in text are done, and *processing* phase in which using the information obtained from the previous phase, the most important and coherent sentences are selected to put in the summary. Each of these phases consist of some steps that we will discuss them in detail.

3.2.1 Preprocessing

In this phase the basis for producing summary is created in form of some steps. Fig. 1 shows the steps in this phase. As it can be seen, the text intended to be summarized is the input and graph representation of it is the output. The steps in this phase are respectively:

A. **Tokenization:** In this step, which is the first step of the system architecture, input text which is a stream of text is parsed into tokens (words, signs and other meaningful elements).
B. **Sentence Boundary Detection:** In this step, sentence boundaries (start and finish) as the text unit are detected. For this purpose, the punctuation marks that come at the end of Persian sentences {., ! and ?} are used.

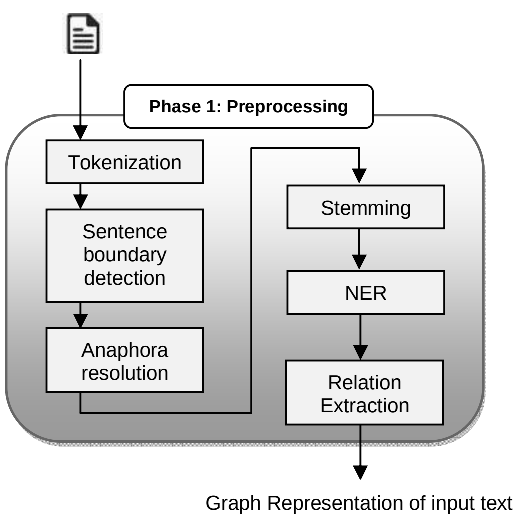

Fig. 1. Preprocessing phase

C. Anaphora Resolution: *Anaphora* is the language phenomenon of referring to an entity (object or an event) that has been mentioned before and *anaphora resolution* in the process of detecting these entities [45]. Imagine a document that its subject is about Grahame Bell, the inventor of telephone. It is possible that *his* name has been mentioned only once at the beginning of the text, and after that pronouns (such as "his") or referential phrases (such as "this inventor") are used to refer to him. Because of that sentences are taken as considered text units, these references will certainly break. So sentences must have semantic independence as much as possible; if not, these pronouns can affect the results when matching elements of each sentence (words and phrases) with ontology entries. Therefore in each high level task of natural language processing, anaphora resolution is performed as a necessary step. In Persian language, pronoun, as a syntactic word, only refers to person and number and is divided to seven types of *personal*, *adjoining personal*, *demonstrative*, *reflexive/emphatic*, *reciprocal*, *question* and *indefinite* pronouns.

Unfortunately, because there aren't any modules for replacing pronouns and referential phrases with corresponding references in Persian language, we used a simple and rudimentary solution. For this purpose we only replaced personal and demonstrative pronouns and refused to replace other types of pronouns and referential phrases. For better understanding, look at this example:

[Grahame Bell] is the [first inventor of practical telephone]. [He] was one of the founding members of the National Geographic Society. [The inventor of telephone], [who] was born in Edinburgh, Scotland, was educated at the University of London.

By identifying personal pronouns {I, you, he/she, we, you and they} and also name of people in Persian language, personal pronouns are detected and are replaced with the latest person name that has been mentioned in the previous sentence. We used Persian Wikipedia for identifying names of people. In the above example, [He] will be replaced with [Grahame Bell]. For the adjoining type of personal pronouns that come with another word; like "کتابم" (my book), "کتابت" (your book), "کتابش" (his/her book), based on the stemming done in the next step, the adjoining pronoun is removed from the word and only the stem, i.e. "کتاب" (book) remains. Also for demonstrative pronouns "این" (this) and "آن" (that), replacement is done with the first name which is mentioned before (nearest name before pronoun). For the remaining types of pronouns such as reflexive/emphatic pronouns (myself, himself), question pronouns (which, whom), indefinite pronouns (any, none) and reciprocal pronouns (each other), no replacement is done.

However, because the system has an ontology-based method, not replacing the referential phrases with their respective references, is justified. For example, in the example above, [Grahame Bell] is a sample of "inventor" and is specifically "inventor of telephone", so it is semantically related to [The inventor of telephone] and their entities are related and adjacent in the graph representation (which is created in the last step of preprocessing phase). Therefore most types of referring to Grahame Bell by referential phrases are somehow identified using semantic network existing in ontology, and their relationships are maintained.

Without doubt the approach of this paper for anaphora resolution has some flaws; but as said before, because there is no module for this purpose and because of the importance of anaphora resolution, this rudimentary approach is chosen. However, results show that using this approach improves the summarization process. Obviously using more comprehensive and reliable approaches for anaphora resolution will improve the results further.

D. Stemming: Stemming means reducing morphologically alike words to a simple term which is called *stem* or *root* [46]. For example the word "مشکلات" (difficulties), is reduced to the root "مشکل" (difficulty). FarsNet knowledgebase consists of stems corresponding to entities, not their different morphological shapes, therefore reducing different morphologies of words to their stems seems necessary. For this purpose we used the automatic stemmer for Persian words that is created by Nojavan, Ramezani and Feizi-Derakhshi [47]. This automatic stemmer uses a combination of linguistic rules and a database for identifying roots of Persian words.

E. **Named Entity Recognition (NER):** NER is identifying important names in text such as names of individuals, organizations and locations [48]. As we know, ontology is a hierarchical database of beings. For identifying the beings that have a name, we searched all the words and phrases of the input text in the FarsNet ontology; so if there is a match for each search, a named entity is identified. FarsNet version 2.0 is the basis for ontology-based calculation in this paper. Obviously, the number of entries in ontology affects the results of this step. The identified entities will be the vertices of the graph that shapes the conceptual basis for selecting sentences of the summary. By detecting the relationships between the entities, this graph which is a representation of the input text, will be completed in next step.
F. **Relation Extraction:** This is the last step of the preprocessing phase which its purpose is completion of the graph topology and creating a graph representation of the input document. By identifying the entities in previous step, vertices of the graph are created. In this step, by extracting semantic relations between entities from ontology, these relations are drawn as edged between the vertices, and the graph gets completed. For this purpose, a set of semantic relations consisting of synonymy, antonymy, meronymy, hypernymy, hyponymy that are extracted from FarsNet ontology, are used.

If one vertex exists in another synset, a directed edge would be drawn between them (originating from first vertex). At the same time, the synsets themselves relate to each other by means of semantic relations such as hypernymy, hyponymy or meronymy. If the word X is a subtype or an instance of Y, then X is a hyponym of word Y (or Y is a hypernym of word X). For example "Grahame Bell" is a hyponym for "Inventor" (also "inventor" is a hypernym for "Grahame Bell"). Also if X is a part of Y, Y is meronymy of Y. For example finger is a meronymy of hand. If there is hypernymy, hyponymy or meonymy relation between a pair of vertices, a directed edge would be drawn between them. Also by extracting the antonymy relation between entities from FarsNet, if there is antonymy between two vertices, another directed edge would be drawn between them. The idea behind this approach is that existence of each of these relations between vertices, represents the semantic relation between them. By doing this, a graph representation of the input document would be created that includes semantic relations between its entities, and semantic schema of the document. The resulting graph is a small sub-graph of the entities in the ontology and relations between them. It must be noted that since there might be more than one of these relationships between two vertices, in the resulting graph it is possible two draw more than one edge between two vertices (directed multigraph).

3.2.2 Processing

Fig. 2 demonstrates the steps in this phase. In this phase by processing the information obtained from previous phase, most important and coherent sentences would be selected for summary. As it can be seen in Fig. 2, the input of this phase is graph representation of input text, and the output is the summary of the original text. This phase has 4 steps that are respectively:

A. **Selecting the Most Connected Component:** Usually in different texts, despite the existence of a main/central topic, there are many side topics. The graph representation resulting from preprocessing phase, is a complete replacement for the input text and contains all the topics in it. Existence of different topics in the input text would create different connected components. The goal of this paper is to produce coherent summaries of input text. Without a doubt, given the size of final summary (which is usually 30% of the input text), putting sentences from different topics in the input text, would not result in a coherent summary. Because of that, in this part, we want to select the most connected sub-graph, to identify the most important/central topic discussed in the input text. The selected sub-graph creates the basis for scoring sentences and their placement in summary. By doing this, side topics (less related topics) would be eliminated and only most important/central topics of the input text will remain, and the summarization system will concentrate on the core topic. For this purpose, the component that has the most vertices (entities) and the highest sum for the degrees of its vertices, would be selected as the most connected component.

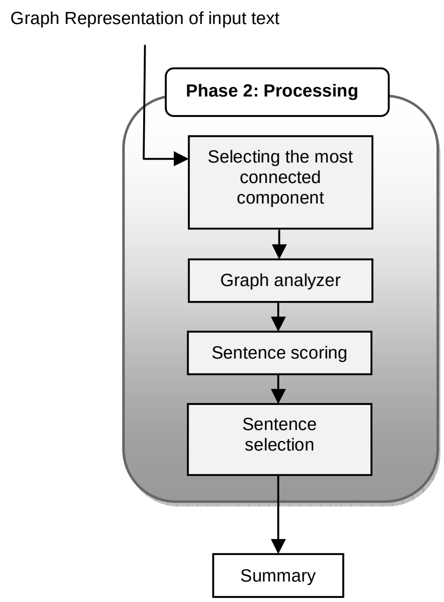

Fig. 2. Processing phase

B. Graph Analyzer: In extractive summarization, summary is produced by selecting a subset of sentences of original text. For this purpose, a set of the most central sentences would be selected that contain the most important information of the original text. Usually the degree of centrality of a sentence is determined based on the centrality of its words [49]. After identifying the most important topics in a text, we want to determine the degree of centrality for each one the entities, and by using it, calculate the degree of centrality for each of the sentences. For this purpose we use 3 graph-based evaluation measures, including *degree centrality* [50], *eigenvector centrality* [51] and *barycenter centrality* [52], for evaluating the importance/centrality of each of vertices (entities).

- **Degree centrality:** Based on this measure, centrality of each vertices/node is equal to its degree, or number of its connections. Since this is a directed graph, degree of each vertex is equal to sum of the inward and outward edges of it. Based on this measure, degree of a vertex being high means that it is more connected to other vertices, so its semantic importance is higher.
- **Eigenvector centrality:** This measure is an expansion of the degree centrality, but unlike degree centrality, the centrality of a vertex/node is not merely dependent on its degree, but also degrees of the vertices that are connected to it affect its centrality. In other words, based on this measure, centrality of a node would be higher if it is connected to other nodes with high centrality. Therefore a vertex with few connections to nodes with high degrees, has more centrality than a vertex that has more connections, but with low degree vertices. Therefore centrality of each node equals to sum its degree with degrees of vertices that have direct connection (direct edge) with it.
- **Barycenter centrality:** Unlike the two measures before, that centrality of vertices was based on its degree or its adjacent vertices degree, in this measure, the number of edges in the shortest path represents its centrality. Based on this measure, the amount of barycenter centrality of vertex v equals to 1/ total distance from vertex v to all other vertices. Based on this measure if the sum of distances of the vertex from other vertices is high, then the vertex has less centrality, because it does not have direct (semantic) relation with other vertices and vice versa. The highest amount of this centrality is achieved when distance of the vertex with every other vertex is one, which means there is a direct edge between the two vertices. In this case, certainly of the entity in question is of high importance that has direct semantic relations with other entities. Also in this measure, if there is no path between two nodes, distance between them would be taken as infinite.

C. **Sentence Scoring:** In this step we are decided to determine the centrality of each sentence and achieving a ranking of them. For this purpose, 2 types of scores would be calculated for each of the sentences; *static centrality score*, and *dynamic centrality score*. Based on the static centrality score, score of each sentence is calculated individually. In other words, in this approach, score of each sentence is calculated merely based on its relation with the most important topic of the text. For doing this, after determining the centrality of each of the entities in previous step, its static centrality score would be equal to the sum of centrality of all the entities inside it, that has been calculated based on a measure (of the measures mentioned in previous step) selected by the user. It must be noted that based on the selection of the most connected part of the graph, numerous entities has been removed. Functional basis of the summarization system would be this selected sub-graph, and centrality of the entities that are not in it would be taken as zero; since we want to achieve more coherent sentences, this approach is justified. Also, in order to prevent short sentences being sacrificed at the expense of longer sentences, this score is normalized with the length of sentences.

Dynamic centrality score of the sentences is suggested in order to increase coherence of the summaries, and based on it, score of the sentences are calculated not only based on the relation of them with the most important topic of text, but also score of the previous and next sentences would affect their chance of being selected. In other words, with this technique, selection of a sentence for putting in summary would increase the chance of selection for its next and previous sentences. The idea behind this approach is that since sentences with high scores are more related to main topic of text, and considering the fact that in well written text, adjacent sentences have high semantic correlation, then if previous and next sentences of a sentence are selected for summary, their score would affect the score of the sentence, therefore increasing the semantic relation between sentences of summary. Static centrality score of the sentence affect their dynamic centrality score. For calculating this score, in first step static centrality score of the sentences must be calculated, and sentences are sorted based on this score. Then considering the *compression rate*, which represents the length of summary in relation to the original document, sentences with highest static score are nominated for putting in the summary. Final decision for their selection, is made based on dynamic centrality score of the sentences. Dynamic score of sentence S_i equals to:

$$\begin{aligned} DCScore\ (S_i) = {} & (X_{i-1} \times IF_{i-1} \times SCScore\ (S_{i-1})) \\ & + SCScore\ (S_i) \\ & + (X_{i+1} \times IF_{i+1} \times SCScore\ (S_{i+1})) \end{aligned} \tag{1}$$

Which in this equation, X is a coefficient that if the previous (S_{i-1}) or next sentence (S_{i+1}) is selected it would be 1 and if not selected, it would be 0. $SCScore\ (S_i)$ is static centrality score of the i^{th} sentence that is calculated based on one of the graph evaluation methods discussed before. IF is impact factor that is used to control the amount of influence current sentence takes from previous and next sentences in order to prevent fast convergence the selected sentences to sentences with high scores. The value of IF is calculated based on the equation 2 proportional to occurrence of entities of current sentence S_i in previous (next) sentence and also adjacency of these entities to the entities of previous (next) sentence. This impact factor that implicitly refers to relation of the current sentence to previous (next) sentence, is normalized with length (number of words) of current sentence and sum of the number of edges between entities in current sentence.

$$IF_{i-1} = \frac{Number\_of\_Common\_Entities\_Between\_S_i\_and\_S_{i-1}}{Length(S_i)} + \frac{Number\_of\_Common\_Edges\_Among\_Entities\_in\_S_i\_and\_S_{i-1}}{Total\_Number\_of\_Edges\_Among\_Entities\_in\_S_i} \tag{2}$$

D. **Sentence selection and summary production:** As the last step of the summarization process, in this step sentences are sorted based on their *DCScore* and considering the compression rate, sentences with highest score are selected for putting in the summary. In this paper we have used 30% compression rate for all the summarization processes, which is a common compression rate.

4 Evaluation Results and Discussion

Approaches for evaluating automatic summarization systems are generally divided in two types: *extrinsic* evaluations and *intrinsic* evaluations [53]. In extrinsic evaluation approaches, the quality of summary is evaluated based on performing specific tasks (such as information retrieval), while in intrinsic evaluation approaches, quality of summaries are evaluated independently and based on analysis of summary ([53] for more study). It can be said that in intrinsic evaluation approaches for automatic summarization systems, in fact it is coherence and informativeness of the summary that is being evaluated, while in extrinsic approaches, it is the effect of produced summaries in other tasks such as text categorization, that is being evaluated [54].

The basis for evaluation and comparison of results of automatic summarization systems (*system summary*), are summaries that are produced by humans, which are referred to as *golden* or *reference summaries*. As mentioned before, summaries produced by a generic summarization system, must cover all the topics discussed in original text; also in golden summaries produced by humans for purpose of evaluating generic summarization systems, it is tried to cover most important sentences from all the topics as much as possible and this means variety of topics in summaries and less coherency; and this contradicts our goals. Therefore golden summaries that contain different topics from original document and suffer from lack of coherence, are not a good measure for evaluating success of a system that is trying to achieve coherent summaries. In other words for evaluating coherence of a system summary, we must have a coherent golden summary. Hence we have produced coherent golden summaries from a set of documents that only contain sentences related to most important/central topics discussed in the main documents; we have used these human produced summaries to evaluate the results of the suggested system.

For evaluating the results of summarization we have used *Precision* and *Recall* intrinsic measures and also *F-score* as a combination of these two. Precision measure is the fraction of retrieved instances that are relevant, and recall is the fraction of relevant instances that are retrieved [55]. With regard to the fact that we deal with sentences as desired text units, it can be said that precision measure equals to number of common sentences between golden summary and system summary divided by number of sentences in system summary. Also recall measure is equal to number of common sentences between golden summary and system summary divided by number of sentences in golden summary [53]. F-score measure is equal to harmonic average of precision and recall measure and would be equal to $(2\times P\times R)/(P+R)$.

The objective of the ontology based summarization system designed in this research, is to produce coherent summaries. For this purpose, by using ontology and semantic relations in it, based on a graph structure extracted from ontology that its vertices are entities and its edges are relations between these entities, most central sentences are identified based on their static centrality score, and are nominated for being put in the summary. As it was discussed in 3.2.2, by using three measures of degree centrality, eigenvector centrality and barycenter centrality, it is possible to evaluate centrality of graph vertices and calculate static score of the sentences in the text. After identifying nominated sentences for being in the summary by using static centrality score, dynamic centrality score (equation 1) of each of the sentences is calculated and a ranking of them is achieved based on this score. Finally, according to the compression rate, sentences with highest dynamic centrality score are selected to be in the summary. Table 1 (and its corresponding diagram in Fig. 3) shows the evaluation results for the summaries produced by the suggested summarization system; it contains three measures of precision, recall and F-score based on the centrality evaluation measure used for evaluation. These values presented in Table 1 are the results of average values, obtained from evaluation of a set of documents (110 documents).

Table 1. Evaluation results using dynamic centrality score

Centrality evaluation metric	Precision (%)	Recall (%)	F-score (%)
Degree centrality	64.84	61.92	**63.38**
Eigenvector centrality	**69.80**	**66.18**	**67.98**
Barycenter centrality	61.20	57.64	**59.40**

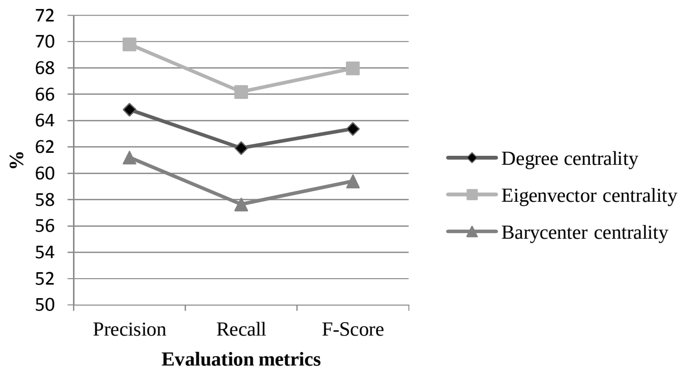

Fig. 3. Evaluation results using dynamic centrality score

As it can be seen, eigenvector centrality has achieved the best results among these three measures. In other words, among three measures for evaluating centrality, the measure that considers the degree of the adjacent vertices in addition to degree of each vertex, has better performance in evaluation of importance of different textual parts. Also, summaries produced using degree centrality measure have attained better results than barycenter centrality measure for precision, recall and F-score measures. It can be deduced that in evaluating centrality of graph vertices, measures that are based on the degree of vertices (in first place the measure that considers the degree of the current vertex and degree of its adjacent vertices, and in second place, the measure that only considers degree of the current vertex) have better performance than measures that are based on the distance of the vertices from each other.

And for comparing the success of the suggested technique for producing coherent summaries, we have also done the summarization only based on the static centrality score. Table 2 (and its corresponding diagram in Fig. 4) shows the evaluation results of the produced summaries, for precision, recall and F-score measures considering three measures of calculating the centrality of sentences. Summarization results of this approach, same as dynamic score approach, are best when eigenvector centrality measure has been used, and after that, degree centrality and barycenter centrality respectively have the best results. Likewise, these values are the average of results of each approach for a set of summaries in each of the three evaluation measures.

Fig. 5 shows a schematic comparison of F-score of summaries produced using both dynamic and static centrality score. As it can be seen, while both approaches have a similar pattern, summarization based on dynamic score is more successful in achieving more coherent summaries.

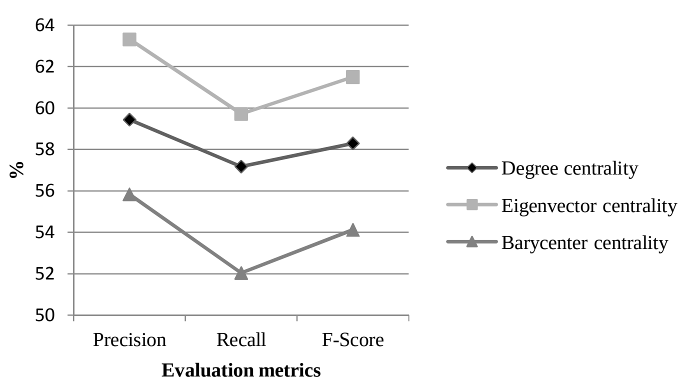

Fig. 4. Evaluation results using static centrality score

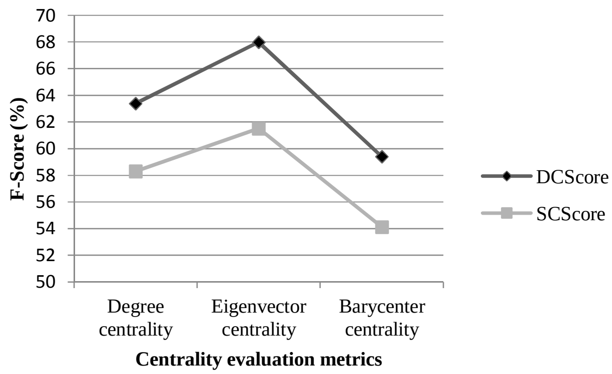

Fig. 5. Comparing summaries produced using dynamic and static centrality scores

Table 2. Evaluation results for summarization using static centrality score

Centrality evaluation metric	Precision (%)	Recall (%)	F-score (%)
Degree centrality	59.44	57.18	**58.30**
Eigenvector centrality	**63.32**	**59.72**	**61.50**
Barycenter centrality	55.84	52.04	**54.12**

5 Conclusion

In this paper with goal of achieving coherent summaries in automatic text summarization, an extractive and generic summarization system is designed that uses ontology as basis for producing summaries and increasing relation between its sentences. The idea behind this approach is that by selecting a sentence to be in the summary, the chance of its adjacent sentences to be in the summary would be increased. Evaluation results of summaries produced, shows that this idea has been successful in achieving coherent texts for generic summarization without any queries. But considering the flaws in the suggested approach for anaphora resolution, it is expected that using more comprehensive approaches for this purpose, and more complete versions of FarsNet ontology would be effective in improving the results.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Jones KS. Automatic summarizing: factors and directions. In Advances in automatic text summarization, Cambridge, MA: MIT Press. 1999;1–12.

[2] Ježek K, Steinberger J. Automatic text summarization (The state of the art 2007 and new challenges). In Proceedings of Znalosti. 2008;1–12.

[3] Lloret E. Text summarization: An overview. Pap. Support. by spanish Gov. under Proj. TEXT-MESS; 2008.

[4] Nenkova A, McKeown K. A survey of text summarization techniques. In Mining Text Data. Aggarwal CC, Zhai C. Eds. Boston, MA: Springer. 2012;43–76.

[5] Lloret E, Palomar M. Text summarisation in progress: A literature review. Artif. Intell. Rev. 2012;37(1):1–41.

[6] Saggion H, Poibeau T. Automatic text summarizatio: Past, present and future. In Multi-source, Multilingual Information Extraction and Summarization, Springer. 2013;3–21.

[7] Gupta V, Lehal GS, A survey of text summarization extractive techniques. J. Emerg. Technol. Web Intell. 2010;2(3):258–268.

[8] Lee JH, Park S, Ahn CM, Kim D. Automatic generic document summarization based on non-negative matrix factorization. Inf. Process. Manag. 2009;45(1):20–34.

[9] Mani I, Maybury MT, Sanderson M. Advances in automatic text summarization. The MIT Press. 1999;26(2):280–281.

[10] Bamberg B. What makes a text coherent?. Coll. Compos. Commun. 1983;417–429.

[11] Landauer TK, Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol. Rev. 1997;104(2):211–240.

[12] McKoon G, Ratcliff R. Inference during reading. Psychol. Rev. 1992;99(3):440.

[13] Barzilay R, Lapata M. Modeling local coherence: An entity-based approach. Comput. Linguist. 2008;34(1):1–34.

[14] Lawson T. A conception of ontology. Mimeogr. Univ. Cambridge; 2004.

[15] Gruber TTR. What is an Ontology; 1993.

[16] Lapata M. Probabilistic text structuring: Experiments with sentence ordering. In Proceedings of the 41st Annual Meeting on Association for Computational Linguistics. 2003;1:545–552.

[17] Althaus E, Karamanis N, Koller A. Computing locally coherent discourses. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. 2004;399.

[18] Lapata M, Barzilay R. Automatic evaluation of text coherence: Models and representations. In IJCAI. 2005;5:1085–1090.

[19] Lin Z, Ng HT, Kan MY. Automatically evaluating text coherence using discourse relations. In proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies. 2011;1:997–1006.

[20] Crossley SA, McNamara DS. Cohesion, coherence, and expert evaluations of writing proficiency. In Proceedings of the 32nd annual conference of the Cognitive Science Society. 2010;984–989.

[21] Luhn HP. The automatic creation of literature abstracts. IBM J. Res. Dev. 1958;2(2):159–165,.

[22] Dehkordi P, Kyoumarsi F. Using gene expression programming in automatic text summarization. Middle East J Sci Res. 2013;13(8):1070–1086.

[23] McKeown K, Radev DR. Generating summaries of multiple news articles. In Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval. 1995;74–82.

[24] Lin CY, Hovy E. From single to multi-document summarization: A prototype system and its evaluation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. 2002;457–464.

[25] Chowdary CR, Sravanthi M, Kumar PS. A system for query specific coherent text multi-document summarization. Int. J. Artif. Intell. Tools. 2010;19(05):597–626,.

[26] Mithun S, Kosseim L. Discourse structures to reduce discourse incoherence in blog summarization. In RANLP. 2011;479–486.

[27] Christensen J, Mausam SS, Soderland S, Etzioni O. Towards coherent multi-document summarization. In HLT-NAACL. 2013;1163–1173.

[28] Miller T. Latent semantic analysis and the construction of coherent extracts. Recent Adv. Nat. Lang. Process. III. 2004;260:277–286.

[29] Landauer TK, McNamara DS, Dennis S, Kintsch W. Handbook of latent semantic analysis. Psychology Press; 2013.

[30] Celikyilmaz A, Hakkani-Tür D. Discovery of topically coherent sentences for extractive summarization. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011;1:491–499.

[31] Jiang J, Zhai C. Extraction of coherent relevant passages using hidden markov models. ACM Trans. Inf. Syst. 2006;24(3):295–319,.

[32] Zhang R, Li W, Gao D. Generating coherent summaries with textual aspects. In AAAI. 2012;1727–1733.

[33] Yoo I, Hu X, Song IY. A coherent graph-based semantic clustering and summarization approach for biomedical literature and a new summarization evaluation method. BMC Bioinformatics. 2007; 8(Suppl 9):S4.

[34] Shamsfard M, Akhavan T, Joorabchi ME. Persian document summarization by parsumist. World Appl. Sci. J. 2009;7:199–205.

[35] Kamyar H, Kahani M, Kamyar M, Poormasoomi A. An automatic linguistics approach for persian document summarization. In Asian Language Processing (IALP), 2011 International Conference on. 2011;141–144.

[36] Zhang R. Coherence-targeted text summarization. The Hong Kong Polytechnic University; 2013.

[37] Nishikawa H, Arita K, Tanaka K, Hirao T, Makino T, Matsuo Y. Learning to generate coherent summary with discriminative hidden semi-markov model. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics. 2014;1648–1659.

[38] Schiffrin D, Tannen D, Hamilton HE. The handbook of discourse analysis. John Wiley & Sons; 2008.

[39] Zienkowski J, Östman JO, Verschueren J. Discursive pragmatics. John Benjamins Publishing. 2011;8.

[40] Taboada M, De los Ángeles Gómez-González M. Discourse markers and coherence relations: Comparison across markers, languages and modalities. Linguist. Hum. Sci. 2012;6(1–3):17–41.

[41] Wang Y, Guo M. A short analysis of discourse coherence. J. Lang. Teach. Res. 2014;5(2):460–465.

[42] Shamsfard M, Hesabi A, Fadaei H, Mansoory N, Famian A, Bagherbeigi S, Fekri E, Monshizadeh M, Assi SM. Semi automatic development of farsnet; the persian wordnet. In Proceedings of 5th Global WordNet Conference, Mumbai, India; 2010.

[43] Miller G, Fellbaum C. Wordnet: An electronic lexical database. MIT Press Cambridge; 1998.

[44] Shamsfard M. Developing FarsNet: A lexical ontology for Persian. In 4th Global WordNet Conference, Szeged, Hungary; 2008.

[45] Mitkov R. Anaphora resolution. Routledge; 2014.

[46] Gupta V. Automatic stemming of words for punjabi language. In advances in signal processing and intelligent recognition systems. Springer. 2014;73–84.

[47] Nojavan MB, Ramezani M, Feizi-Derakhshi MR. Automatic stemming of persian words using an optimal combination of lexical rules and database. In 8th International Conference of Iranian Society for Promotion of Persian Language and Literature. 2013;1–11.

[48] Mohit B. Named entity recognition. In Natural Language Processing of Semitic Languages. Zitouni I, Ed. Springer. 2014;221–245.

[49] Erkan G, Radev DR. LexRank: Graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res.(JAIR). 2004;22(1):457–479.

[50] Freeman LC. Centrality in social networks conceptual clarification. Soc. Networks. 1979;1(3):215–239.

[51] Bonacich P. Some unique properties of eigenvector centrality. Soc. Networks. 2007;29(4):555–564.

[52] Beveridge A. Centers for random walks on trees. SIAM J. Discret. Math. 2009;23(1):300–318.

[53] Steinberger J, Ježek K. Evaluation measures for text summarization. Comput. Informatics. 2012; 28(2):251–275.

[54] Jagarlamudi J. Query-based multi-document summarization using language modeling. Master's thesis, IIIT Hyderabad; 2006.

[55] Rüger S. Multimedia information retrieval. Synth. Lect. Inf. Concepts, Retrieval, Serv. 2009;1(1):1-171.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/16891