Importance of Surface Methods in Human and Automatic Text Summarization

Nives Mikelić Preradović, Damir Boras, Marta Vlainić

***Abstract*—** Both human and automatic summaries enable a concise display of the most important information from the original text. Summaries written by the author of the document, expert in the field, professional summarizer or generated by the automatic summarization system use the same shallow feature of the text (such as word frequency or location) to create a high-quality summary. In this paper, we describe these features and compare summary written by human with a summary created by automatic text summarization systems: Microsoft Word, SweSum, SHVOONG and Online Brevity Document Summarizer. Research results show that although all these automatic summarizers rely heavily and only on the shallow features of the text, they all generate informative extracts satisfying quality expectations of the human users.

***Keywords*—** human summary, automatic summary, surface methods of summarization, Microsoft Word, SweSum, SHVOONG, Online Brevity Document Summarizer

I. Introduction

Text summarization represents a method of extracting relevant portions of the input document, presenting the main ideas of the original text. It is a process of condensing a source document into its shorter version preserving the information content. Automatic text summarization borrows techniques from information retrieval (IR) and information extraction (IE) and has been under development for many years.

Recently, it has received more attention due to the increased use of the Internet.

For example, it is used for summarizing news to SMS or WAP-format for mobile phones or PDAs or for TTS (text to speech) systems, where the summarized text is synthetically read to the user.

It is also used for compressed descriptions of the search results in the search engines as well as in keyword directed subscriptions of news which are summarized and sent to the user. Summaries are also useful for orientation (e.g. Google Maps) and in a decision making process (e.g. TV guide). Finally, summaries are utilized for the document retrieval in a foreign language to obtain an automatically translated summary of the automatically summarized text.

Automatic summary can contain full sentences, but may also include only key words, pictures or diagrams. Different types of summaries exist: indicative or informative summary, extract or abstract, generic summary, query-oriented summary [1] or topic-focused summary [2].

Moreover, there are background summaries or just-the-news summaries, as well as single-document summaries or multi-document summaries [3].

Furthermore, user-focused summaries consider information needs of an individual user, while storing parameter values of his needs in the user's profile [4].

Already in 1982 Rowley [5] proposed the typology containing seven different types of summaries:

- *extract* (set of paragraphs that represent a summary of the original document)
- - *summary* (occurs at the end of the paper and concisely conveys the most important discoveries)
- *abridgement* (reduction of the document through which secondary facts or opinions are left out)
- *precis* (display of the main points of the discussion)
- *digest* (book summary or a summary of a newspaper article)
- *highlight* (comment integrated into the parts of the document to alert the reader) and
- *synopsis* (in cinematography, summary which presents content of the movie).

Summaries can be created by the authors of the documents, experts in the field or summarizing professionals. According to Hovy & Lin [6], the unsophisticated methods of automatic summarization were developed already in late 1950's and early 1960's.

The authors (ibid) state that these methods primarily relied on the automated counting of words (repeated words were mostly considered relevant and were included in the summary) and the position of sentences within the text.

After several decades and a large increase in computer speed and memory, as well as in automatic Natural Language Processing (NLP) techniques, automatic summarization of text started to grow in a rapid manner.

Regarding the automatic summarization, the most common summarization systems use either statistical techniques [7] or linguistic approach [8, 9].

There are also approaches which use a kind of linear combination of the above mentioned techniques [10, 11, 12, 13, 14, 15].

Marta Vlainić is a PhD candidate at the Department of Information and Communication Sciences, University of Zagreb, Zagreb, Croatia (email: marta.vlainic@gmail.com)

Nives Mikelić Preradović is an assistant professor at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia (nmikelic@ffzg.hr)

Damir Boras is a dean of the Faculty of Humanities and Social Sciences and full professor at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia (dboras@ffzg.hr)

Generally, automatic text summarization methods can be classified into three levels: surface methods, entity level and discourse level methods.

Surface level methods process the document source to locate the relevant sentences based on shallow text features, treating the most frequent words and the best positioned sentences as the most relevant.

On the other hand, entity level methods model semantic, syntactic and logical relations between entities in the text, trying to obtain the deep understanding of the text. Finally, the discourse level methods reffer to modeling of the global text structure, requiring a knowledge-rich approach to text processing.

Liu [16] proposes a framework for text summarization (Fig. 1), where the first phase encompasses the extraction of text segments using only the shallow text features resulting in a generic or query-specific summary, while the second phase consists of an intensive text analysis and language understanding process.

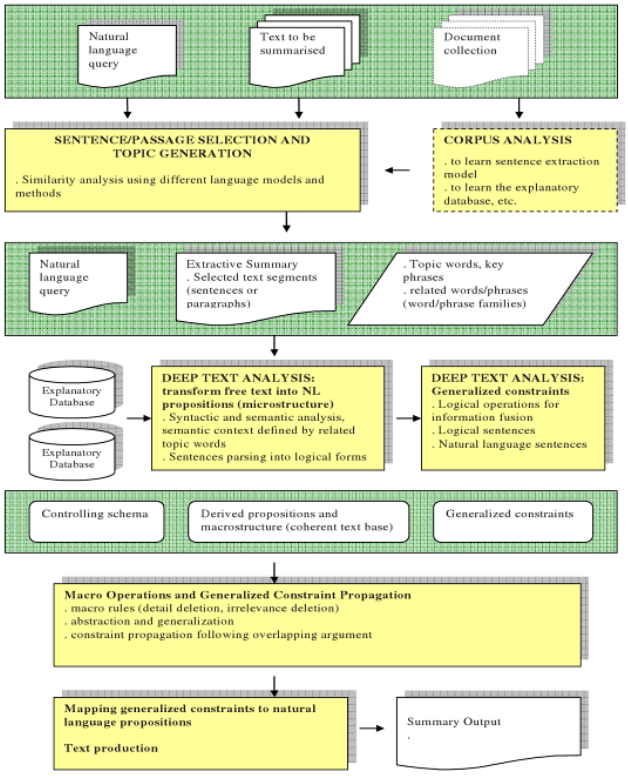

Fig 1. A Framework for Text Summarization

Surface level summarization methods were the most popular since the 1950's, relying on the frequency of words [17], sentence position, words in the title or to the presence of cue phrases in text [18].

The word frequency method (or keyword method) assigns a higher score to the sentences cointaning uncommon words (statistically salient terms), that appear frequently in the document (where frequency is larger than a certain threshold).

The keyword method used by Edmundson [18] assigns a score to each word in the text equal to its frequency, calculating the sentence score by summing the scores of words it contains.

The other keywod method used by Luhn [17] additionaly takes into account a relative position of the relevant words, specifying a maximum number of irrelevant words.

The method selects a range of words in a sentence with a relevant word at the beginning and at the end; the range must not containg more irrelevant words than the given threshold. The sentence score is calculated as a square of the number of relevant words divided by the total number of words in the range.

The sentence position method [18] assumes that the sentences which appear very early or very late in the source text and its sections tend to be more relevant.

Additionaly, the appearance of relevant sentences at the beginning and/or the end of the text depends on the genre of the source text. For example, in scientific articles, relevant information can be found in the specific sections of the document, such as the introductory part and conclusion, while the newspaper articles usually contain the most important information in the very beginning of the text.

It is generally believed that the title of the source text and the top two positions of the sentences are covering 60% of the document's subject matter. Therefore, the title method relies heavily on headings and subheadings of the source text, assigning a positive score to a sentence based on the occurrence of its words in the title or in one of the document's headings.

According to the Edmundson's research [18], the usefulness of words in titles and subtitles in the automatic summarization systems is statistically proven. Words in titles are considered relevant, as well as bolded words and phrases. Therefore, the summrization systems add points to the sentences containing them.

The cue method assumes the relevance of a sentence based on the presence of specific words and phrases that indicate either the positive relevance (bonus words) or the negative relevance of a sentence (stigma words).

The bonus words may be adjective comparatives, superlatives or expressions such as "it is important to note that", "in conclusion", etc.

The stigma words or phrases indicate that sentences containing them should be ranked lower and usually have the form of negations, anaphora or expressions such as "for example", "difficult", "impossible" and so on.

Apart from bonus and stigma words, the Cue method also distinguishes the null words (numbers, pronouns, prepositions, adjectives) that are considered irrelevant for the and do not influence the sentence score.

Entity level summarization methods [10, 19, 20, 21, 22, 23] model text entities and the semantic relationships between these entities.

The relations between entities are based on similarity, proximity and cohesion. Furthermore, similarity refers to words in the text that have similar stem (e.g. words *white* and *whiteness*), while proximity refers to the distance between the text units in which the entity occurs.

Cohesion refers to the connection between relevant units of text which contain entities strongly connected into a semantic structure. Connections between sections, sentences and phrases are made with the grammatical and lexical links.

Cohesion provides mechanical connections on the language level and ensures that the text makes sense as a whole.

Furthermore, the discourse level summarization methods [24, 25, 26] model the document's global structure and its relation to the communicative goals, taking into account the rhetorical structure of the text (argumentation and/or narrative structure).

Finally, the purpose of automatic text summarization is to provide concise and accurate content representation of the source document. Automatic text summarizers extract the most relevant information from the information source, summarizing them and constructing output according to the user's needs.

II. Methodology

Providing a quick overview of the document's content, summaries save time enabling a reader to easily decide which documents require his detailed attention. Along with human authors and summarization experts, automatic document summarizers are also able to create a high-quality summary.

In this paper, authors have compared some of the techniques used in both human and automatic summarization, such as using of the shallow text features, local context and heuristics. Regarding the shallow text features, both humans and automatic systems rely on the average word length, average sentence length and absolute number of words in the text.

As to the local context of the words and sentences, they both analyse the absolute and relative position of the sentences in the paragraph as well as paragraphs in text. Finally, regarding the heuristic features, they rely on number of words that start with an uppercase letter, number of words written in all-caps, number of date and time tokens, etc.

In this chapter we bring the results of the comparision of the automatically generated summary in Microsoft Word to the human summary of the source text, generated by authors. Additionaly, we have compared the four machine-generated summaries (created by Microsoft Word, SweSum, SHVOONG and Online Brevity Document Summarizer) respectively.

In this way, we have determined which surface methods were used by automatic text summarizers, whether created summaries show coherence, if they have maintained the cohesion and which of the summaries is of the best quality.

III. Comparison of the Human Summary and the Automatic Summaries

Microsoft Office Word Summarizer 2007 determines key points in a text by analyzing the document and assigning a score to each sentence (Fig. 2).

It assigns a higher score to sentences that contain words used frequently in the document.

User can then choose a percentage of the highest-scoring sentences to display in the summary and select whether to highlight the key points in a document, insert an executive summary or abstract at the top of a document, create a new document and put the summary there, or hide everything but the summary.

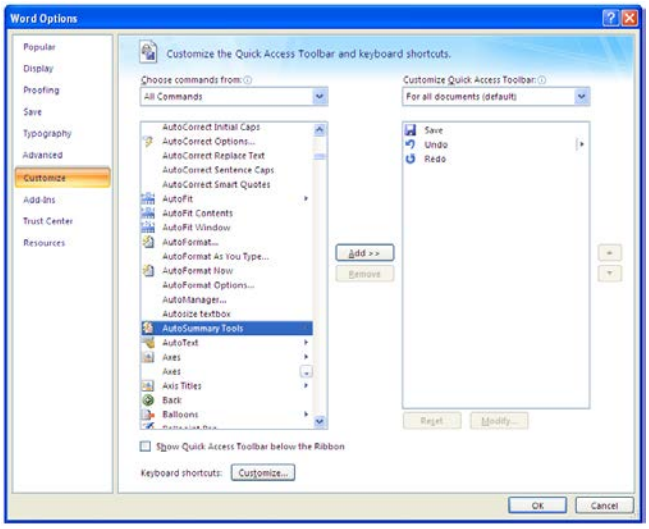

Fig. 2. Microsoft Office Word Summarizer 2007

SweSum [27] is a web-based text summarizer developed at the Royal Institute of Technology (KTH). It uses text extraction based on statistical and linguistic as well as heuristic methods to obtain text summarization and its domain is Swedish HTML-tagged newspaper text (Fig. 3).

Fig. 3. SweSum Text Summarization system

Since the newspaper text contains the most relevant information at the very top and since newspaper text even tends to be written to be cuttable from the bottom, SweSum assigns a higher score to the sentences that belong to the beginning of the text.

It also utilizes several different topic identification schemes (the bold tag which is often used to emphasize contents and headings). Furthermore, it scores higher the sentences that contain keywords, where keyword is defined as an open class word with a high term frequency (tf).

SweSum is available for Swedish, Norwegian, Danish, Spanish, English, French and German. It has been evaluated and its performance for English, Danish, Norwegian and

Swedish is considered to be state-of-the-art. The French, German and Spanish versions are in the prototype states.

Online Brevity Document Summarizer [28] was developed by Lextek International Company that specializes in full-text search technologies and generation of the text summaries (Fig. 4).

Brevity generates document extracts of the desired length and can also be used to highlight key sentences or words in the document.

According to its designers, the key benefits of Brevity are: accurate generation of automated document summaries, quick determination of a document's content, highlighting of significant words and sentences in a document and discovery of the key parts of a document.

Fig. 4. Online Brevity Document Summarizer

Shvoong Summarizer [29] is a free summarizing tool which offers a quick summary of document's content in a clean manner, with the summarized portion highlighted or summarized (Fig. 5).

Fig. 5. Shvoong Summarizer

All the automatically generated summaries are informative, just like the human summary made by the authors. After reading such an informative summary, the reader gets a glimpse into a piece of content, and is able to describe parts of the original text [6].

Furthermore, the summary created in Microsoft Word contains only general information. According to Hovy & Lin [6], this type of summary assumes the reader's weak background knowledge of the original text and therefore provides only general information from the source text.

The type of summary made by the human authors can be classified as just-the-news summary, since it provides only fresh news.

Human summarizers rely on the fact that their user possesses enough knowledge to interprete the information out of the context. Summaries generated by SweSum, SHVOONG and Online Brevity Document Summarizer can also be classified as just-the-news summaries.

While making the summary, authors used three types of reading techinques suggested by Cremmins [30].

In the first stage, authors read the original text looking for all the relevant information and focusing on the basic features of the text to be abstracted (search reading).

In the second stage, the authors created a first version of the summary by extracting important data, organizing it and reducing the relevant information into a coherent unit (creative reading).

In the third stage (called critical reading) authors have completed the summarization process, refining the completed summary htrough editing.

Although, according to a study conducted by Saggion & Lapalme [31], professional summarizers copy 11% of sentences from the source text without changes, every sentence in a human-made abstract was changed (i.e., eight sentences in the human abstract were completely redefined) as a result of the third stage.

According to Hovy & Lin [6], a coherent summary is written in grammatical sentences which are fluently and logically connected. Hence, a good summary must possess cohesion (correlation of the text units in a semantic relationship), which is a driving factor of the text comprehension.

Authors (ibid) claim that incoherent summary tends to be fragmented and consists of parts of the text which are not found in grammatically correct sentences or aren't inserted into coherent paragraphs.

The summary created with the the Online Brevity Document Summarizer is fully coherent and cohesive (Table 1).

Online Brevity Document Summary
In recent years, internet piracy or file sharing of copyrighted material has gained much popularity.
It is a topic that has incessantly sparked debate and has even received global attention.
Although on the surface the act of file sharing may seem harmless, it is far from it.
Each year, film and record production companies as well as software and video game development companies suffer from billions of dollars in lost profits.
This loss comes as a direct result of internet piracy.
Many agree that file sharing is unethical and consider it to be no different than downright theft.
On the other hand, there are those that support file sharing and claim that their actions are easily justified and ethical.

Table 1. Online Brevity Document Summary

The extract created in Microsoft Word is almost entirely coherent and the cohesion of the source text is maintained (Table 2).

The only sentence that stands out is the last one, undermining the coherence, since it does not belong to the summary.

Microsoft Word Summary

In recent years, internet piracy or "file sharing" of copyrighted material has gained much popularity.

File sharing, put simply, is the process of making available and distributing files via the internet.

The growth of file sharing on P2P networks and other file sharing methods can probably be attributed to the advances in audio compression in the mid 90's.

Today, file sharing and internet piracy has grown to new proportions.

Proponents of file sharing often suggest that information should be "free."

Opponents of file sharing heavily criticize and question the ethics of those that support the aforementioned arguments.

It is my personal opinion that file sharing is unethical.

Contrary to the arguments of its proponents, file sharing is not a victimless crime.

File sharing of copyrighted material is nothing more than anonymous theft, and like most, I consider theft unethical.

File sharers do not possess such rights.

Table 2. Microsoft Word Summary

The summaries generated by SweSum and SHVOONG are not entirely coherent, since they both contain two semantically redundant phrases, but they both maintained the cohesion (Table 3 and Table 4).

Swesum Summary

On the other hand, there are those that support file sharing and claim that their actions are easily justified and ethical.

The growth of file sharing on P2P networks and other file sharing methods can probably be attributed to the advances in audio compression in the mid 90's.

Shortly after this format became available, individuals around the world began "backing up" their music collections and sharing these small MP3 files with friends.

Today, file sharing and internet piracy has grown to new proportions.

Thus, is the sharing of copyrighted files also illegal or should it be considered theft?

Contrary to the arguments of its proponents, file sharing is not a victimless crime.

File sharing of copyrighted material is nothing more than anonymous theft, and like most, I consider theft unethical.

Clearly, internet piracy is a significant issue.

Table 3. Swesum Summary

SHVOONG Summary

In recent years, internet piracy or "file sharing" of copyrighted material has gained much popularity.

The growth of file sharing on P2P networks and other file sharing methods can probably be attributed to the advances in audio compression in the mid 90's.

Shortly after this format became available, individuals around the world began "backing up" their music collections and sharing these small MP3 files with friends.

Without a doubt, Napster is responsible for the growth in file sharing popularity as its main focus was to offer an easy way for users to share their music collection.

Today, file sharing and internet piracy has grown to new proportions.

File sharers rationalize their actions based on the argument that they "would not have purchased the content anyway."

Last, file sharers argue that the content that they are freely obtaining (stealing) is not worth the price.

File sharing of copyrighted material is nothing more than anonymous theft, and like most, I consider theft unethical.

Table 4. SHVOONG Summary

Finally, regarding the evaluation, the summary created by four different automatic summarizers (Microsoft Word summarizer, SweSum summarizer, SHVOONG summarizer and Online Brevity Document Summarizer) were evaluated in terms of how well they match the human-made abstract.

Out of ten sentences in the summary created by Microsoft Word, only five sentences matched the human summary and were classified as relevant and salient.

Tthe other five sentences were classified as entirely redundant, since none of them matched the sentences in the human summary (not even the partial match).

However, all seven sentences in the Online Brevity Document Summarizer extract were salient and relevant, while the summary contained no redundant information.

Unlike the authors who made an abstract summary, the automatic summarizers literally took all sentences from the original text and included them in the final extract.

Furthermore, while creating their abstract, authors have deleted the irrelevant and repetitive words, shortened the sentences, took only sentence parts from the source text creating a new short sentence, and synthesized two or more sentences into one.

All sentences included in the abstract were based on the original source, but authors changed the order of some sentences in the abstract to obtain a better quality and cohesion.

All automatic summarizers, on the other hand, kept the original sentence order. It is interesting to point out that in the summary created in Microsoft Word, five out of ten extracted sentences were found at the beginning of a new paragraph in the source text.

Summing up all the above remarks, authors have concluded that although all automatic summarizers rely heavily on the surface methods of text summarization, they all generate informative extracts satisfying quality expectations of the human users.

IV. Comparison of the automatic summaries: Microsoft Word, SweSum, SHVOONG and Brevity

All four automatically created summaries are generic summaries. The summary created in Microsoft Word contains general information, while the other three can be classified as just-the-news summaries. When creating summaries, all four automatic summarizers used the word frequency method (or keyword method), although Online Brevity Document Summarizer did it to a much lesser extent.

The automatic summarizer Microsoft Word used the sentence position method and the keyword method.

SweSum also used the sentence position method, but, compared to Microsoft Word, did not extract only the sentences from the beginning of each section in the source text. The reason for this may be the additional feature in SweSum - sentences with bold text are given a higher score than the ones without it.

Furthemore, Swesum assigns a higher score to sentences containing numerical data. All these parameters (keywords, sentence position, bold text and numbers) are normalized and put in a naïve combination function with no special weighting to obtain the total score of each sentence in SweSum. Additionaly, the user of SweSum can also specify his own keywords to the system and thus obtain a more user-centered summarization approach.

Furthermore, SHVOONG summarizer included only two sentences from the beginning of the each section of the original document in its summary.

The Online Brevity Document Summarizer relied more on the sentence position method than the keyword method while creating the summary. None of the systems used a cue method.

We believe that the best summary was created with the Online Brevity Document Summarizer.

The reason for such opinion is that all seven sentences in the extract can be regarded as sentences that actually belong to the summary (i.e., TP - true positives).

Every sentence included in the summary is salient and relevant, while the summary has no redundant information.

The extracted text is also fully coherent which means that it makes sense as a whole.

Furthermore, summary is cohesive which means that its mechanical connections are correct. Hence, the extract has fluently and logically related sentences.

For the above explained reasons, authors have concluded that the extract made by the Online Brevity Document Summarizer is of high-quality and comparable to the human abstract by its informativity, coherence and cohesion.

V. Evaluation of the Word, SweSum and SHVOONG automatic summaries against the Brevity summary

The automatic text summarization systems are usually evaluated with the common information retrieval metrics of precision and recall.

The human selects sentences from the source text that best convey its meaning and the sentences extracted by the automatic system get evaluated against the human selection.

Although all four machine-generated summaries represented 10% of the length of the source text, they did not contain the same number of sentences.

The summary created by the Online Brevity Document Summarizer had seven sentences.

The Microsoft Word summary had ten sentences, while the two summaries created by SweSum and SHVOONG both contained eight sentences.

Since we already manually determined that the Online Brevity Document Summarizer generates extracts of the quality comparable to the human extracts, we evaluated the three automatic summarization systems (Microsoft Word, SweSum and SHVOONG) against the Brevity.

The recall was calculated as the fraction of the sentences extracted by the Brevity that were also correctly identified by other systems.

Furthermore, the precision was calculated as the fraction of sentences extracted by the other systems that were correct.

The evaluation procedure determined the number of:

- **TP (true positives -** sentences that should belong to the summary and were included in it)
- **FN (false negatives -** sentences that should belong to the summary but were not included in it) and
- **FP (false positives -** sentences that should not belong to the summary but were included) for all three systems.

Results obtained are given in Table 5.

Precision: $$\frac{TP}{TP + FN} \quad (1)$$

Recall: $$\frac{TP}{TP + FP} \quad (2)$$

The precision of SweSum was 0.125, as well as the precision of SHVOONG.

The same result might indicate that both automatic systems use the same summarization features. The precision of Microsoft Word was 0.1.

The recall for all three systems was 0.143.

It is interesting that SweSum and SHVOONG have chosen the four identical sentences in their summaries (out of 8).

While SweSum and Word share four identical sentences, SHVOONG and Word share three equal sentences.

Sentence Number	SweSum			SHVOONG			Word		
	TP	FN	FP	TP	FN	FP	TP	FN	FP
1		+	+	+			+		
2		+	+		+	+		+	+
3		+	+		+	+		+	+
4		+	+		+	+		+	+
5		+	+		+	+		+	+
6		+	+		+	+		+	+
7	+				+	+		+	+
8			+			+			+
9									+
10									+
Total	1	6	7	1	6	7	1	6	9

Table 1. Number of true positives (TP), false positives (FP) and false negatives (FN)

VI. Conclusion

In today's information age, text summaries represent a necessity. Due to huge savings in time, they represent a useful tool for managing the vast available online literature.

The process of summarization reduces the complexity and length of the original document, providing the visibility of the subject matter and key ideas of the work.

The professional summaries help users to make a decision regarding the genre of the document and the need to read it. Summaries can either be generic or query-oriented summaries intended for a specific domain.

The purpose of the summary depends mostly on the end user, who is the ultimate consumer, and on the context in which the summary will be used.

While creating a summary, humans usually produce the abstract, changing the order of words in a sentence, deleting the irrelevant and repetitive words, shortening the sentences and synthesizing two or more sentences into one.

Automatic summarizers compared in this paper, on the other hand, produce extracts, keeping the original sentences from the source document and generating informative and generic summaries.

They all use the sentence position method and the word frequency keyword method, but some of them (like SweSum) also use the additional features, such as the bold text and numbers.

In this paper we have compared four machine-generated summaries created with the Microsoft Word, SweSum, SHVOONG and Online Brevity Document Summarizer with the human-made summary.

We found out that the best summary was created by Brevity, since all seven sentences in the extract were salient and relevant, while the extract contained no redundant information. Since the extract was fully coherent and cohesive, we concluded that its quality was comparable to that of the human extracts.

Finally, we can conclude that although the automatic creation of a human-like summary requires different methods of both deep and shallow natural language analyisis, surface summarization methods used by the above mentioned text summarization systems proved to be computationally effective, universally applicable and effective in generating informative extracts satisfying quality expectations of the end users.

References

[1] Berger, A., Mittal, V.O. (2000) Query-relevant summarization using faqs. In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL '00). Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 294-301

[2] Conroy J.M., Schlesinger J.D., O'Leary D.P. (2006) Topic-Focused Multi-document Summarization Using an Approximate Oracle Score. Proceedings of the COLING/ACL.

[3] Jiaming Z. (2008) Exploiting Textual Structures of Technical Papers for Automatic Multi Document Summarization. PhD Thesis. Department of Mechanical Engineering, National University of Singapore.

[4] Berkovsky S., Baldwin T., Zukerman I. (2008) Aspect-Based Personalized Text Summarization. Proceedings of 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. pp. 267-270.

[5] Rowley, J. (1982) Abstracting and Indexing. Clive Bingley, London.

[6] Hovy, E., Lin, C-Y. (1998) Automated Text Summarization And The Summarist System, TIPSTER III, Final Report (SUMMAC). pp. 1-17

[7] Strzalkowski, T., Wang J., Wise B. (1998) A robust practical text summarization system. In: Proceedings of the Fifteenth National Conference on AI. pp. 26-30

[8] April, Rochester, NY, USA. Marcu, D. (1997) From discourse structures to text summaries. In the Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, (ACL'97/EACL'97)

Workshop on Intelligent Scalable Text Summarization. Madrid, Spain. pp. 82-88.
[9] Nakao, J. (2000) An algorithm for one-page summarization of a long text based on thematic hierarchy detection. In: Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. pp. 302-209.
[10] Barzilay, R., Elhadad, M. (1997) Using lexical chains for text summarization. In: Proceedings of the ACL (Association for Computational Linguistics) Workshop on Intelligent Scalable Text Summarization. Madrid. pp. 10-17.
[11] Leite, D.S.; Rino, L.H.M.; Pardo, T.A.S.; Nunes, M.G.V. (2007). Extractive Automatic Summarization: Does more linguistic knowledge make a difference? In C. Biemann; I. Matveeva; R. Mihalcea and D. Radev (eds.). Proceedings of the Workshop on TextGraphs 2: Graph-Based Algorithms for Natural Language Processing, pp.17-24.
[12] Tadashi, N., Yuji, M. (2001) A New Approach to Unsupervised Text Summarization. In: Proceedings of ACM SIGIR (Special Interest Group on Information Retrieval). pp. 26–34
[13] Preradovic Mikelic, N, Lauc, T, Boras, D. (2007) CROXMLSUM – the system for XML document summarization in Croatian. International Journal of Mathematics and Computers in Simulation, 2007 (1), pp. 81–89.
[14] Lauc, T, Mikelic, N, Boras, D. (2005) Croatian Text Summarizer (CROSUM). In Proceedings of ITI 2005 (Proceedings of the 27th International Conference on Information Technology Interfaces) / Budin, L., Luzar-Stiffler, V., Bekic, Z., Hljuz Dobric, V. (eds). Zagreb: SRCE, pp. 651-657.
[15] Mikelic Preradovic, N., Ljubešić, N., Boras, D. (2010) Croatian web text summarizer (CroWebSum). Proceedings of the ITI 2010 32nd International Conference on INFORMATION TECHNOLOGY INTERFACES. University Computing Centre, University of Zagreb, pp. 109-114.
[16] Liu, S. (2007) Computing with Words Using Fuzzy Logic: Possibilities for Application in Automatic Text Summarization. In Proceedings of IFSA (2), pp. 151-160.
[17] Luhn, H.P. (1958) The automatic creation of literature abstracts. IBM Journal of Research and Development, 2, pp. 159–165.
[18] Edmundson, H. P. (1969) New Methods in Automatic Extracting, Journal of the Association for Computing Machinery, Vol. 16, No. 2. pp. 264-285
[19] Demartini, G., Missen, M. M. S., Blanco, R., Zaragoza, H. (2010) Entity summarization of news articles. In Proceedings of the SIGIR '10 (Proceedings of the 33rd International ACM SIGIR conference on research and development in Information Retrieval), ACM, pp. 795-796.
[20] Barzilay, R., Lapata, M. (2008) Modeling Local Coherence: An Entity-Based Approach. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL '05). Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 141-148.
[21] Dang, C., Luo, X. (2008) WordNet-based Document Summarization. In Proceeding of the 7th WSEAS Int. Conf. on APPLIED COMPUTER & APPLIED COMPUTATIONAL SCIENCE (ACACOS '08), Hangzhou, China, pp. 383-387.
[22] Dang, C., Luo, X, Zhang, H. (2008) WordNet-based Summarization of Unstructured Document. Journal of WSEAS Transactions on Computers, 9 (7), pp. 1467-1472.
[23] Enokizu, H., Murakam. S., Kumasaka, M., Uenosono, K., Komiya, S. (2008) Automatic Extraction of Important Sentences from Story Based on Connecting Patterns of Propositions. In Proceedings of the 7th WSEAS International Conference on ARTIFICIAL INTELLIGENCE, KNOWLEDGE ENGINEERING and DATA BASES (AIKED '08), University of Cambridge, UK, pp. 41-49.
[24] Radev, D. (2000) A Common Theory of Information Fusion from Multiple Text Sources. Step One: Cross Document Structure. In Proceedings of 1st SIGdial Workshop on Discourse and Dialogue, pp. 74–83.
[25] Erkan, G., Radev, D. R. (2004) LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research, 22, pp. 457–479.
[26] Litvak, M., Last. M. (2008) Graph-based Keyword Extraction for Single document Summarization. In Proceedings of MMIES 08' (Proceedings of the Workshop on Multi-source Multilingual Information Extraction and Summarization).
[27] Dalianis, H. (2000) SweSum - A Text Summarizer for Swedish. Technical report TRITA-NAP0015, IPLab-174, NADA, KTH. (http://swesum.nada.kth.se/index-eng.html)
[28] Online Brevity Document Summarizer. http://www.lextek.com/cgi-bin/Brev.cgi. (URL viewed 2.03.2012.)
[29] Shvoong Summarizer. http://www.shvoong.com/summarizer/. (URL viewed 1.03.2012.)
[30] Cremmins, Edward.T . (1982) The art of abstracting. Philadelphia: ISI Press.
[31] Saggion, H., Lapalme, G. (1998) Where does Information come from? Corpus Analysis for Automatic Abstracting. Rencontre Internationale sur l'extraction, le filtrage et le résumé automatique (RIFRA'98). pp. 72–83.

Marta Vlainić, is a PhD candidate at the Department of Information and Communication Sciences, University of Zagreb, Zagreb, Croatia (email: marta.vlainic@gmail.com)

Nives Mikelić Preradović is associate professor at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia (nmikelic@ffzg.hr).
She obtained her MA in Croatian language and literature and Information sciences at the University of Zagreb.
She also obtained MPhil in Natural Language and Speech Processing at Cambridge University, United Kingdom. She obtained her PhD in 2008 at the Zagreb University on the development of the Croatian Valency Lexicon – CROVALLEX (theta.ffzg.hr/crovallex/) and accentual-derivational models for Croatian nouns and adjectives as well as accentual-conjugational model for verbs.
Her research interests include developing multilingual valency lexicons, morphosyntactic annotation, sentiment analysis (opinion mining), computer-assisted language learning and text summarization.
She participated in several international and national projects: ACCURAT (Analysis and evaluation of Comparable Corpora for Under Resourced Areas of machine Translation), CESAR (CEntral and South-east europeAn Resources), Abu-MaTran (Automatic Building of Machine Translation), Typology of Knowledge and Information Processing Methods and Design and Management of Public Knowledge in the Information Space.
She published a book, 10 book chapters and about 30 scientific papers in international journals and conference proceedings.
Also, in ac. y. 2005-2006 she introduced a new teaching method called service learning (community-based learning), which became the measure of the National Program for Youth (2009-2013), approved by the Croatian Government. The Program emphasizes the importance of the introduction of this method in the university and high school curriculum to increase the community involvement of young people and education for the civil society, linking educational institutions and the community. Since 2005 she mentored and administrated over 60 service learning projects at the Faculty of Humanities and Social Sciences, University of Zagreb.

Damir Boras is full professor with tenure at the Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb.
He is also the dean of the Faculty of Humanities and Social Sciences, University of Zagreb.
He is the head and the founder of the Chair for lexicography and encyclopaedic science at the Department of Information and Communication Sciences.
His main recent research interests are dictionary heritage and dictionary knowledge presentation, linguistic databases, computer assisted language learning, morphosyntactic annotation and terminology extraction.
He has been principal investigator or research team member in many projects related to the application and introduction of the information technology, high and elementary education, publishing and jurisdiction (Sources for Croatian Heritage and Croatian European Identity, Croatian Dictionary Heritage and Croatian European Identity, Croatian dictionary heritage and dictionary knowledge presentation etc.).
He is a member of many committees and boards (Committee for Science, Education and Culture of the Croatian Parliament, Board of the National Center for External Evaluation of Education, National Scientific-field board for Social sciences at the National board for science, etc.).