

A Turkish Wikipedia Text Summarization System for Mobile Devices

Akif Hatipoglu

Yapı Kredi Bank, Content and Document Management Software Programming Department, Software Developer,
Levent/İstanbul, 34330, Turkey
E-mail: akif.hatipoglu@yapikredi.com.tr

Sevinç İlhan Omurca

Kocaeli University Computer Engineering Department, Kocaeli, 41000, Turkey
E-mail: silhan@kocaeli.edu.tr

Abstract—Today Wikipedia provides a very large and reliable domain-independent encyclopedic repository. With this study a mobile system which summarizes Turkish Wikipedia text is presented. The presented system selects the sentences due to structural features of Turkish language and semantic features of the sentences. The performance evaluation is made based on judgments of human experts. The results are tested due to precision and recall values of a ranked sentence list and it is concluded that, the summarization results are promising.

Index Terms—Turkish text summarization, Turkish Wikipedia, Latent semantic analysis, Helmholtz principle, Mobile application.

I. INTRODUCTION

Automatic text summarization has become a critical and popular research area for finding relevant information quickly and effectively. There are two main reasons for this: first, there is an exponential and uncontrolled growth of online information, and second, small computing devices such as mobile phones constrained by screen space are becoming increasingly popular. Today when a topic is researched, a lot of document resources are easily accessed through the internet. The huge number of available documents in digital media makes it difficult to obtain the necessary information related to the needs of users. Therefore, there is an increasing need for new technologies that help users to access relevant and core information effectively. Automatic text summarization systems reduce the access time for the relevant information and allow users to understand main theme of the document easily. Radev et al. [22] define three aspects that characterize research on automatic summarization like that, summaries may be produced from a single document or multiple documents, summaries should preserve important information and summaries should be short.

There are basically two types of automatic summarization systems in natural language processing which are defined as summarization by selecting sentences and summarization by interpreting sentences.

The first group of techniques involves selecting the most important sentences, whereas the second group of techniques involves generating novel sentences from the given documents. Interpreting sentences approach requires a deeper semantic analysis of documents according to selecting sentences approach. Furthermore, in contrast to interpreting sentences approach, selecting sentences approach is more practical [11]. The summarization by selecting sentences method is also separated into two sub-methods. The first one is the method of scoring the sentences statistically due to their structural features, such as, the position of the sentence, the number of words in the sentence, how many title words the sentence has. The second one is latent semantic analysis (LSA) of the sentences which integrates the semantic relations between words, and their syntactic composite.

In this study, we present a mobile text summarization system for Turkish Wikipedia documents. There hasn't been a mobile application system which summarizes Turkish documents automatically before this study. Furthermore, this application is not for only Wikipedia texts but also can be easily adapted for any kind of documents. The proposed text summarization system ranks the sentences from relevant to irrelevant and then the summary of a given document is generated with the top n sentences of this ranked list. Here, n is a user defined parameter and in this paper it is defined as 40% of sentences in the document. As in the Information Retrieval (IR) and web search, in document summarization usually no certain decision is made on whether a sentence is relevant or irrelevant to a summary. Therefore in this study we evaluate the ranking of the sentences for the first time as a summary evaluation metric.

The AHP model is used to integrate the scores of structural and semantic features into an overall sentence score as in [12]. The structural and semantic features are linearly combined using the weights determined by AHP. Length, position, and the number of title terms are considered as structural features of a sentence. Semantic features are examined in two groups, latent semantic based features and Wikipedia semantic based features. In

the similar studies which summarize Wikipedia text, the Wikipedia keywords are not used as a separate feature. In this study we propose to use Wikipedia keywords as a semantic feature of sentences.

In the literature, evaluating summary process is reported as a difficult task due to the absence of a standard human or automatic evaluation metric. This drawback makes an automatic summary system very hard to evaluate. In this paper, in order to see the performance of the applied summarization system, a human-generated summary corpus is used. The precision and recall values which determine the coverage between the human-generated and automatically generated summaries is used as performance evaluator. The computational results show that the proposed summarization method is a promising approach to create a valid summary of Turkish Wikipedia documents.

The rest of the paper is organized as follows: Section 2 refers the related work, section 3 explains the proposed system. The experimental results presented in section 4. Finally, discussion and conclusions for the future work are summarized in section 5.

II. RELATED WORK

In the literature there are some studies which represent documents with their structural features as in [29]. However, most of the related studies represent documents with structural features concomitantly with semantic relations.

Yeh et al. [30] proposed a trainable text summarizer which considers position, positive keyword, negative keyword, centrality, and the resemblance to the title as document features to generate summaries. These features are combined by a weighting score function using genetic algorithm (GA). And they also used LSA to derive the semantic matrix of a document. Binwahlan et al. [6] proposed a hybrid model based on fuzzy logic, swarm intelligence and diversity selection for text summarization problem. They use sentence centrality, title words, word sentence score, keyword terms, and similarity to the first sentence as the features of sentences and the features are combined in a linear combination to generate summary sentences.

The automatic text summarization has not been extensively studied for the Turkish language. Güran et al. [11] used non-negative matrix factorization method as a feature reduction method and summarized 100 news documents. Güran et al. [12] presented a summarization system that combines some structural and semantic features of sentences by using analytical hierarchical process (AHP) and artificial bee colony algorithm. Cıgır et al. [7] generated summaries by ranking sentences due to their scores calculated by combining the features such as term frequency, title similarity, key phrases, position of the sentence in the document, and centrality of the sentence. Ozsoy et al. [19] presented LSA based

summarization algorithms for Turkish documents.

In our daily lives we want to reach the core information of any subject as fast as possible. Mobile devices and applications provide fast access to data and many benefits for accessing data from everywhere you want. While the people are searching for a topic, they usually might first want to look at a short summary, and then request the full document. Therefore in this paper, we decided to realize the text summarization as a mobile application.

Wikipedia, currently the world's largest online encyclopaedia is perhaps the best reliable information resource which comes to mind first, whereby the entire information is every time available to everyone. It is an online free-content encyclopaedia which is very often referenced by the people for accessing the information. In the literature a few recent studies have leveraged Wikipedia for summarization tasks. Ramanathan et al. [23] mapped sentences to semantic Wikipedia concepts and select sentences for the summary based on the concept frequency thresholds. Gong et al. [10] and Bawakid et al. [5] used a thesaurus extracted from Wikipedia in finding the most important concepts within a document or a set of documents. Pourvali and Abadeh [20, 21] extracted sentences with high importance from several independent graphs using Wikipedia. Sentences are considered as nodes and the relationships between words within a sentence with other sentence words are considered to be edges in the graph. They asserted that, the graph which is higher than others contains main topic of the text. Miao and Li [18] presented a WikiSummarizer system by examining sentences in the feature space of Wikipedia concepts considering their semantic relatedness. Sankarasubramaniam et al. [25] presented a Wikipedia-based multi-document summarization algorithm that the users first view an initial summary, and then request additional content if interested. Sakhare and Kumar [26] presented a hybrid approach to text summarization by combining features and syntactic structure. They calculated the score of sentences by combining two Neural Networks.

III. THE PROPOSED SYSTEM FOR TEXT SUMMARIZATION

The first main task of summarization system is pre-processing of the original document. The pre-processing steps consider eliminating stop words, document segmentation by sentences and stemming processes. After sentences of the original document are extracted, the stop words are eliminated then the stemming is applied by Zemberek [31] which is an open source NLP framework for Turkish language. Finally the structural and semantic analysis of pre-processed document is realized and the extracted features combined using a sentence score function. The determined features of a sentence are shown in Fig. 1.

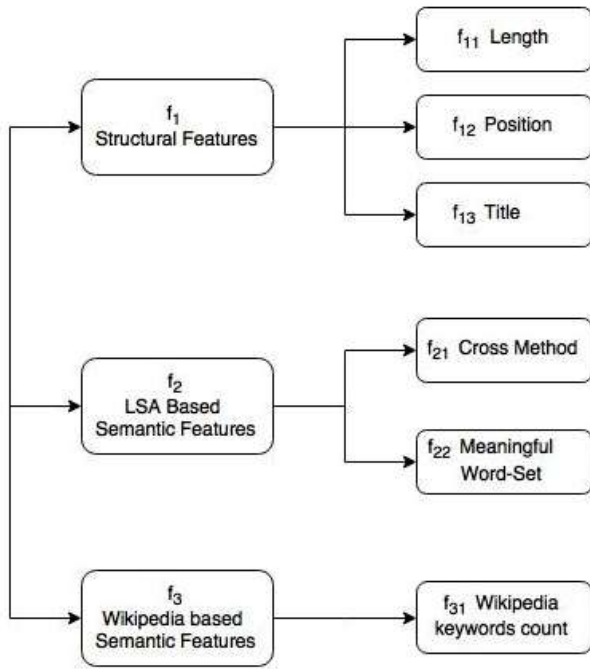


Fig.1. The structural and semantic features of a sentence.

In our proposed summarization model, the features of a sentence are determined in three main parts: structural features, LSA based semantic features and Wikipedia based semantic features which are indicated as f_1 , f_2 , f_3 respectively. The structural features consist of: length of the sentence, position of the sentence and title information contained by the sentence. The LSA based semantic features are extracted by using cross method and meaningful word-set method. Wikipedia based semantic feature is calculated by counting the Wikipedia keywords in the sentence.

Structural features: Structural features that indicate some structural rules of a natural language are particularly valuable in text summarization. The structural features, f_1 , for a sentence $S^{(i)}$ included in a given document $D = \{S^{(1)}, S^{(2)}, \dots, S^{(m)}\}$ is comprised of the following features.

Length: Each sentence in the original document is given a length score which represent the number of words contained in the sentence. The importance of this feature can be explained like that, the average length sentences considered more important in terms of information extraction. To formulate this consideration, first the average sentence length is calculated. The sentence which has closer length score to the average is more important for the summary. The short and long sentences are considered as less relevant. The formulation of this acceptance is shown in (1) and (2).

$$S_{score_{f_{11}}}^{(i)} = (\text{length}_{S^{(i)}} - 1/2m (\sum_{j=1}^m \text{length}_{S^{(j)}}) - 1/2)^2 \quad (1)$$

$$S_{score_{f_{11}}}^{(i)} = -1 \times S_{score_{f_{11}}}^{(i)} + \max_{S_{score_{f_{11}}}^{(i)}} \{S_{score_{f_{11}}}^{(1)}, \dots, S_{score_{f_{11}}}^{(m)}\} \quad (2)$$

Here, $S_{score_{f_{11}}}^{(i)}$ is the summarization score of the sentence i in a document D and $\text{length}_{S^{(i)}}$ is length of the sentence.

Position: In Turkish, documents are composed three main parts which are introduction, body and conclusion. Sentences at the beginning of the documents considered as introduction sentences and describe the main issues of the documents. Similarly sentences at the end of the documents considered as the gist of text and may also describe the main issue of the documents. In the body of the text, idea of introduction part is extended by supporting examples. Thus body is longer and more detailed than introduction and conclusion. To evaluate the importance of different sentence positions, each sentence in a document is ranked as in:

$$S_{score_{f_{12}}}^{(i)} = (\text{position}_{S^{(i)}} - 1/2(\text{position}_{S^{(m)}}) - 1/2)^2 \quad (3)$$

Here, $S_{score_{f_{12}}}^{(i)}$ is the summarization score of the $S^{(i)}$ in a given document D and $\text{position}_{S^{(m)}}$ is the final sentence position of this document

Title: The title feature defines how many words in a sentence are also located in the title sentence. If the indicated sentence involves title words, then this sentence considered as an important sentence for the summary text [12]. For each sentence, the involved title words are directly proportional to the summarization score of the sentence. Let, $T = \{t^{(1)}, t^{(2)}, \dots, t^{(k)}\}$ represent the set of title words for a given document D , then the summarization score of $S^{(i)}$ is calculated as in:

$$S_{score_{f_{13}}}^{(i)} = \begin{cases} 1 \{S^{(i)} \cap T = T\} \text{ if } S^{(i)} \text{ contains all title words together} \\ \sum_{j=1}^k 0.5 \times 1 \{S^{(i)} \cap t^{(j)} \neq \emptyset\} \text{ else} \end{cases} \quad (4)$$

If sentence $S^{(i)}$ contains all of the title words together then it's $S_{score_{f_{13}}}^{(i)}$ is calculated as +1. If sentence $S^{(i)}$ doesn't contain all of them together then it takes 0.5 point for every title word included in the sentence.

Semantic Features: In this study, for a sentence $S^{(i)}$ in a document D , not only structural features but also the semantic features are considered. The semantic features are split in half; LSA based and Wikipedia based semantics.

The LSA approach [9] uses singular value decomposition (SVD) method to decompose a large scaled term-document matrix into a set of k orthogonal factors. SVD preserves the relative distances in the sentence-term matrix, while projecting it into a Semantic Space Model, which has a lower dimensionality. This allows keeping just the minimum information needed to define the appropriate representation of the sentences. Furthermore, the SVD of the Tf-Idf (Term frequency-Inverse document frequency) matrix reveals the

underlying semantic relationship between terms and documents [14].

A $m \times n$ dimensional term-sentence matrix, A , which represents the correlation between each term to the whole sentence set of a given document, can be written in a form of equation as in:

$$A = USV^T \quad (5)$$

Here, U and V^T are orthogonal matrices and S is a diagonal matrix. For dimension reduction, an appropriate k value is selected. Then, the first k columns of matrix U are selected as U_k ; the first k rows of matrix V^T are selected as V_k^T ; the k factors of matrix S as S_k . The matrix A' is considered to capture the most important relationship between terms and sentences.

$$A' = U_k S_k V_k^T \quad (6)$$

SVD reduces the original matrix A into k topics that contain remarkable patterns of word combinations in a given document. Each column i of matrix A corresponding to the sentence i is mapped to column i of V_k^T , which represents sentence i in the space of k topics. Matrix U_k emphasizes mapping between the space of k topics and the space of the m terms. Matrix S specifies the importance of the selected document topics [17].

After input matrix creation and SVD execution steps, sentence selection is realized by scoring sentences due to V_k^T matrix. In this study, LSA based semantic sentence scoring is realized by combining cross method and meaningful word-set method. The calculated score represents the first main semantic metric for a sentence.

Cross Method: In this method which is an extension to enhanced LSA Summarization approach [27] and proposed by [19], first the SVD of term-sentence matrix need to be computed. Then, the length of each sentence vector in matrix V_k^T is computed. The columns of the V_k^T matrix represent the sentences and the k rows of it represent the semantic concepts. For each concept, the average score is calculated and this score considered as a threshold value for the concept. Then the values which are less to the average score are set to zero. The new cell values are multiplied with their eigenvalues taken from the S_k matrix. The length of a sentence is found by summing up concepts values of that sentence. These processes are formulated as in (7). Finally, the calculated length value for each sentence is assigned to that sentence as the semantic score f_{21} .

$$S^{(i)}_score_{f_{21}} = \sqrt{V_{(i,j)}^T \times S_{(j,j)}} \quad (7)$$

Meaningful Word-Set Method: In this paper, the meaning score of a sentence is calculated by summing up the meaning values of its terms. The meaning values of the terms are calculated due to the Helmholtz principle [16] from Gestalt theory [1, 2, 3, 4]. A document consists of natural structures in the form of sentences, paragraphs and topics. In [2], meaningfulness of these natural

structures is defined using the Helmholtz principle from Gestalt Theory. The Helmholtz principle in human perception supposes that an observed geometric structure is perceptually meaningful if it has a very low probability to appear in noise [3]. In the context of text mining, Balinsky et al. [3] defined the Helmholtz principle as the statement that meaningful terms in a document appear as large deviations from randomness. We calculate the meaning score of each term in a sentence in our study.

Suppose that we are given a document $D = \{S^{(1)}, S^{(2)}, \dots, S^{(m)}\}$ with m sentences. Let P denote a family of sentences of the document. Wikipedia documents do not have natural paragraph structures. Thus, in our experiments four consecutive sentences are considered as a separate natural structure, P , of a document. We can define a measure of meaningfulness of a term t from D inside P as follows. If a term t appears m times in the first sentence $S^{(1)}$ and K times in the other sentences, first, the number of false alarms $NFA(t, S^{(1)})$ is defined by the following equation:

$$NFA(t, S^{(1)}) = \frac{K!}{m!(k-m)!} \times \frac{1}{N^{m-1}} \quad (8)$$

where N is defined by using the length of D and the length of P in the number of terms.

$$N = \frac{\text{length}(D)}{\text{length}(P)} \quad (9)$$

The meaningfulness of a term t in P defined as follows:

$$\text{Meaning}(t, P) = -\frac{1}{m} \log NFA(t, P) \quad (10)$$

For each P , a $\text{MeaningfulSet}(\epsilon)$ is defined as a set of all terms with $\text{Meaning}(t, P) > \epsilon$ and in our study $\epsilon = 0$. The meaning score of each term is defined as follows:

$$\text{Meaning}(t) = \max_{P \in D} \text{Meaning}(t, P) \quad (11)$$

The set $\text{MeaningfulSet}(\epsilon)$ is empty for a sufficient large positive ϵ and contains all the words from D for a large negative ϵ . The sum of $\text{Meaning}(t)$ scores of the terms included in $\text{MeaningfulSet}(\epsilon)$ is defined as the meaning score of a given sentence.

Wikipedia Based Semantic Features: The semantic similarity between concepts (or words) is an important issue in computer linguistic research area. Some concepts have the same or related meaning but they are not lexicographically similar. Wikipedia provides a semantic network for lots of domain-independent concepts [13]. Thus, the second main semantic metric for a sentence is defined as how many Wikipedia concepts it has. For all sentences in a document we construct a sentence-concept mapping scheme. If we have a high overlap of Wikipedia concepts across sentences, then we decide that the sentence is strong candidate to be a summary sentence.

Let $S = \{S^{(1)}, S^{(2)}, \dots, S^{(m)}\}$ denote the m sentences in an input document D and each sentence $S^{(i)} =$

$\{s_1^{(i)}, s_2^{(i)}, \dots, s_n^{(i)}\}$ has n distinct terms; $Z = \{z_1, z_2, \dots, z_l\}$ denote the Wiki concepts which are included in document D . For every sentence $S^{(i)}$, we calculate the Wikipedia semantic score of the sentence as in the following equation.

$$S^{(i)}_{score_{f_{31}}} = \sum_{j=1}^n \sum_{k=1}^l 1\{s_j = z_k\} \quad (12)$$

Combining Features of Sentences: To generate the summary of a given Wikipedia document, all of its sentences are ranked due to their summarization scores and then the sentences which are early in the list are chosen for the summary text. For each sentence $S^{(i)}$ in document D , a weighted score function based on AHP (Analytic hierarchy process) is employed to combine the entire features by the following linear model.

$$S^{(i)}_{score} = \sum_{j=1}^3 w_{1j}f_{1j} + \sum_{j=1}^2 w_{2j}f_{2j} + \sum_{j=1}^1 w_{3j}f_{3j} \quad (13)$$

Here, w_{ij} weights are determined by AHP model such in the [12]. To generate a summary, all sentences are ranked due to their scores calculated in (9), and a number of the top-score sentences are included in the summary. In our study 40% of sentences are selected as summary sentences however, this threshold can be changed by the user.

The Analytic Hierarchy Process (AHP) which is introduced by Saaty [24], is a multi-criteria decision making method that helps the decision-maker modelling a complex problem as a hierarchy of multiple conflicting criteria.

In our study the summarization problem is modelled as a hierarchy of sentence features. After the hierarchy is constructed, the decision makers determine the importance of each feature at each level of the hierarchy. The sentence features are compared to each other in a pair wise comparison matrix constructed by decision makers [29].

Table 1. Comparison matrix of structural, semantic and Wikipedia features

| Main Features | f_1 | f_2 | f_3 |
|---------------|-------|-------|-------|
| f_1 | 1 | 1/1.8 | 1/1.2 |
| f_2 | 1.8 | 1 | 1.1 |
| f_3 | 1.2 | 1/1.1 | 1 |

Table 2. Comparison matrix of length, position and title features

| Structural Features (f_1) | f_{11} | f_{12} | f_{13} |
|-------------------------------|----------|----------|----------|
| f_{11} | 1 | 1/1.2 | 1/1.3 |
| f_{12} | 1.2 | 1 | 1/1.1 |
| f_{13} | 1.3 | 1.1 | 1 |

Table 3. Comparison Matrix of Cross method and Meaningful Word-Set method

| Semantic Features (f_2) | f_{21} | f_{22} |
|-----------------------------|----------|----------|
| f_{21} | 1 | 2 |
| f_{22} | 1/2 | 1 |

Table 4. Comparison Matrix of Wikipedia Keywords Feature

| Wikipedia Semantic (f_3) | f_{31} |
|------------------------------|----------|
| f_{31} | 1 |

In the Fig. 1, the considered features of sentences are given in a hierarchical structure. Due to this structure, four pairwise comparison matrices are formed and shown in Table 1, Table 2, Table 3 and Table 4 respectively.

These comparison matrices are specified by using Expert Choice software [8] and the average pair wise comparison is computed. Finally the importance values of all features are identified and these values are shown in Fig. 2.

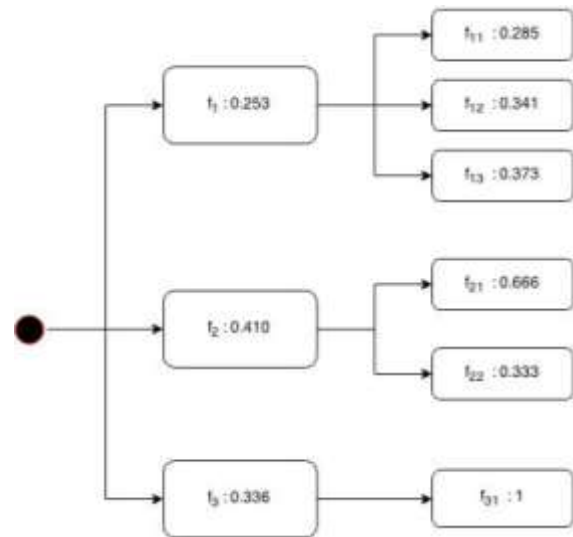


Fig.2. The structural and semantic features of a sentence.

The calculated AHP weights of features are as follows: 0.072 is for length, 0.086 is for position, 0.094 is for title, 0.273 is for cross-method, 0.136 is for meaningful word-set method and 0.336 is for Wikipedia based semantic. Thus the weighted score function of sentence $S^{(i)}$ is defined as in (14).

$$S^{(i)}_{score} = (0.072 \times f_{11} + 0.086 \times f_{12} + 0.094 \times f_{13}) + (0.273 \times f_{21} + 0.136 \times f_{22}) + (0.336 \times f_{31}) \quad (14)$$

IV. EXPERIMENTAL RESULTS

In IR usually no decision is made on whether a document is relevant or irrelevant to another document. Instead, a ranking of the documents is produced [15].

Document summarization is also an IR system where the most important sentences are extracted for the summary text.

Given a Wikipedia document D , this method first computes relevance scores for all sentences in D and then produce a ranking R_{summary} of the sentences based on the relevance scores,

$$R_{\text{summary}} = \langle S_1^{\text{summary}}, S_2^{\text{summary}}, \dots, S_N^{\text{summary}} \rangle.$$

Where $S_1^{\text{summary}} \in D$ is the most relevant sentence to the summary text and $S_N^{\text{summary}} \in D$ is the most irrelevant sentence to the summary text. The precision and recall values at each S_i^{summary} in the ranking are computed.

Recall at position i denoted by $r(i)$ is the fraction of relevant sentences from S_1^{summary} to S_i^{summary} in R_{summary} . Let the number of relevant sentences in this range be total_relevant_i . Then recall value is computed as in:

$$r(i) = \frac{\text{total_relevant}_i}{|R_{\text{summary}}|} \quad (15)$$

Precision at position i , denoted by $p(i)$, is the fraction of documents from S_1^{summary} to S_i^{summary} in R_{summary} and computed as in:

$$p(i) = \frac{\text{total_relevant}_i}{i} \quad (16)$$

The computed precision $p(i)$ and recall $r(i)$ values enable the evaluation of the sentence coverage among the manually and automatically generated summaries. Further, an average precision can be computed based on the precision at each relevant document in the ranking R_{summary} as in:

$$p_{\text{avg}} = \frac{\sum_{S_i^{\text{summary}} \in R_{\text{summary}}} p(i)}{|R|} \quad (17)$$

Performance analysis is conducted on the proposed system by using randomly selected Wikipedia documents. The manually generated summaries of the selected documents are compared with the automatically generated summaries. The empirical analysis of the proposed model is done by ten Wikipedia documents which are listed in Table 5.

Evaluating a document summarization system is a difficult task due to the absence of an automatic evaluation metric. In this paper, in order to analyze the performance of the applied system, the automatically generated and the human-generated summaries are compared. The length of the summary part is decided as 40% of the whole document's sentences. The precision and recall values which determine the coverage between the manually and automatically generated summaries is used as performance evaluator. Let assume that M is the manual summary and S is the automatically generated

summary, the measurements are defined as following; $P = \frac{|S \cap M|}{|S|}$ and $R = \frac{|S \cap M|}{M}$.

Table 5. Comparison matrix of structural, semantic and Wikipedia features

| ID | Title | Sentence Count | Word Count | The link of the document |
|----|----------------------------|----------------|------------|---|
| 1 | Solar system | 10 | 134 | http://tr.wikipedia.org/wiki/Güneş_Sistemi |
| 2 | Solar cooker | 6 | 66 | http://tr.wikipedia.org/wiki/Güneş_ocağı |
| 3 | Shine On You Crazy Diamond | 5 | 66 | http://tr.wikipedia.org/wiki/Shine_On_You_Crazy_Diamond |
| 4 | Elephant | 32 | 372 | http://tr.wikipedia.org/wiki/Fil |
| 5 | United States of America | 14 | 153 | http://tr.wikipedia.org/wiki/Amerika_Birleşik_Devletleri |
| 6 | Star wars | 13 | 187 | http://tr.wikipedia.org/wiki/Yıldız_Savaşları |
| 7 | Java programming language | 13 | 203 | http://tr.wikipedia.org/wiki/Java_(programlama_dili) |
| 8 | Computer | 13 | 150 | http://tr.wikipedia.org/wiki/Bilgisayar |
| 9 | Africa | 10 | 147 | http://tr.wikipedia.org/wiki/Afrika |
| 10 | Evolution | 27 | 470 | https://tr.wikipedia.org/wiki/Evrım |

In Table 6, the sentence ranking of the document titled as "elephant" is reported. The first column of the table represents the rank order; the second column represents the sentences which are listed due to their ranking scores in descending order. In the third column, the "+" sign means that the manually-selected and the automatically-selected sentence is coincided; "-" sign means that the manually-selected and the automatically-selected sentence is not coincided. The precision $p(i)$ and recall $r(i)$ values at each position i are given in columns 4 and 5.

The document titled as "elephant" has 32 sentences, and the first 12 of them are selected for the summary text. As seen in Table 6, our system ranked the first sentence as the most important sentence and ranked the 13th sentence as the less important sentence. The precision $p(i)$ and recall $r(i)$ values for the 12th level of the ranked list are 75%. This point is the breakeven point which is also commonly used as a performance metric. Thus the performance of the summarization for this document can be accepted as 75%. Or average precision value can be calculated instead of breakeven point. Our proposed system selects the top-ranked 40% of sentences, thus, the average precision value is:

$$p_{\text{avg}} = \frac{100\%+100\%+100\%+100\%+100\%+100\%+100\%+88\%+72\%+75\%}{12} = 69.5\% .$$

If the user decides to select first six sentences as the

summary text then the average precision of document summarization process is 100%. If the user decides to select ten sentences then the average precision values is:

$$P_{avg} = \frac{100\%+100\%+100\%+100\%+100\%+100\%+88\%}{10} = 68.8\%$$

Table 6. Sentence ranking of the document titled “elephant”

| Rank order (i) | Sentence number | +/- | p(i) | r(i) |
|----------------|-----------------|-----|-------------|--------------|
| 1 | 1 | + | 1/1 = 100% | 1/12 = 9% |
| 2 | 2 | + | 2/2 = 100% | 2/12 = 16% |
| 3 | 3 | + | 3/3 = 100% | 3/12 = 25% |
| 4 | 28 | + | 4/4 = 100% | 4/12 = 33% |
| 5 | 12 | + | 5/5 = 100% | 5/12 = 41% |
| 6 | 27 | + | 6/6 = 100% | 6/12 = 50% |
| 7 | 4 | - | 6/7 = 86% | 6/12 = 50% |
| 8 | 6 | + | 7/8 = 88% | 7/12 = 58% |
| 9 | 15 | - | 7/9 = 77% | 7/12 = 58% |
| 10 | 21 | - | 7/10 = 70% | 7/12 = 58% |
| 11 | 23 | + | 8/11 = 72% | 8/12 = 66% |
| 12 | 32 | + | 9/12 = 75% | 9/12 = 75% |
| 13 | 5 | + | 10/13 = 77% | 10/12 = 83% |
| 14 | 26 | - | 10/14 = 71% | 10/12 = 83% |
| 15 | 29 | - | 10/15 = 66% | 10/12 = 83% |
| 16 | 8 | - | 10/16 = 62% | 10/12 = 83% |
| 17 | 30 | - | 10/17 = 58% | 10/12 = 83% |
| 18 | 31 | + | 11/18 = 61% | 11/12 = 91% |
| 19 | 14 | - | 11/19 = 57% | 11/12 = 91% |
| 20 | 24 | + | 12/20 = 60% | 12/12 = 100% |
| 21 | 7 | - | 12/21 = 57% | 12/12 = 100% |
| 22 | 19 | - | 12/22 = 54% | 12/12 = 100% |
| 23 | 25 | - | 12/23 = 52% | 12/12 = 100% |
| 24 | 10 | - | 12/24 = 50% | 12/12 = 100% |
| 25 | 9 | - | 12/25 = 48% | 12/12 = 100% |
| 26 | 18 | - | 12/26 = 46% | 12/12 = 100% |
| 27 | 11 | - | 12/27 = 44% | 12/12 = 100% |
| 28 | 16 | - | 12/28 = 42% | 12/12 = 100% |
| 29 | 22 | - | 12/29 = 41% | 12/12 = 100% |
| 30 | 20 | - | 12/30 = 40% | 12/12 = 100% |
| 31 | 17 | - | 12/31 = 38% | 12/12 = 100% |
| 32 | 13 | - | 12/32 = 37% | 12/12 = 100% |

The obtained interpolated precisions at all 32 recall levels in Table 6 are shown as precision recall curve in Fig.3.

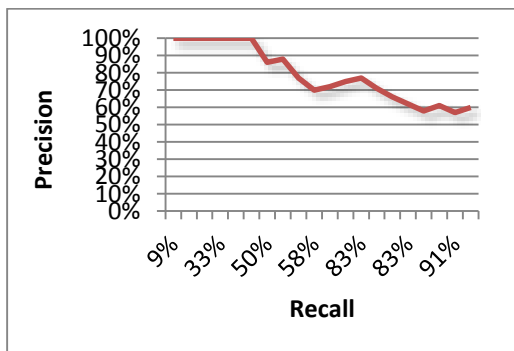


Fig.3. The precision-recall curve for document titled “elephant”

Table 7 shows the summarization results of another document which is titled “evolution”. This document has 27 sentences, and the first 10 of them are selected for the summary text. The precision and recall values for the 10th level of the ranked list are both 80%. Apart from that, as another evaluation criterion, the average precision of document summarization process is calculated as 73%. However, if the user prefers to select five sentences as the summary text, then the average precision is 100%.

Table 7. Sentence ranking of the document titled “evolution”

| The Rank order (i) | The Sentence number | +/- | p(i) | r(i) |
|--------------------|---------------------|-----|-------------|--------------|
| 1 | 24 | + | 1/1 = 100% | 1/10 = 10% |
| 2 | 5 | + | 2/2 = 100% | 2/10 = 20% |
| 3 | 1 | + | 3/3 = 100% | 3/10 = 30% |
| 4 | 20 | + | 4/4 = 100% | 4/10 = 40% |
| 5 | 4 | + | 5/5 = 100% | 5/10 = 50% |
| 6 | 25 | - | 5/6 = 83% | 5/10 = 50% |
| 7 | 16 | - | 5/7 = 71% | 5/10 = 50% |
| 8 | 18 | + | 6/8 = 75% | 6/10 = 60% |
| 9 | 27 | + | 7/9 = 77% | 7/10 = 70% |
| 10 | 17 | + | 8/10 = 80% | 8/10 = 80% |
| 11 | 7 | - | 7/11 = 63% | 8/10 = 80% |
| 12 | 6 | + | 8/12 = 66% | 9/10 = 90% |
| 13 | 22 | - | 9/13 = 69% | 9/10 = 90% |
| 14 | 15 | - | 10/14 = 71% | 9/10 = 90% |
| 15 | 11 | - | 10/15 = 66% | 9/10 = 90% |
| 16 | 3 | + | 10/16 = 62% | 10/10 = 100% |
| 17 | 10 | - | 10/17 = 58% | 10/10 = 100% |
| 18 | 26 | - | 11/18 = 61% | 10/10 = 100% |
| 19 | 2 | - | 11/19 = 57% | 10/10 = 100% |
| 20 | 8 | - | 12/20 = 60% | 10/10 = 100% |
| 21 | 13 | - | 12/21 = 57% | 10/10 = 100% |
| 22 | 19 | - | 12/22 = 54% | 10/10 = 100% |
| 23 | 21 | - | 12/23 = 52% | 10/10 = 100% |
| 24 | 12 | - | 12/24 = 50% | 10/10 = 100% |
| 25 | 23 | - | 12/25 = 48% | 10/10 = 100% |
| 26 | 9 | - | 12/26 = 46% | 10/10 = 100% |
| 27 | 14 | - | 12/27 = 44% | 10/10 = 100% |

The obtained interpolated precisions at all 27 recall levels in Table 7 are shown as precision recall curve in Fig. 4.

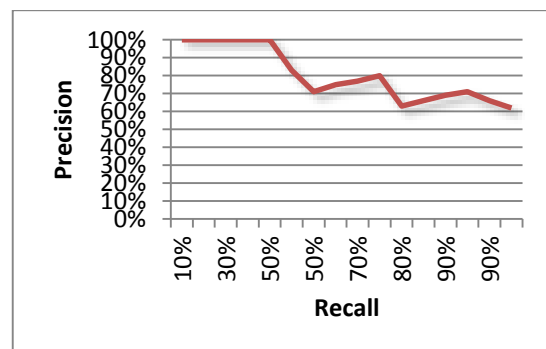


Fig.4. The precision-recall curve for document titled “evolution”

Table 8 shows the performance measures of the ten sample documents while the summary text rate is 40% of the original document.

Table 8. The performance of the proposed system by selecting the top-ranked 40% of sentences

| Document ID | Precision (%) | Recall (%) | Average Precision (%) |
|-------------|---------------|------------|-----------------------|
| 1 | 100 | 100 | 100 |
| 2 | 100 | 100 | 100 |
| 3 | 100 | 100 | 100 |
| 4 | 66 | 66 | 69.5 |
| 5 | 80 | 80 | 91 |
| 6 | 60 | 60 | 87 |
| 7 | 80 | 80 | 76.6 |
| 8 | 60 | 60 | 47.75 |
| 9 | 75 | 75 | 75 |
| 10 | 80 | 80 | 73 |

When we investigate the table 8, it is observed that, the computed precision and recall values of the documents varied in the range of 60% to 100%. While the precision and recall values are both 60%, the average precision value is 47.75% for 8th document. This can be explained as follows. The human expert selected the 1, 6, 8, 10 and 13th sentences as the summary set. In spite of this, the computer's ranked list considers 12, 13, 8, 10 and 7th sentences respectively. In this case, they agree on 13, 8 and 10th sentences for the summary text. Above all, the 12th sentence which is selected by the computer as the most relevant sentence to the summary is not in the human expert's selected set. Therefore the average precision value decreases considerably. Consequently, if the ranked list starts with right selected sentence and the longer continues like this, then the average precision increases. The exact opposite situation of this is seen when we examine 6th document. The f-measure value is 60% however; average precision value is 80% for this document. The reason for that is explained as follows. There are five computer selected sentences and three of them are in the first three entry fields of the ranked list. So the remainder two sentences are wrong selected.

Given the evaluation results it can be said that the proposed summarization method is a promising approach to create a valid summary of Turkish documents. The system produces admissible results when it is compared against a human generated summary.

When the number of sentences in the summary text changes, the precision and recall values are also changes. For example when we decrease the selecting rate of sentences from 40% to lower rates such as 30%, then the performances of the summarization process for all documents are change as in Table 9.

Android is the preferred platform for our mobile application due to widely usage among the people. Today the processors of mobile devices cannot process some algorithms which are computationally expensive such as the NLP algorithms with a good performance. Therefore the applied summarization methods are realized by a web service mechanism; and the mobile client communicates with the summarization system by this web service which is using SOAP (Simple Object Access Protocol). For the

exchange of data between the mobile application and the web service the Android kSOAP2 library is used. The web service is hosted by using Apache Axis2 in Tomcat Web Server. The mobile interface and an output screen of the summarization application are shown in Fig.5.

Table 9. The performance of the proposed system by selecting the top-ranked 30% of sentences

| Document ID | Precision (%) | Recall (%) | Average Precision (%) |
|-------------|---------------|------------|-----------------------|
| 1 | 100 | 100 | 100 |
| 2 | 100 | 100 | 100 |
| 3 | 100 | 100 | 100 |
| 4 | 77 | 77 | 76.4 |
| 5 | 75 | 75 | 94 |
| 6 | 100 | 100 | 100 |
| 7 | 100 | 100 | 100 |
| 8 | 66 | 66 | 38.6 |
| 9 | 100 | 100 | 100 |
| 10 | 75 | 75 | 71.8 |

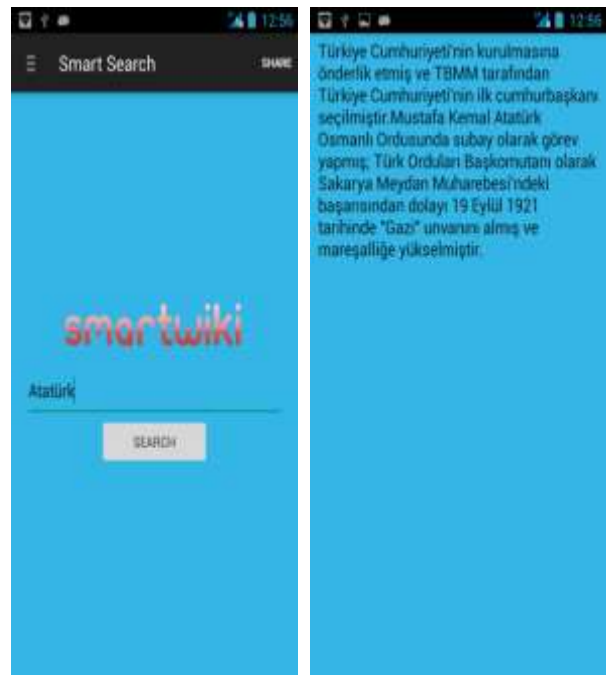


Fig.5. The mobile interface and the output screen

V. CONCLUSIONS

In this paper, we present a mobile Turkish text summarization system which combines some structural and semantic features of the sentences to produce accurate summarization results. The evaluated features of a sentence are considered in three main parts: structural features such as length, position and title information of the sentence, LSA based semantic features and Wikipedia based semantic features.

The LSA based cross method, the Helmholtz principle based meaningful word-set method and the Wikipedia based semantic scoring methods are combined to constitute the semantic features of the sentences. One of the contributions of the paper is to combine these methods in a hybrid structure for text summarization

problem. Another contribution is that, we implement an existing performance evaluation method for text summarization process. In this context, the average precision and breakeven point computed by using ranking of the sentences are used as evaluation metrics for the document summaries. By this consideration, we not only obtain the summary sentences but also the ranking of them due to their relevance scores. Other than these, this is the first time a mobile application summarizes Turkish Wikipedia documents by using the presented methods.

In this study, while combining features, AHP generates the feature weights depending on an expert judgment. Different feature weights influence the summary results directly. Therefore the analytic network process (ANP) which is an alternative method used in multi-criteria decision analysis is also tried. However, we couldn't obtain promising results as obtained in AHP method. The applied manual process for combining features should be change by an optimization model and optimal feature weights can be computed automatically as a feature work.

REFERENCES

- [1] Balinsky A., Balinsky H. and Simske S., On the Helmholtz principle for documents processing, *In: Proceedings of the 10th ACM Document Engineering*, 2010.
- [2] Balinsky A., Balinsky H. and Simske S., On the Helmholtz principle for data mining, *In: Proceedings of the Conference on Knowledge Discovery*, 2011a.
- [3] Balinsky A., Balinsky H. and Simske S., Rapid change detection and text mining, *In: Proceedings of the 2nd Conference on Mathematics in Defence*, 2011b.
- [4] Balinsky H., Balinsky A. and Simske, S. Document sentences as a small world, *In: Proceedings of the IEEE International Conference on Systems*, 2011c, pp. 2583-2588.
- [5] Bawakid A., and Oussalah M., Summarizing with Wikipedia, *In Proceedings of the text analysis conference*, 2010.
- [6] Binwahlan M.S., Salim N. and Suanmali L., Fuzzy swarm diversity hybrid model for text summarization, *Information Processing and Management*, 46 (2010), 571-588.
- [7] Cıgır C., Kutlu M. and Cicekli I., Generic Text Summarization for Turkish, *The Computer Journal*, 53, 8 (2010), 1315-1323.
- [8] Collaboration and Decision, Support Software for Groups and Organizations 2011-2012, available at: <http://www.expertchoice.com/>.
- [9] Deerwester S.C., Dumais S.T., Landauer T.K., Furnas G.W. and Harshman R.A. Indexing by latent semantic analysis, *Journal of the American Society of Information Science*, 41, 6 (1990), 391-407.
- [10] Gong S., Qu Y., and Tian S., Summarization using Wikipedia, *In Proceedings of text analysis conference*, 2010.
- [11] Güran A., Bayazit N.G. and Bekar E., Automatic Summarization of Turkish Documents Using Non-negative Matrix Factorization, *Innovations in Intelligent Systems and Applications (INISTA)*, IEEE, 2011, pp. 480 - 484.
- [12] Güran A., Bayazit N.G. and Gürgüz M.Z., Efficient feature integration with Wikipedia-based semantic feature extraction for Turkish text summarization, *Turkish Journal of Electrical Engineering & Computer Sciences*, 21 (2013), 1411 - 1425.
- [13] Jiang Y., Zhang X., Tang Y. and Nie R., Feature-based approaches to semantic similarity assessment of concepts using Wikipedia, *Information Processing and Management*, 51 (2015), 215-234.
- [14] L'Huillier G., Hevia A., Weber R. and Rios S., Latent Semantic Analysis and Keyword Extraction for Phishing Classification, *IEEE International Conference on Intelligence and Security Informatics*, 2010, pp. 129 - 131.
- [15] Liu B. *Web Data Mining Exploring Hyperlinks, Contents, and Usage Data*, First Edition, Springer, 2006.
- [16] Lowe D., *Perceptual Organization and Visual Recognition*, Amsterdam: Kluwer Academic Publishers, 1985.
- [17] Mashechkin I.V., Petrovskiy M.I., Popov D.S. and Tsarev D.V., Automatic text summarization using latent semantic analysis, *Programming and Computer Software*, 37 (2011), 299-305.
- [18] Miao Y. and Li C., WikiSummarizer - A Wikipedia-based summarization system, *In Proc. text analysis conference (TAC)*, 2010.
- [19] Ozsoy M.G., Cicekli I. and Alpaslan F.N., Text Summarization of Turkish Texts using Latent Semantic Analysis, *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 869-876.
- [20] Pourvali M., and Abadeh M. S., A new graph based text segmentation using Wikipedia for automatic text summarization, *International Journal of Advanced Computer Science and Applications*, 3,1 (2012a).
- [21] Pourvali M. and Abadeh M. S., Automated text summarization base on lexicales chain and graph using of WordNet and Wikipedia knowledge base, *International Journal of Computer Science Issues*, 9, 1 (2012b).
- [22] Radev D.R., Hovy E. and McKeown K., Introduction to the special issue on summarization, *Computational Linguistics*, 28, 4 (2002), 399-408.
- [23] Ramanathan K., Sankarasubramaniam Y., Mathur N. and Gupta, A, Document summarization using Wikipedia, *In Proceedings of the first international conference on intelligent human computer interaction*, Springer, 2009, pp. 254-260.
- [24] Saaty T.L., A Scaling Method for Priorities in Hierarchical Structures, *Journal of Mathematical Psychology*, 15 (1997), 57-68.
- [25] Sankarasubramaniam Y., Ramanathan K. and Ghosh S., Text summarization using Wikipedia, *Information Processing and Management*, 50 (2014), 443-461.
- [26] Sakhare and Kumar, Syntactic and Sentence Feature Based Hybrid Approach for Text Summarization, *I.J. Information Technology and Computer Science*, 03, (2014), 38-46.
- [27] Steinberger and Jezek, Latent Semantic Analysis in Text Summarization and Summary Evaluation, *Proceedings of ISIM*, 2004, pp. 93-100.
- [28] Utkin L.V. and Simanova N.V., The DS/AHP Method Under Partial Information About Criteria and Alternatives by Several Levels of Criteria, *International Journal of Information Technology & Decision Making*, 11, 2, (2012), 307-326.
- [29] Wong K-F., Wu M. and Li W., Extractive Summarization Using Supervised and Semi-supervised Learning, *Proceedings of the 22nd International Conference on Computational Linguistics*, (2008), pp. 985-992.
- [30] Yeh J-Y., Ke H-R, Yang W-P and Meng I-H., Text summarization using a trainable summarizer and latent

semantic analysis, *Information Processing and Management*, 41 (2005), 75–95.

[31] Zemberek website: <http://code.google.com/p/zemberek/>.

Authors' Profiles



Akif Hatipoglu was born in Balıkesir in Turkey at 27th may 1993. He has graduated from Kocaeli University Computer Engineering Department in July 2015. He is still working as a software developer in Yapı Kredi Bank in İstanbul/Turkey.



Sevinç İlhan Omurca was born in Burdur in Turkey at 23th December 1979. She is an assistant professor at the Kocaeli University Computer Engineering Department in Turkey. She has a Ph.D. at the Kocaeli University Electronics and Communications Engineering.

How to cite this paper: Akif Hatipoglu, Sevinç İlhan Omurca, "A Turkish Wikipedia Text Summarization System for Mobile Devices", *International Journal of Information Technology and Computer Science(IJITCS)*, Vol.8, No.1, pp.1-10, 2016. DOI: 10.5815/ijitcs.2016.01.01