Using the F-measure as similarity measure for automatic text summarization

R. M. Aliguliyev
Institute of Information Technology
of the National Academy of Sciences of Azerbaijan, Baku
e-mail: a.ramiz@science.az

Цель статьи — показать, что качество реферирования текстов непосредственно зависит от выбора меры подобия, и значит, использование F-меры как меры подобия является эффективным подходом. Эффективность выбора F-меры в качестве меры подобия подтверждена экспериментальным путем. Эксперименты показывают, что использование F-меры, с точки зрения точности реферирования, дает лучшие результаты, чем мера косинуса.

Introduction

Text mining is a research area that is currently extremely active. Automatic text summarization is one important task in this field. Automatic text summarization plays an important role in information retrieval (IR). The technology of automatic text summarization is maturing and may provide a solution to the information overload problem [1, 2]. With a large volume of texts, presenting the user with a summary of each document greatly facilitates the task of finding the desired documents. Similar to the tasks of IR, automatic text summarization can be regarded as how to find out salient pieces (here piece could be a phrase, sentence, or a paragraph of the document) from a document. Its goal is to include in that summary the most significant pieces in the text. In general, automatic text summarization takes an original text(s) as input, extracts the essence of the original text(s), and presents a well-formed summary to the user. Mani and Maybury [1] formally defined automatic text summarization as a process that produces a condensed version of its input's for user(s) consumption while preserving the main information content of source text(s).

1. Related Work

A variety of automatic text summarization methods have been proposed and evaluated. They can be broadly categorized into two approaches: abstraction and extraction. The goal of abstraction is to understand the text using knowledge-based methods and compose a coherent summary comparable to a human authored summary. This is very difficult to achieve with current natural language processing (NLP) techniques. In contrast to abstraction that requires heavy machinery from NLP, extraction can be easily viewed as the process of selecting salient excerpts from the source document [1]. Extraction systems analyze a source document using techniques derived from IR (e. g. frequency analysis and keyword identification) to determine significant sentences that constitute the summary.

A summary can also either be a user-oriented (or query-based) or generic [3]. A generic summary locates the main topics and key contents covered in the source text. A query-based summary locates the contents pertinent to user's seeking goals [2]. Query-based text summaries are useful for answering such questions as whether a given document is relevant to the user's query, and if relevant, which part(s) of the document is relevant. On the other hand, a generic summary provides an overall sense of the text's contents. A good generic summary should contain the main topics of the document while keeping redundancy to a minimum.

Sentence based summarization techniques are commonly used in automatic text summarization to produce extractive summaries [4–11]. The generic summarization methods that extract the most relevance sentences from the source document to form a summary in papers [7–10] are proposed. The proposed methods are based on clustering of sentences. Effective techniques for sentence extraction have been proposed in papers [4–6, 12]. The techniques first break a document into a list of sentences (paragraphs). Important sentences are then detected by some sentence weighting scheme, and the highly weighted sentences are selected to form a summary. A sentence weighting scheme can be variously formulated by employing many components and distributing them with different parameters. For example, Term Frequency, Sentence Order and Sentence Length are common components. The paper [13] focus on investigating and comparing effectiveness between Query Term Frequency (QTF) and Query Term Order (QTO). QTF in sentence weighting algorithm means the number of times the query terms appear in a sentence, and each term is equally weighted. QTO means the number of times the query terms appear in a sentence, with those terms appearing earlier in the query being assigned higher scores than those appearing later. Various criteria maybe used to associate importance with paragraphs, giving rise to different paths. To achieve automatic text summarization in paper [11] proposed two novel methods: modified corpus based approach (MCBA), and LSA-based TRM (Text Relationship Map) approach (LSA+TRM). The first is based on a score function combined with the analysis of salient features, and the genetic algorithm is employed to discover suitable combinations of feature weights. The second one exploits LSA and a TRM to derive semantically salient structures from a document. Both approaches concentrate on single-document summarization and generate extract-based summaries. The method TRM proposed by Salton et al. [12] is a graphical represent at, on of textual structure, in which paragraphs (in general, pieces of text) are represented by nodes on a graph and related paragraphs are linked by edges.

In this paper, we propose a simply and effective sentence extractive technique to achieve automatic text summarization. This method is based on evaluation of relevance score of sentence. The relevance score of each sentence is calculated in relation to all other sentences. We concentrate our presentation on choice a similarity measure.

2. Extractive Generic Summarization by Relevance Measure

Extractive summarization works by choosing a subset of the sentences in the original document. This process can be viewed as identifying the most salient sentences in a document, that give the necessary and sufficient amount of information related to the main theme of the document. To assess the importance of the sentences use a several similarity measures. One of the similarity measure widely used in text mining is the cosine measure. The cosine similarity between two sentences S_i and S_l is defined as:

$$\cos(S_i, S_l) = \frac{\sum\limits_{j=1}^{m} w_{ij} w_{lj}}{\sqrt{\sum\limits_{j=1}^{m} w_{ij}^2 \sum\limits_{j=1}^{m} w_{lj}^2}}, \quad i, l = 1, ..., n, \tag{1}$$

where w_{ij} is the term weight of the t_j in the sentence $S_i = (w_{i1}, w_{i2}, ..., w_{im})$, $i = 1, 2, ..., n$, $j = 1, 2, ..., m$.

A typical weighting scheme is to use the frequency-based formula, such as [14]:

$$w_{ij} = f_{ij} \log\left(\frac{n}{n_j}\right), \tag{2}$$

where f_{ij} — the number of occurrences of term t_j in the sentence S_i and n_j — the number of sentences containing the term t_j.

In our paper to evaluate the importance of the sentences we use classical IR precision and recall measures. To calculate the similarity measure each sentence must at first be represented in a suitable form. In our method, a sentence S_i is represented as bag-of-terms, instead of the term-based frequency vector. Let a document D is decomposed into individual sentences $D = (S_1, S_2, ..., S_n)$, where n is the number of sentences in a document D. Let $T = (t_1, t_2, ..., t_m)$ represent all the terms occurred in a document D, where m is the number of terms. Let a sentence S_i is represented as bag-of-terms $S_i = (t_1, t_2, ..., t_{m_i})$, where m_i is the number of terms in a sentence S_i.

The similarity between pair of sentences S_i and S_l is evaluated to determine if they are semantically related. The similarity between sentences S_i and S_l we define as

$$F(S_i, S_l) = \frac{2P(S_i, S_l)R(S_i, S_l)}{P(S_i, S_l) + R(S_i, S_l)}, \quad i \neq l = 1, 2, ..., n. \tag{3}$$

In formula (3) $P(S_i, S_l)$ and $R(S_i, S_l)$ are classical IR precision and recall measures, which we compute as follows:

$$P(S_i, S_l) = \frac{|S_i \cap S_l|}{|S_i|} = \frac{|S_i \cap S_l|}{m_i}, \quad i \neq l = 1, 2, ..., n, \tag{4}$$

$$R(S_i, S_l) = \frac{|S_i \cap S_l|}{|S_l|} = \frac{|S_i \cap S_l|}{m_l}, \quad i \neq l = 1, 2, ..., n, \tag{5}$$

where $|A|$ is the cardinality of a set A.

In view of (4) and (5), the formula (3) becomes:

$$F(S_i, S_l) = 2\frac{|S_i \cap S_l|}{m_i + m_l}, \quad i, l = 1, 2, ..., n. \tag{6}$$

Our approach to text summarization allows generic summaries by scoring sentences. Each sentence is scored according to the formula (7). The relevance score of S_i with regard to all sentences in a document$D$ as (based on F-measure), we compute as:

$$F_{\text{score}}(S_i) = \sum_{\substack{l=1 \\ l \neq i}}^{n} F(S_i, S_l), \quad i = 1, 2, ..., n. \tag{7}$$

Since the main purpose to show the effectiveness of the application of an F-measure as the similarity measure, that analogously we determine the relevance score of S_i with regard to all sentences in a document D (based on cosine measure) [4–6]:

$$C_{\mathrm{score}}(S_i) = \sum_{\substack{l=1 \\ l \neq i}}^{n} \cos(S_i, S_l), \quad i = 1, 2, ..., n. \tag{8}$$

Finally, as to selection of sentences to generate a summary all sentences are ranked according to their relevance scores calculated from formula (7) ((8)), and a designated number of top-weight sentences are picked out to form the summary.

Thus the generation summary process consist the following steps:

1. Decompose the document into individual sentences.
2. Represent each sentence as bag-of-terms.
3. Using the formulae described in (6) for each pair of sentences S_i and S_l compute the similarity measure.
4. Using the formula (7) ((8)) for each sentence S_i, compute the relevance score.
5. Rank all sentences according to their relevance score.
6. Starting with the sentence which has a highest relevance score the sentences add to the summary. If the compression rate (CR), which is defined as ratio of summary length to original length, reaches the predefined value, terminate the operation; otherwise, continue the process adding of the sentences to the summary.

The methods (7) and (8) are conditionally we call Method1 and Method2, respectively.

3. Experiments and discussion

In this section, we describe the experiment results to evaluate our text summarization algorithm. In our experiments, using human-generated and NewsInEssence-generated summaries, we employed four text-summarization methods — Method1, Method2, MS Word Summarizer and Copernic Summarizer [15]. The document collection used in this experiment consisted of fourteen documents, partitioned into two groups. The first group contained four documents (doc1. . .doc4), taken from http://oswinds.csd.auth.gr, www.actapress.com, and http://www.mitre.org. The second group contained ten news articles (news1. . .news10), were randomly selected from the NewsInEssence [16]. For first group we compared the summaries produced by methods Method1, Method2, MS Word Summarizer, and Copernic Summarizer against the human-generated summaries. For second group we compared the summaries produced by methods Method1, Method2, MS Word Summarizer, and Copernic Summarizer against the summaries produced by NewsInEssence Summarizer. These are an important point, since there is no standard measure of summary quality. To quote [2]: "Text summarization is still an emerging field, and serious questions remain concerning the appropriate methods and types of evaluation".

3.1. Preprocessing

Each document in the first group has been transformed to text format and the abstracts, keywords and references have been removed. One of the major problems in text mining is that a document can contain a very large number of words. If each of these words is represented as a vector coordinate, the number of dimensions would be too high for the text mining algorithm. Hence, it is crucial to apply preprocessing methods that greatly reduce the number of dimensions (words) to be given to the text mining algorithm. Our system can apply several preprocessing methods to the original documents, namely stemming and removal stopwords.

Stemming (i. e. removing word affixes such as 'ing', 'ion', 's') consists of converting each word to its stem, i. e. a natural form with respect to tag-of-speech and verbal/plural inflections. In essence, to get the stem of a word it is necessary to eliminate its suffixes representing tag-of-speech and/or verbal/plural inflections. We have used Porter's algorithm [17], originally developed for the English language.

Stopwords (i. e. insignificant words like 'can', 'in', 'this', 'from', 'then', 'or', 'the', 'by') are words that occur very frequently in a document. Since they are so common in many documents, they carry very little information about the contents of a document in which they appear.

3.2. Comparison between human-generated and automatically-generated summaries

Four independent professional evaluators were employed to conduct manual summarization. For each document doc1…doc4, each professional evaluator was requested to select 15 and 30 % sentences which (s)he deemed the most relevant for summarizing the document. Table 1 shows the statistics of the documents and the summarization results.

We employ the standard measures to evaluate the performance of summarization, i. e. precision, recall and F-measure. We assume that a human would be able to identify the most important sentences in a document most effectively. If the set of sentences selected by an automatic extraction method has a high overlap with the human-generated extract, the automatic method should be regarded as effective. Assume that S_{man} is the manual summary and S_{auto} is the automatically-generated summary, the measurements are defined as [14]:

$$P = \frac{|S_{\text{man}} \cap S_{\text{auto}}|}{|S_{\text{auto}}|}, \tag{9}$$

$$R = \frac{|S_{\text{man}} \cap S_{\text{auto}}|}{|S_{\text{man}}|}, \tag{10}$$

$$F = \frac{2PR}{P + R}. \tag{11}$$

The evaluation results are shown in Tables 2 and 3. Tables 2 and 3 show a summary of precision (P), recall (R) and F-measure (F) for each system, when CR is 15 and 30 %, respectively. The MS Word Summarizer reaches an average of 0.433 (0.511) P, 0.540 (0.538) R and 0.479 (0.524) F, when CR is 15 % (30 %). The Copernic Summarizer reaches an average of 0.514 (0.535) P, 0.540 (0.560) R and 0.527 (0.547) F, when CR is 15 % (30 %). Our approach using cosine measure (Method2) achieves an average of 0.530 (0.510) P, 0.560 (0.532) R and 0.544 (0.520) F, while using the F-measure as similarity measure achieves an average of 0.615 (0.614) P, 0.649 (0.666) R and 0.630 (0.652) F, when CR is 15 % (30 %). Interestingly, the Method1 gives the best results in both cases when CR is 15 and 30 %, than other methods. It can be observed that when F-measure is considered, on average, Method1 outperforms Method2 about 15.8 % F and 25.4 % F when CR is 15 and 30 %, respectively.

T a b l e 1. Statistics of the documents doc1…doc4 and summaries

Documents	Number of sentences in the										
	Documents	summaries created by summarizers									
		Human		Method2		MS Word		Copernic		Method1	
		CR =		**CR =**		**CR =**		**CR =**		**CR =**	
		15 %	**30 %**	**15 %**	**30 %**	**15 %**	**30 %**	**15 %**	**30 %**	**15 %**	**30 %**
doc1	158	23	47	23	47	29	50	23	47	23	47
doc2	151	22	45	23	45	23	43	24	48	23	45
doc3	111	17	34	17	34	21	36	17	34	17	34
doc4	195	27	55	32	64	40	63	30	61	32	64

T a b l e 2. Evaluation measures for automatic extraction methods, CR = 15 %

Documents	**Overlap with the human-generated extracts**											
	Method1			Method2			MS Word summarizer			Copernic summarizer		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
doc1	0.652	0.652	0.652	0.478	0.478	0.478	0.379	0.478	0.423	0.435	0.435	0.435
doc2	0.565	0.591	0.578	0.522	0.545	0.533	0.478	0.500	0.489	0.500	0.545	0.522
doc3	0.647	0.647	0.647	0.588	0.588	0.588	0.476	0.588	0.526	0.588	0.588	0.588
doc4	0.594	0.704	0.644	0.531	0.630	0.576	0.400	0.593	0.478	0.533	0.593	0.561
avg.	0.615	0.649	0.630	0.530	0.560	0.544	0.433	0.540	0.479	0.514	0.540	0.527

T a b l e 3. Evaluation measures for automatic extraction methods, CR = 30 %

Documents	**Overlap with the human-generated extracts**											
	Method1			Method2			MS Word summarizer			Copernic summarizer		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
doc1	0.617	0.617	0.617	0.447	0.447	0.447	0.400	0.426	0.412	0.383	0.383	0.383
doc2	0.644	0.644	0.644	0.533	0.533	0.533	0.558	0.533	0.545	0.583	0.622	0.602
doc3	0.676	0.676	0.676	0.529	0.529	0.529	0.611	0.647	0.629	0.618	0.618	0.618
doc4	0.625	0.727	0.672	0.531	0.618	0.571	0.476	0.545	0.508	0.557	0.618	0.586
avg.	0.641	0.666	0.652	0.510	0.532	0.520	0.511	0.538	0.524	0.535	0.560	0.547

T a b l e 4. Performance evaluation (*F*-measure) compared between Method1 and other methods

Documents	**CR = 15 %**			**CR = 30 %**		
	Method2	MS Word	Copernic	Method2	MS Word	Copernic
doc1	36.4 (+)	54.1 (+)	49.9 (+)	38.0 (+)	49.8 (+)	61.1 (+)
doc2	8.4 (+)	18.2 (+)	10.7 (+)	20.8 (+)	18.2 (+)	7.0 (+)
doc3	10.0 (+)	23.0 (+)	10.0 (+)	27.8 (+)	7.5 (+)	9.4 (+)
doc4	11.8 (+)	34.7 (+)	14.8 (+)	17.7 (+)	32.3 (+)	14.7 (+)
avg.	15.8 (+)	31.5 (+)	19.5 (+)	25.4 (+)	24.4 (+)	19.2 (+)

Hereafter, we use relative improvement $\frac{(\text{Method1} - \text{other methods})}{\text{other methods}} \cdot 100$ for comparison. Table 4 reports the performance results compared between Method1 and other methods.

3.3. Comparison between NewsInEssence-generated and automatically-generated summaries

We assume that the NewsInEssence summarizer would be able to identify the most relevant sentences in a document most effectively. For each news article news1…news10 by NewsInEssence summarizer were created two summaries, at CR = 20 and CR = 30 %. Table 5 shows the statistics of the news articles and the summarization results. If the set of sentences selected by an automatic extraction method has a high overlap with the NewsInEssence-generated extract, the automatic method should be regarded as effective. Assume that S_{NIE} is the NewsInEssence-generated summary and S_{auto} is the automatically generated summary, the measurements are defined as:

$$P = \frac{|S_{\text{NIE}} \cap S_{\text{auto}}|}{|S_{\text{auto}}|}, \tag{12}$$

$$R = \frac{|S_{\text{NIE}} \cap S_{\text{auto}}|}{|S_{\text{man}}|}, \tag{13}$$

$$F = \frac{2PR}{P+R}. \tag{14}$$

Tables 6 and 7 show the evaluation results. When CR is 20 and 30 % the MS Word Summarizer reaches an average of 0.396 and 0.410 P, 0.436 and 0.444 R, 0.413 and 0.422 F, respectively. The Copernic Summarizer reaches an average of 0.479 (0.567) P, 0.425 (0.512) R and 0.447 (0.532) F, when CR is 20 % (30 %). The Method2 achieves an average of 0.376 (0.410) P, 0.360 (0.402) R and 0.367 (0.404) F, and the Method1 achieves an average of 0.512 and 0.575 P, 0.488 and 0.583 R, 0.498 and 0.575 F, when CR is 20 and 30 %, respectively. It can be observed that when F-measure is considered, on average, Method1 outperforms Method2 about 35.7 % F and 42.3 % F when CR is 20 and 30 %, respectively. Table 8 gives the performance results compared between Method1 and other methods. In the Tables 4 and 8 "+" means the result outperforms and "–" means the opposite.

T a b l e 5. Statistics of the documents news1…news10 and summaries

News articles	Number of sentences in the										
	News articles	summaries created by summarizers									
		NewsInEssence		Method2		MS Word		Copernic		Method1	
		CR =		**CR =**		**CR =**		**CR =**		**CR =**	
		20 %	**30 %**	**20 %**	**30 %**	**20 %**	**30 %**	**20 %**	**30 %**	**20 %**	**30 %**
news1	26	5	8	5	8	6	12	5	8	5	8
news2	36	7	10	7	10	9	12	7	10	7	10
news3	13	3	4	3	4	3	4	2	3	3	4
news4	13	3	5	3	4	3	5	2	3	3	4
news5	39	8	12	8	12	9	12	7	11	8	12
news6	36	7	10	7	11	9	13	7	10	7	11
news7	40	9	14	8	12	10	14	8	12	8	12
news8	25	5	7	5	8	7	10	5	8	5	8
news9	26	5	7	5	8	7	9	5	7	5	8
news10	26	7	11	5	8	5	8	5	7	5	8

T a b l e 6. Evaluation measures for automatic extraction methods, CR = 20 %

News articles	Overlap with the NewsInEssence-generated extracts											
	Method1			Method2			MS Word summarizer			Copernic summarizer		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
news1	0.800	0.800	0.800	0.400	0.400	0.400	0.500	0.600	0.545	0.400	0.400	0.400
news2	0.429	0.429	0.429	0.572	0.572	0.572	0.333	0.429	0.392	0.714	0.714	0.714
news3	0.667	0.667	0.667	0.333	0.333	0.333	0.667	0.667	0.667	0.500	0.333	0.400
news4	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.500	0.333	0.400
news5	0.375	0.375	0.375	0.375	0.375	0.375	0.333	0.375	0.353	0.286	0.250	0.267
news6	0.286	0.286	0.286	0.572	0.572	0.572	0.222	0.286	0.250	0.286	0.286	0.286
news7	0.625	0.556	0.589	0.375	0.333	0.353	0.400	0.444	0.421	0.500	0.444	0.470
news8	0.400	0.400	0.400	0.200	0.200	0.200	0.143	0.200	0.167	0.600	0.600	0.600
news9	0.600	0.600	0.600	0.200	0.200	0.200	0.429	0.600	0.500	0.600	0.600	0.600
news10	0.600	0.429	0.500	0.400	0.286	0.334	0.600	0.429	0.500	0.400	0.286	0.334
avg.	0.512	0.488	0.498	0.376	0.360	0.367	0.396	0.436	0.413	0.479	0.425	0.447

T a b l e 7. Evaluation measures for automatic extraction methods, CR = 30 %

News articles	Overlap with the NewsInEssence-generated extracts											
	Method1			Method2			MS Word summarizer			Copernic summarizer		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
news1	0.625	0.625	0.625	0.625	0.625	0.625	0.333	0.500	0.400	0.375	0.375	0.375
news2	0.400	0.400	0.400	0.400	0.400	0.400	0.417	0.500	0.455	0.500	0.500	0.500
news3	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.667	0.500	0.572
news4	0.600	0.750	0.667	0.500	0.400	0.444	0.400	0.400	0.400	0.667	0.400	0.500
news5	0.583	0.583	0.583	0.417	0.417	0.417	0.500	0.500	0.500	0.636	0.583	0.608
news6	0.455	0.500	0.476	0.364	0.400	0.381	0.231	0.300	0.261	0.400	0.400	0.400
news7	0.583	0.500	0.538	0.167	0.143	0.154	0.429	0.429	0.429	0.500	0.429	0.462
news8	0.625	0.714	0.667	0.250	0.286	0.267	0.111	0.143	0.125	0.500	0.572	0.534
news9	0.625	0.714	0.667	0.500	0.571	0.533	0.556	0.714	0.625	0.857	1.000	0.923
news10	0.750	0.545	0.631	0.375	0.273	0.316	0.625	0.455	0.527	0.571	0.364	0.446
avg.	0.575	0.583	0.575	0.410	0.402	0.404	0.410	0.444	0.422	0.567	0.512	0.532

T a b l e 8. Performance evaluation (*F*-measure) compared between Method1 and other methods

News articles	**CR = 15 %**			**CR = 30 %**		
	Method2	MS Word	Copernic	Method2	MS Word	Copernic
news1	100.0 (+)	46.8 (+)	100.0 (+)	0.0	56.2 (+)	66.7 (+)
news2	−25.0 (−)	9.4 (+)	−39.9 (−)	0.0	−12.1 (−)	−20.0 (−)
news3	100.3 (+)	0.0	66.8 (+)	0.0	0.0	−12.6 (−)
news4	0.0	0.0	−16.8 (−)	50.2 (+)	66.8 (+)	33.4 (+)
news5	0.0	6.2 (+)	40.4 (+)	39.8 (+)	16.6 (+)	−4.1 (−)
news6	−50.0 (−)	14.4 (+)	0.0	24.9 (+)	82.4 (+)	19.0 (+)
news7	66.9 (+)	39.9 (+)	25.3 (+)	249.4 (+)	25.4 (+)	16.5 (+)
news8	100.0 (+)	139.5 (+)	−33.3 (−)	149.8 (+)	433.6 (+)	24.9 (+)
news9	200.0 (+)	20.0 (+)	0.0	25.1 (+)	6.7 (+)	−27.7 (−)
news10	49.7 (+)	0.0	49.7 (+)	99.7 (+)	19.7 (+)	41.5 (+)
avg.	35.7 (+)	20.6 (+)	11.4 (+)	42.3 (+)	36.3 (+)	8.1 (+)

Conclusion

In this paper, we propose a practical approach for extracting the most relevant sentences from the original document to form a summary. The proposed text summarization method creates generic summaries by scoring and extracting sentences from the source documents. For sentence scoring most summarization systems use cosine measure as well a similarity measure. The idea of our approach is to exploit the classical IR precision and recall measures as similarity measure. In this paper, we show that using the classical IR precision and recall measures as similarity measure is a viable and effective technique. We provide experimental evidence that our approach achieves reasonable performance. The experimental result shows that the similarity measure may bias the score, and make the summarizer misjudge the importance of sentences. The effect of F-measure as similarity measure in text summarization is illustrated with an example shown in Tables 4 and 8. The experiments justify our assumption that the relevance score of a sentence directly depends on choice of similarity measure. We conclude that F-measure can be employed as similarity measure to promote text summarization.

References

[1] Mani I., Maybury M.T. Advances in automated text summarization. Cambridge: MIT Press, 1999. 442 p.

[2] Zhang Y., Zincir-Heywood N., Milios E. World Wide Web site summarization // International J. of Web Intelligence and Agents Systems. 2004. Vol. 2, N 1. P. 39–53.

[3] Afantenos S., Karkaletsis V., Stamatopoulos P. Summarization from medical documents: a survey // Artificial Intelligence in Medicine. 2005. Vol. 33, N 2. P. 157–177.

[4] Alguliev R.M., Aliguliyev R.M. Effective summarization method of text documents // Proc. of the 2005 IEEE/WIC/ACM Intern. Conf. on Web Intelligence. France: Compiegne, 2005. P. 264–271.

[5] Alguliev R.M., Aliguliyev R.M. Improvement of Text Information Classification by Definition Importance of Sentences // Information Technologies. 2006. N 3. P. 62–68 (in Russian).

[6] Alguliev R.M., Aliguliyev R.M. A new summarization method of text documents and evaluation of classification result in three aspects // Telecommunications. 2006. N 3. P. 7–16 (in Russian).

[7] Alguliev R.M., Aliguliyev R.M., Bagirov A.M. Global optimization in the summarization of text documents // Automatic Control and Computer Sci. N.Y.: Allerton Press, Inc. 2005. Vol. 39, N 6. P. 42–47.

[8] Alguliev R.M., Aliguliyev R.M. Summarization of text-based documents with a determination of latent topical sections and information-rich sentences // Automatic Control and Computer Sci. N.Y.: Allerton Press, Inc. 2007. Vol. 41, N 3. P. 132–140.

[9] Aliguliyev R.M. A novel partitioning-based clustering method and generic document summarization // Proc. of the 2006 IEEE/WIC/ACM Intern. Conf. on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops) (WI-IATW'06). China: Hong Kong, 2006. P. 626–629.

[10] Aliguliyev R.M. Automatic document summarization by sentence extraction // Comput. Technologies. 2007. Vol. 12, N 5. P. 5–15.

[11] Yeh J.-Y., Ke H.-R., Yang W.-P., Meng I.-H. Text summarization using a trainable summarizer and latent semantic analysis // Information Processing and Management. 2005. Vol. 41, N 1. P. 75–95.

[12] Salton G., Singhal A., Mitra M., Buckley C. Automatic text structuring and summarization // Information Processing and Management. 1997. Vol. 33, N 2. P. 193–207.

[13] Liang S.F., Devlin S., Tait J. Investigating sentence weighting components for automatic summarization // Information Processing and Management. 2006. Vol. 43, N 1. P. 146–153.

[14] Baeza-Yates R., Ribeiro-Neto R. Modern Information Retrieval. N.Y.: Addison Wesley. ACM Press, 1999. 513 p.

[15] :(CopernicSummarizer: http://www.copernic.com//en/products/summarizer)

[16] :(NewsInEssence Summary: http://lada.si.umich.edu)

[17] Porter M. An algorithm for suffix stripping // Program. 1980. Vol. 14, N 3. P. 130–137.

Received for publication 11 October 2007