Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 1387–1396

DOI: 10.15439/2015F188
ACSIS, Vol. 5

The automatic summarization of text documents in the Cognitive Integrated Management Information System

Marcin Hernes
Wrocław University of Economics
ul. Komandorska 118/120, 53-345
Wrocław, Poland
Email: marcin.hernes@ue.wroc.pl

Marcin Maleszka
Wroclaw University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370
Wrocław, Poland
Email: marcin.maleszka@pwr.edu.pl

Ngoc Thanh Nguyen
Wroclaw University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370
Wrocław, Poland
Email:
ngoc-thanh.nguyen@pwr.edu.pl

Andrzej Bytniewski
Wrocław University of Economics
ul. Komandorska 118/120, 53-345
Wrocław, Poland
Email:
andrzej.bytniewski@ue.wroc.pl

***Abstract*—This paper presents issues related to a process of the automatic summarization of the text documents connected with economic knowledge performed by the cognitive agents in an integrated management information system. In contemporary companies, the unstructured knowledge is essential, mainly due to the possibility of obtaining better flexibility and competitiveness of the organization. Therefore more often the decision are taken in the enterprises on the basis of the summaries. The first part of the paper shortly presents the state-of-the-art in the considered field; next, the summarization process in the Cognitive Integrated Management Information System is characterized; the case study related with the summaries generating agent is presented in the last part of this paper.**

I. Introduction

IN contemporary companies the unstructured knowledge is essential, mainly due to the possibility of obtaining better flexibility and competitiveness of the organization. The unstructured knowledge supports structuralized knowledge to a high degree. It is mainly stored in natural language, so it is processed with symbols (not numbers). One example of unstructured knowledge is experts' opinion about a predicted currency trading. Some experts may argue that the exchange rate of the currency will rise, others that it will decrease, and still others that it will remain unchanged. In addition, expert opinions include the reasons for these predictions. The number of such opinions in the Internet is usually very large (hundreds, thousands of the opinions). An investor who makes a decision, for example, on the financial markets, needs to analyze and summarize these opinions to formulate the correct decision. However, the manual realization of these processes is extremely difficult, and often impossible, due to time constraints [14]. Thus, often the processes of the analysis and summarization of the text documents are made automatically by computer systems, including the integrated management information systems. They may be constructed, for example, on the basis of the number of the cognitive agents [16]. Generally speaking, the cognitive agent is a smart program that not only concludes on the basis of the data received, takes specific actions to achieve the desired objective (this can be, for example, decision support), but also learns at the same time gaining experience. An example of a cognitive agent's architecture is The Learning Intelligent Distribution Agent (LIDA) [35]. This is a hybrid architecture that allows for symbolic and emergent knowledge processing and uses the semantic net with node and links activation level (the "slipnet") [19] to represent a knowledge.

Broadly understood, the analysis of the text documents is mainly based on document retrieval, information extraction, text mining and natural language processing. Summarization, instead, can include the contents of a document or set of documents. The basic idea of summarization is to get a summary that contains the most important information from the source document. One of the parameters of this process is the text volume. A good document summary frees the system user (investor, manager) from the need to read and analyze all of the text documents, and give the opportunity to focus his attention on aspects of the rapid and effective decision.

So far, the methods of summarization of the electronic text documents, containing economic knowledge and represented by the "slipnet" (semantic net with node and links activation level), has not been developed. It should be noted, however, that this type of representation allows the processing of both

978-83-60810-66-8/$25.00

knowledge represented in a symbolic way and knowledge represented in a numerical way.

The aim of this paper is to develop a method for automatic summarization of the electronic text documents, containing economic knowledge. This method will be implemented in the architecture of the cognitive agents running in the Cognitive Integrated Management Information System (CIMIS). The agent-based approach is used mainly because it enables taking automatic decisions and performing actions on the basis of summarization results.

This paper is organized as follows: the first part shortly presents the state-of-the-art in the considered field; next, the summarization process in CIMIS is characterized; the case study related with summaries generating agent is presented in the last part of paper.

II. Related works

The problems of automatic summarization have been widely considered in many papers and practical solutions. The simplest method for creating summaries is based on the assumption that the weight of the sentence depends on the weight of its words, calculated on the basis of their frequency in the text. In addition to the counted weight of words, other factors are also taken into account, such as the position of sentences in the text and the occurrence of words in the title or header [30]. An important work on the problem, was the project Parseval/GEIG, where the phrase structure grammars have been compared [4]. In this project using an anaphoric relations to create appropriate groups (consisting of paragraphs) has been proposed. The authors propose is that either all members of the group belong to the summary, or no element of the group appears in the summary. The solution proposed by [24] was based on several standard heuristics (e.g. sentence length, sentence position, keywords). However, the poor quality of the summaries generated using these tools has led to the search for alternatives. Another solution was tf-ifd system, used in ANES (Automatic News Extraction System), which determined the weight of words based on the number of their instances in the analyzed document [5]. R.Barzilay and M.Elhadad [3], instead, used an algorithm based on lexical chain for automatic creation of summaries. In the Tipster project a series of activities focusing on tasks such as summarization, translation and searching, were realized [18]. The paper [20] presents an overview of the different methods of summarization. Description of the elements of natural language processing and information retrieval methods have been widely presented in the works [2] and [7]. D.Weiss [43] presents algorithms of lemmatization and stemming, and a hybrid solution combining both approaches. P.Sołdacki [38] devoted his doctoral dissertation to an automatic processing of documents in natural language, in particular the use of text analysis methods for the shallow processing of documents in Polish. The problems with the automatic creation of summaries of texts document in Polish were well represented in the work of A.Dudczak [10]. Also E. Branny, M. Gajecki [46] present an algorithm for text summarization in Polish. Most of the mentioned works have been related to the creation of a single document summaries. Issues related to the creation the summaries of a set of documents can be found, for example, in the works [9], [30]. The papers [14], [27, 28] present two main streams of research related to the analysis of natural language texts:

- the formal description of the language and the real world's right,
- the statistical methods that bypass the problem of understanding of the text for the analysis of the prevalence of selected dependences.

J. Gramacki and A. Gramacki [15] used selected algebraic methods of analysis of textual data for automatic creation of summaries. Their paper presents the data structures and data modeling, which showed the essence of the reduced vectors' space of the mapped texts.

The paper [40] however, concerns the use of semantic roles in the process of summarization. The work [25] proposes a system of summarization of text documents by using a semantic net (semantic graph) for knowledge representation and mechanisms of the support vectors in the learning process. The methods rely on creating a new node for each sentence. (Fig.1).

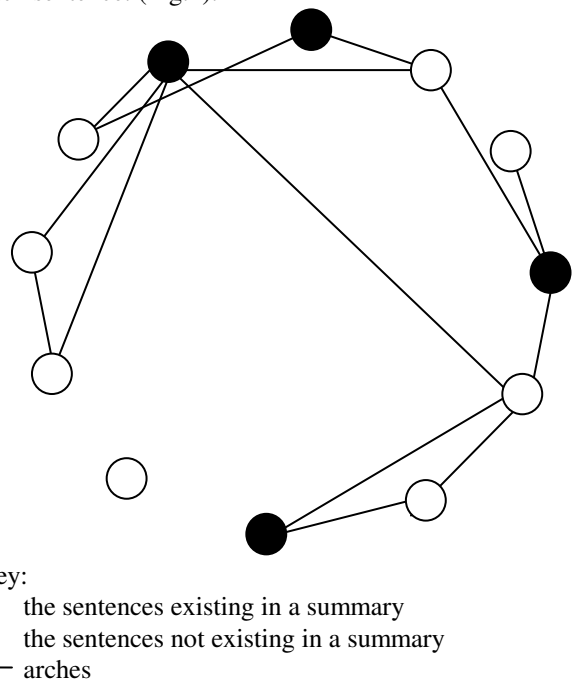

Fig 1. The example of semantic graph.

If two sentences have common word then these sentences' nodes are connected by an arc (if the sentences have more than one common word - for each pair a separate arc is created). Sub-graphs whose nodes are connected by arches with other sub-graphs contain statements that define a good topic, while the nodes of sentences that have the highest cardinality (most arches) are the most significant sentences

in the considered text, and should be included in the summary. However the applied semantic net does not include the activation levels of nodes and arcs.

Taking into consideration issues related to a text documents analysis process, nowadays the hybrid methods for processing unstructured knowledge are used; the methods involve structuralization of knowledge, followed by symbolic processing (e.g. with the use of expert systems or genetic algorithms) or converting knowledge into numerical representation followed by numerical processing (e.g. with the use of neural networks or fuzzy logic systems). In both cases, for knowledge processing, the following methods are used [32]:

- Information Retrieval,
- Information Extraction,
- Text Mining,
- Natural Language Processing.

There are two categories of analysis of text documents in the literature of the subject (for example [38]):

1. Deep Text Processing, that is a linguistic analysis of all possible interpretations and grammatical relationships occurring in natural text. The full analysis can be very complex. Moreover in many cases, the information obtained in this way may not be necessary. For this reason, more and more often there is a tendency to carry out only a partial analysis of the text, which may be much less time-consuming and is a compromise between precision and performance.
2. Shallow Text Processing, is defined as the analysis of the text to which the effect is incomplete in relation to the deep analysis of the text. The usual limitation lies in the identification of non-recursive structures or of limited recursion level, which can be diagnosed with a large degree of certainty. The structures requiring a complex analysis of the many possible solutions are overlooked or analyzed in part. The analysis is addressed mainly at recognizing proper names, noun phrases, verb groups without resolving their internal structure and function in a sentence. In addition, recognized are some main parts of sentences, for example the judgment or the judgment group.

There are several different methods within these categories.

For information retrieval the Boolean Logic Model (BML) or ranked-output systems are used [39]. A BML query consists of words or phrases concatenated with logical operators, such as AND, OR, and NOT. As a result, the set of documents is divided into two sub-sets: the first sub-set consists of documents matched with the query and the second sub-set consists of documents mismatched with the query. A ranking system, using vector algebra, assesses the probability of the content of documents matching the content of the query, and on this basis the ranking of the found documents is created. In this approach, for example, the Vector Space Model, Probabilistic Model or Interface Network Model are used [34].

One of the methods used in the shallow analysis of text documents is machine learning [33]. Under this method there are used, among other things, the naive Bayesian classifiers [12] and support vector machines [21]. In these approaches an analysis is made of the prevalence of individual words (terms) in the documents concerned. For example, in work [42] with the use of support vector machines it was determined, that the product attribute is considered part of the text, whereas in work [43] polarity of the opinion was expressed during the passage of the text.

Another method of both deep and shallow analysis of text documents is the use of rules, on the basis of which identification (annotation) of pieces of text for a specific topic is performed. Such rules are based on templates, taking into account the relationship between words and semantic classes of words [37]. The basis for the generation of rules can be automatic or manual analysis of the annotated corpus [31]. An analysis of documents by using rules rely on identifying the importance of text fragments in accordance with the principles enshrined in the rule. In certain cases it can be thought of as assigning the document to the category. The analysis of text documents with the use of the rules has been used, among others, to the extraction of spatial relationship [45], identification of the requirements for the IT projects, expressed on Internet forums [41] or extraction of information from real estate ads [33]. In turn, the article [1 shows the characteristics of the project Semantic Monitoring of Cyberspace, that uses rules analysis in order to search cyberspace offers regarding illicit trafficking in drugs.

The text files are often represented in databases with key words contained in the document (symbolical representation of knowledge). However, such representation makes it difficult to compare documents, especially when it comes to measuring distance between documents - distance meaning similarity between the documents. Increasingly semantics nets are used to represent text documents, for example the topic maps. In work [11] it was found that the topic map allows to record information ontology and data taxonomy, structured semantically and at the same time it allows for knowledge mapping (both structured and unstructured) on a wide variety of hierarchical dependencies that exist between economic concepts and semantics (the concept of this type include, among others, to text documents in the field of management and economics).

However to match fully the needs of decision makers, a decision structure must consist of the level of certainty because the economic decision most are taken in terms of risk or uncertainty. Nowadays such structures [36] and consensus algorithms as regards these decisions, are elaborated [23, 47, 50].

Thus it is possible to determine a certainty level of semantic relations between nodes (topics). In case of the economic knowledge it is very important issue, because the decisions making on the basis of this knowledge is performed usually takes place under risk and uncertainty conditions.

III. AUTOMATIC SUMMARIZATION IN THE CIMIS

The CIMIS is dedicated mainly for the middle and large manufacturing enterprises operating on the Polish market (because the user language, at the moment, is Polish language). This does not preclude the implementation of other language text documents processing in the CIMIS. The process of analysis of text documents highly depends on the language. In addition, it is more difficult in case Polish language than, for example, in case English or German due to the greater complexity of Polish grammar [32].

The Learning Intelligent Distribution Agent (LIDA) architecture is used in CIMIS construction [8], [13]. In the construction of a LIDA agent mixed-used symbolically-connexionistic organization of memory is used in an attempt to ground the meaning of all symbols. It is necessary to properly process the unstructured knowledge, recorded mostly using natural language, such as customer's opinion about products. Grounding is meant as these cognitive processes, which are responsible for establishing and maintaining a connection between the language and the corresponding objects in the world [22]. The LIDA consists of the following modules [13]:

- Sensory Memory
- Perceptual-Associative Memory,
- Workspace,
- Transient Episodic Memory,
- Declarative Memory,
- Attentional Codelets,
- Global Workspace,
- Procedural Memory,
- Action Selection,
- Sensory-Motor Memory.

In the LIDA architecture it was adopted that the majority of basic operations are performed by the so-called codelets, namely specialized, mobile programs processing information in the model of global workspace.

In 2011, the CCRG group released the Framework LIDA, which allowed the use of this architecture by a wider circle of users. Framework LIDA is a software underlying the implementation of the cognitive agents. The Framework contains the object class (implemented in JAVA), performing operation in the field of agent architecture (definition and methods of handling all types of memory, communication protocols, methods for making reservation by the agent operations on real-world objects, such as searching for and identifying the objects, specify characteristics of objects, specifying associations between objects). The programmer's task is to fill in tools provided by the framework LIDA (write a program code) about aspects related to the specific domain of the problem - for example related to economics, management.

The CIMIS consist of following sub-systems: fixed assets, logistics, manufacturing management, human resources management, financial and accounting, controlling, CRM, business intelligence. These sub-systems are described in detail in [16]. In this paper we focus on summarization module, which is a part of CRM sub-system (Fig 2.). The module consist of following groups of agents:

- document retrieval,
- information extraction,
- text analysis,
- summaries generating.

Document retrieval agents search and retrieve, from the internet sources the documents according to users' needs (consistent with the query retrieval) and save full content of these documents in a database. Retrieving the documents is carried out by codelets operating in an agent environment. In order to achieve a higher level of retrieval effectiveness each agent running on the basis of different document searching engines, for example:

- Java Searching Engine,
- Microsoft BING API,
- Custom Search API,

The documents from the top of its list are stored in database. The number of stored documents is given in the form of a parameter in the agent configuration file (e.g. 10, 20, 50 documents). The format of retrieved documents is not limited (e.g. Pdf, PS, doc, html, xml), while in the database, each document is stored as text, except that html document is saved along with markers.

The role of information extraction agents is to identify essential information in text documents. For example, if opinions of mobile telephone users are to be analyzed, only those pieces of texts which contain opinions (advertisements are to be omitted) need to be extracted from text documents (saved in a database by documents searching agents). Each agent uses a different method of extracting information, for example:

- determining tags of beginnings and ends of essential fragments of texts in a document - the method is mainly used in case of files saved in html/xml format (for example, the series of characters ""/> " is treated as the beginning of an opinion whereas the series of characters " " is treated as an end of an opinion (the characters are determined on the basis of a learning set of text documents);
- identifying essential fragments of texts on the basis of sets of key words or rules - the method is used in case of any document formats (for example an opinion about a phone may be identified on the basis of such

key words as "make", "model", "recommended/not recommended"),

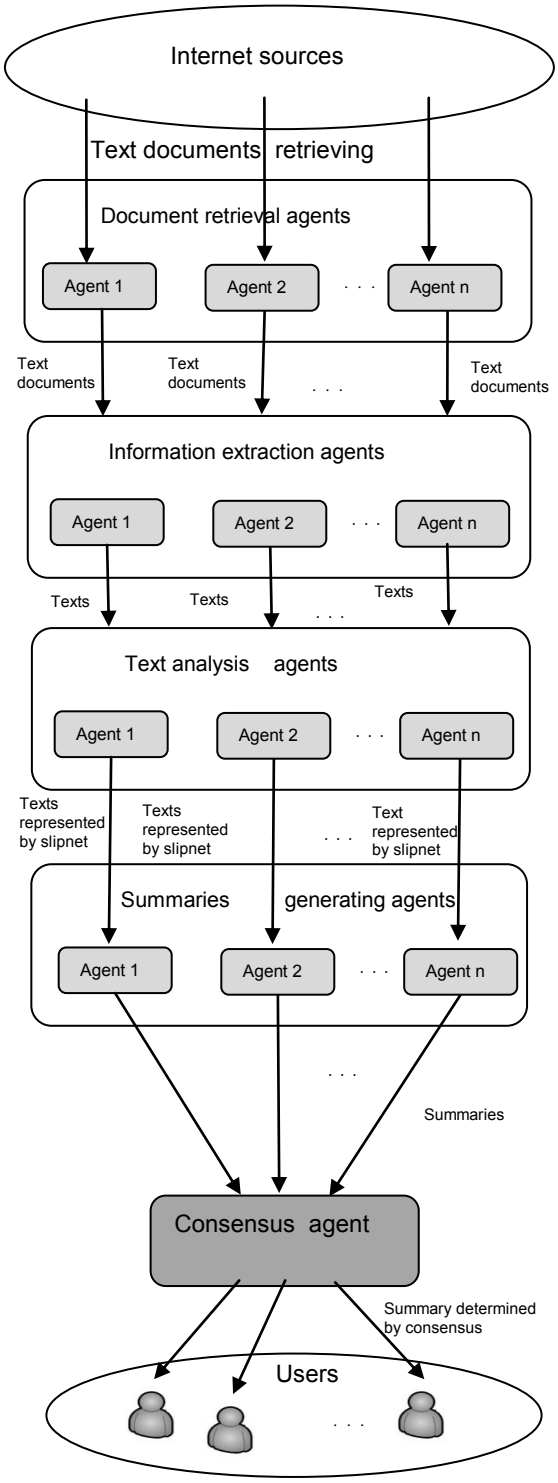

Fig 2. The functional architecture of automatic summarization module.

- identification of essential fragments of texts using documents representations in the form of a semantic network - semantic networks are created on the basis of a learning set, the networks serve as patterns representing particular classes of information (for example a semantic network representing a client's opinion on a given mobile phone), whereas in the process of IE a text document is saved also in the form of a semantic network, and then a cognitive agent searches for a given pattern in a semantic network representing a text document (the degree of similarity is defines as a parameter). This method allows for including a context of extracted information.

The results of the agents' running are stored in the text database.

The environment in which text analysis agents operate constitutes a set of text documents containing information which is the result of operation of agents extracting information (for example opinions of customers about mobile phones) located in a database of a system. Text analysis is performed in the following way:

1. A semantic network containing terms and connections between them is created in the perceptive memory on the basis of a learning set (for example a set of opinions about mobile phones). The perceptive memory stores also synonyms and different variations of words (thesaurus). In the perceptive memory of LIDA agents terms are represented by means of nodes, whereas connections are represented by means of links.

2. Individual text documents are added one by one into the sensory memory.

3. Opinions are analyzed by codelets, i.e. programs which search through texts according to certain criteria specified by means of configuration parameters. Values of the parameters may be indicated by users (parameters are saved in xml file structure and used in the program code of codelets). Codelets have been divided into two groups:

1) Codelets performing shallow analysis of texts. In the frame of this group the following codelets are running:
- tokenization,
- morphological analysis,
- removing the ambiguity,
- recognition of proper names, replacing pronouns,
- the distribution of complex sentences to the simple sentences,

2) Codelets performing in-depth analysis of texts. These codelets, by using thesaurus and results of shallow analysis codelets' operation, search for all possible interpretations and grammatical relations present in an analyzed document, represented in the form of a semantic network ("slipnet") in the perceptive memory of an agent. Results of documents' analysis, represented also in the form of a semantic network, are saved in a database.

4. The next step consists of passing the situation model to the global working memory, and from the procedural memory the following patterns of action are automatically selected: "saving results of opinions analysis into a database" (noSQL database – analysis of results – semantic network – are saved in XML format) and "entering another document into the sensory memory".

The environment in which summary generating agents operate, on the other hand, consists of a set of text documents represented in the form of a semantic network constituting the result of operations of agents analyzing documents. Summary generating agents function in a way similar to documents analysis agents, however, codelets perform tasks connected with summarizing and they are divided into the following groups:

1. Codelets of text level units – responsible for analyzing relations between fragments of a semantic network, e.g.: probability, proximity in a text, common references, language ties (taken from a thesaurus for example), syntactic and semantic relations. Results of analyses are saved in the form of a semantic network with levels of nodes and links activation reflecting the degree of similarity and relations.

2. Discourse level codelets – responsible for performing analysis in the course of which the format of a whole document is taken into account, its rhetoric structure and issues touched upon in the document. The codelets use the method of creating semantic graphs and their subgraphs (graphs and subgraphs are also created in the form of a semantic network "slipnet") using mechanisms of vectors facilitating the learning process. Levels of activating links in a graph depend on the level of the meaning of a given sentence (or phrase) in an analyzed text.

Next, the situation model is transferred to the global workspace, and from the procedural memory, the following patterns of action are automatically selected: "saving results of opinions analysis in a database", "generating summaries in a natural language" (a summary in the form of a text in a natural language, which can be presented to a user, is created on the basis of a summary in the form of a semantic network) and "entering another document into the sensory memory".

The integration of summaries is performing by consensus agent. This agent determines a summary presented to user (users). The use of consensus methods [17], [26], [29], instead, allows for the integration of summaries. The general meaning of the term consensus refers to an agreement. A consensus of a certain set (profile) of text documents may constitute a new document (hypothetical one) created on the basis of documents contained in the profile [48, 49].

Deriving consensus consists of three basic stages. In the first stage, one needs to determine a method of text documents representation. In this paper it has been assumed that the documents are represented in the semantic net. The next step requires defining the function of calculating distance between individual variants. The third stage involves developing consensus deriving algorithms, i.e. determining a representation of a set of documents (profile) where the distance between the representation (consensus) and individual documents of the profile is minimal (according to various criteria). If, for example, the different summaries of the documents describing the given phenomenon have been generated by cognitive agents, then using the consensus methods, on the basis of this summaries, the one variant of summary can be generated and presented to the user. This variant does not need to be one of the summaries generated by the cognitive agents. It can be a new variant determined on the basis of these summaries. Thus, all the summaries on a given phenomenon can be taken into account. Such a solution allows, among other, for shortening the time of determining the target summary (the user does not need to analyze the individual summaries and reflect on their choice - multi-agent system performs these steps automatically for user) and for decreasing the risk related to the choice of the worst summary (because all the summaries are taken into account in the consensus). As a result, the business processes within an organization can be implemented more quickly and efficiently.

IV. Functionality of Summaries Generating Agent – The Case Study

In the frame of this case study the Polish language text documents will be taken into consideration.

The functioning of the module generating summaries of text documents will be presented on the example of experts' opinions present in the Internet. A document-searching agent has received a task to find in the Internet documents whose content matches the following question: "In what stocks to invest during the uncertainty on the market?" For further investigation, one of the opinions obtained by an information-extracting agent was selected (the method of information extraction by determining beginning and end characters or marks of essential pieces of text in a document was used – see section III). The text of the opinion looks as follows: (sentences/phrases) of the opinion have been numbered by authors in order to simplify further analysis):

"(1) The market is in the period of uncertainty (2) therefore investments are to be made into companies representing the defensive branches. (3) Among them, we can first of all mention telecommunication companies stocks, (4) as well as those which are related to public utility. (5) We are talking here about for example energy, (6) or all other companies which make investments using public money. (7) They are the least affected by changes in the economic situation (8). Many companies conduct different types of promotion actions aimed at making people invest in them. (9) However, one should not fall for that. (10) One should not invest in retail sector companies as their prices change greatly during a fluctuation on the financial market".

On the basis of a learning set, a semantic network containing terms and connections between them related to

the investment topic, as well as a thesaurus for the terms were saved in the perceptive memory of a text analysis agent[1]. Next, a sample opinion was entered into the sensory memory of the agent, and a shallow analysis of the text was performed. As a result of the analysis, for example words "invest", "investments" were changed into "investment", and the phrase "changes in the economic situation" were changed into "fluctuation" (tokenization). Complex sentences were also broken into simple sentences (for example the sentence: "The market is in the period of uncertainty therefore investments are to be made into companies representing the defensive branches." was broken into two sentences: "The market is in the period of uncertainty" and "therefore investments are to be made into companies representing the defensive branches). In the next step, codelets performing in-depth analysis of the text (see item 3) saved an opinion in the semantic form. Figure 4 on the other hand shows representation of the following sentence: "especially during of uncertainty on the market".

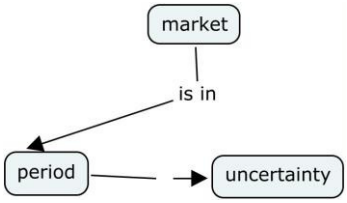

Fig. 3. A section of the network which relates to the following sentence: "especially during of uncertainty on the market"

Figure 4 presents a section of the network which relates to the following sentence: "Investments are to be made into companies representing defensive branches".

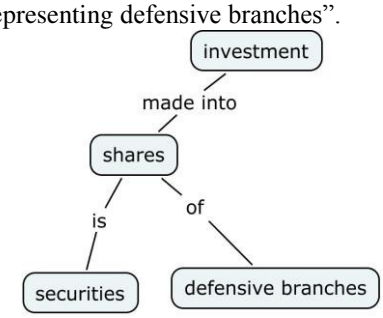

Fig 4. A section of the network which relates to the following sentence: "Investments are to be made into companies representing defensive branches"

It can be observed that the network has been enriched with the meaning of words "shares" and "companies". It has been stated that they refer to "securities".

The sentence: "Many companies conduct different types of promotion campaigns aimed at making people invest in them" (it is a dependent sentence so it has not been divided into two sentences) is represented in a way shown in figure 5. One may notice that the network has been enriched with the meaning of the phrase "promotion actions" – it has been specified that these are "actions"/"campaigns". The remaining sentences or sentences of the opinions are represented in a similar way.

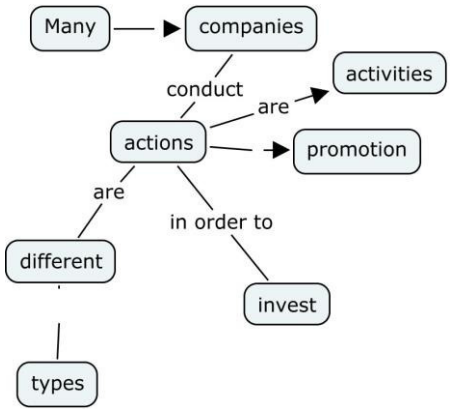

Fig. 5. A section of the network which relates to the following sentence: Many companies conduct different types of promotion actions aimed at making people invest in them".

Next, the opinion, in the form of a semantic network, is sent to the sensory memory of a summary-generating agent. Discourse level codelets generate a semantic graph (figure 6) whose nodes refer to individual sentences of an opinion (the number of sentences is specified after breaking dependent sentences into simple sentences – in case of the particular opinion, there are 10 sentences). This graph corresponds to the opinion in Polish (see an Appendix). The nodes of the graph have been joined using links whose activation levels are directly proportional to the number of words common for the investigated sentences (activation level is defined in the following way: the longest sentence in a particular opinion consists of 10 words (excluding linkers) and the activation level is defined within the range [0..1], so in case of each common word the level of activation increases by 0,1.) [2].

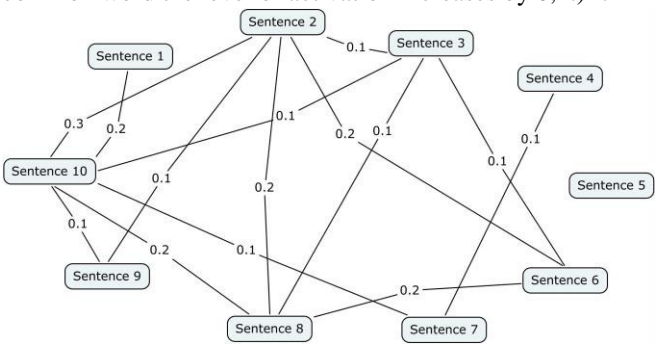

Fig 6. The semantic graph representing the considered opinion.

The most meaningful sentences are marked by nodes which have the highest total number of activation links which come out of them. The sentences should be included in a summary. Taking into account the presented semantic graph, these are the following sentences: "Sentence 2" (total

[1] The way for text analysis by cognitive agent is broader described in the paper [6].

[2] The longest sentence in the present opinion is 10 words (not including conjunctions) and the level of activation is defined in the interval [0..1], therefore, corresponds to one common expression on the activation level rise of 0.1.

of links activation levels: 0.9), "Sentence 6" (total of links activation levels: 0,5), and one of the sentence: "Sentence 10" (the sentences have the total of links activation levels: 1.0).

Apart from discourse level codelets (analyzing semantic graph) there are also text units level codelets. They enable finding sentences of different semantics despite comprising similar vocabulary. "Sentence 2" and "Sentence 8" serve as examples. The semantic graph suggests that it is enough to use just one of the sentences, however, after analysis performed by text units level codelets, it appears that "Sentence 2" ("Investments are to be made into companies representing the defensive branches") refers to investing into shares, whereas "Sentence 8" ("Many companies conduct different types of promotion actions aimed at making people invest in them") refers to taking action by companies (in Polish the words "shares" and "actions" are the same, i.e. "akcje" – see the Appendix) . Discourse level codelets mark sentences which shall be used in a summary by setting the level of activation of nodes for these sentences at the value 1 (value 0 for nodes representing sentences which shall not be placed in a summary).

Consequently, as a result of summary generating agent's operation the following text of a summary of considered opinion has been defined:

"During fluctuation on the financial market, investments are to be made into companies, representing the defensive branches, or all other companies which make investments using public money. One should not invest in retail sector companies".

It needs to be stressed that using representations of knowledge of a cognitive agent in the form of a semantic network "slipnet" enables taking into account a greater number of criteria which determine which sentences (or phrases) are to be included in a summary. For example, one can specify if more important are sentences whose nodes have a smaller number of links but whose levels of links activation are high (the sentence is "strongly" connected with other sentences, however the sentences refer to a small fragment of a summarized text), or whether more important are sentences whose nodes have a very high number of links of low activation level (the sentence is poorly connected with other sentences, however the sentences refer to a large fragment of a summarized text). Additionally, a cognitive agent, on the basis of an obtained summary, is capable of taking automatic decisions (on behalf of a user – investor) concerning abandoning investments in equities or starting investments in, for example gold. An agent will automatically sell equities (if an investor has been investing on a given market) and buy a gold.

On the basis of results of the preliminary research experiment performed by using 30 text documents, it has been state, that in about 70% cases the automatic generated summaries were very similar to summaries generating by users.

However, the main limitations of presented proposal are as follows:

- the results of text document analysis are not always properly, therefore a summaries generated on the basis of these results are also not always correct,
- in many cases, complex sentences weren't broken into simple sentences,
- there are also summaries which consist of not correct generated text (linguistic error).

V. Conclusion

The paper has demonstrated the development of a model of generating summaries of text documents containing mainly economic knowledge, represented by means of the semantic network "slipnet" (containing, apart from arch nodes, also levels of their activation). This sort of representation enables processing symbolically represented knowledge as well as numerically represented knowledge. It is possible then to define the level of probability or strength of semantic connections between terms, sentences, or phrases. Thanks to that, the cognitive agent takes into consideration user's criteria which generated summaries should satisfy. Using cognitive agents in the devised model also enables taking automatic decisions and performing actions (on the basis of knowledge acquired in the process of summarizing documents).

Proper summarization of text documents is of great significance, particularly in integrated management information system design to support decision-making processes. Their functionality has a direct effect on user decisions and – ultimately – affects the organization as a whole.

Further research works will concern, inter alia, the developing, by using a fuzzy logic, a method for determining the activation level of nodes depending on the context of the sentence (or expressions).

Appendix

The considered opinion in Polish is as follows:

„(1) Rynek znajduje się w okresie niepewności, (2) więc należy inwestować w spółki reprezentujące tzw. branże defensywne. (3) Wśród nich można wymienić przede wszystkim akcje spółek telekomunikacyjnych, (4) a także tych, które mają związek z użytecznością publiczną. (5) Chodzi tu np. o energetykę. (6) lub wszystkie inne spółki, które wykonują inwestycje z pieniędzy publicznych (7) W małym stopniu są one narażone na zmiany koniunktury (8) Wiele spółek przeprowadza różnego rodzaju akcje promocyjne aby w nie inwestować. (9) Nie należy im jednak ulegać (10) Nie należy inwestować w spółki działające w handlu detalicznym gdyż ich cena ulega dużym zmianom w okresie fluktuacji na rynku papierów wartościowych".

REFERENCES

[1] W.Abramowicz, E. Bukowska and A. Filipowska, "Zapewnienie bezpieczeństwa przez semantyczne monitorowanie cyberprzestrzeni", *e-mentor*, 3 (50) , 2013.

[2] R. A. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley Longman Publishing Co., Inc., MA, USA, Boston 1999.

[3] R. Barzilay and M. Elhadad , "Using lexical chains for text summarization", *Inteligent Scalable Text Summarization Workshop* (ISTS'97), 1997, pp. 10-17.

[4] E. Black, S. Abney, D. Flickinger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Reukos, B. Santoni and T. Strzalkowski, "A procedure for quantitatively comparing the syntactic coverage of English grammars", *DARPA Speech and Natural Language Workshop*, 1991.

[5] R. Brandow, K. Mitze and L. F. Rau, "Automatic condensation of electronic publications by sentence selection", *Inf. Process. Manage.*, 31(5), 1995, pp. 675-685.

[6] A. Bytniewski, A. Chojnacka-Komorowska, M. Hernes and K. Matouk, "The Implementation of the Perceptual Memory of Cognitive Agents in Integrated Management Information System", in: D. Barbucha, N. T. Nguyen, J. Batubara, *New Trends in Intelligent Information and Database Systems*, Studies in Computational Intelligence Volume 598, Springer International Publishing Switzerland, 2015, pp 281-290. doi: 10.1007/978-3-319-16211-9_29

[7] M. Ciura, D. Grund, S. Kulików and N. Suszczanska, "A System to Adapt Techniques of Text Summarizing to Polish*", International Conference on Computational Intelligence*, 2004 p 117-120.

[8] *Cognitive Computing Research Group*, http://ccrg.cs.memphis.edu/, [29.03.2015]

[9] D. Das and A.F.T. Martins, "A Survey on Automatic Text Summarization", *Literature Survey for the Language and Statistics II course at CMU*, 2007.

[10] A. Dudczak, *Zastosowanie wybranych metod eksploracji danych do tworzenia streszczeń tekstów prasowych dla języka polskiego*, Praca Magisterska Politechniki Poznańskiej, Poznań 2006-2007.

[11] H. Dudycz, *Mapa pojęć jako wizualna reprezentacja wiedzy ekonomicznej*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2013.

[12] E. Frank and R. Bouckaert, "Naive bayes for text classification with unbalanced classes", *Knowledge Discovery in Databases: PKDD* 2006.

[13] S. Franklin, F. G. Patterson, *The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent*, in *Proc. of the Int. Conf. on Integrated Design and Process Technology.* San Diego, CA: Society for Design and Process Science, 2006.

[14] A. Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing,* Springer, Berlin 2012.

[15] J. Gramacki and A. Gramacki, *Automatycznie tworzenie podsumowań tekstów metodami algebraicznymi*, Wyd. PAK nr 07, s. 751-755, Gliwice 2011.

[16] M. Hernes, "A Cognitive Integrated Management Support System for enterprises", w: D. Hwang, J. Jung, N.T. Nguyen N (eds.), *Computational Collective Intelligence Technologies and Applications*, Lecture Notes in Artificial Intelligence, vol. 8733, Springer-Verlag, 2014, pp. 252-261

[17] M. Hernes and N. T. Nguyen, "Deriving Consensus for Hierarchical Incomplete Ordered Partitions and Coverings", *Journal of Universal Computer Science* 13(2)/2007, pp. 317-328.

[18] L. Hirshman, *Language understanding evaluation: lessons learned from MUC and ATIS*, LREC Granada,1998.

[19] D. R. Hofstadter and M. Mitchell, "The copycat project: A model of mental fluidity and analogy-making", in D. Hofstadter, the Fluid Analogies Research group (eds), *Fluid Concepts and Creative Analogies*, Basic Books. Chapter 5, 1995.

[20] P. Jackson and I. Moulinier, *Natural Language Processing for Online Applications - Text Retrieval, Extraction and Categorization*, John Benjamins Publishing Company, Amsterdam/ Philadelphia 2002.

[21] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features", *Machine learning: ECML-98*, 1998.

[22] T.H. Duong, N.T. Nguyen, G.S. Jo, "A Method for Integration of WordNet-based Ontologies Using Distance Measures", in: Proceedings of KES 2008. Lecture Notes in Artificial Intelligence 5177, 2008, pp. 210-219. doi: 10.1007/978-3-540-85563-7_31.

[23] J. Korczak, M. Hernes and M. Bac, "Risk avoiding strategy in multi-agent trading system", [in:] Proceedings of Federated Conference Computer Science and Information Systems (FedCSIS), Kraków, 2013, , pp. 1119 - 1126.

[24] J. Kupiec, J. Pedersen and F. Chen, "A Trainable Document Summarizer", Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval, 1995, pp. , 68 - 73.

[25] J. Leskovec, M. Grobelnik and N. Milic-Frayling, "Learning Sub-structures of Document Semantic Graphs for Document Summarization", in 'Proceedings of the KDD 2004 Workshop on Link Analysis and Group Detection (LinkKDD)', 2004.

[26] M. Maleszka and N.T. Nguyen, "Approximate Algorithms for Solving O1 Consensus Problems Using Complex Tree Structure", *Transactions on Computational Collective Intelligence 8*, 2012, pp. 214-227.

[27] A. Mykowiecka, *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym*, Wyd. Polsko-Japońska WSTK, Warszawa 2007.

[28] A. Nenkova and K. McKeown , "Automatic Summarization", *Foundations and Trends in Information Retrieval*, Vol 5, Issue 2-3, 2011, pp 103-233.

[29] N.T. Nguyen, *Advanced Methods for Inconsistent Knowledge Management*, Springer-Verlag London, 2008.

[30] D. Radev, H. Jing, M. Stys and D. Tam, "Centroid-based summarization of multiple documents", *Information Processing and Management*, pp. 919-938, 2004.

[31] L.V. Pham and S.B. Pham, "Information extraction for Vietnamese real estate advertisements", *Fourth International Conference on Knowledge and Systems Engineering (KSE)*, Danang 2012.

[32] P. Potiopa, „Metody i narzędzia automatycznego przetwarzania informacji tekstowej i ich wykorzystanie w procesie zarządzania wiedzą", *Automatyka* 15(2) pp. 409-419, http://journals.bg.agh.edu.pl/AUTOMATYKA/2011-02/Auto40.pdf.

[33] F. Sebastiani, "Machine learning in automated text categorization", *ACM Computing Surveys (CSUR)*., 34(1), New York 2002.

[34] A. Singhal: "Modern Information Retrieval: A Brief Overview", *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24 (4), 2001, pp. 35-43.

[35] J. Snaider, R. McCall and S. Franklin, "The LIDA Framework as a General Tool for AGI", *The Fourth Conference on Artificial General Intelligence*, 2011.

[36] J. Sobieska-Karpińska and M. Hernes, "Consensus determining algorithm in multiagent decision support system with taking into consideration improving agent's knowledge", Proceedings of Federated Conference Computer Science and Information Systems (FedCSIS), IEEE Xplore Digital Library, Wrocław 2012, pp. 1035-1040.

[37] S. Soderland, "Learning information extraction rules from semi-structured and free text", *Machine Learning*, 34(1-3), 1999.

[38] P. Sołdacki, „Zastosowania metod płytkiej analizy tekstu do przetwarzania dokumentów w języku polskim", Praca Doktorska, Politechnika Warszawska, Warszawa 2006.

[39] S.L. Tomassen, *Semi-automatic generation of ontologies for knowledge-intensive CBR*, Norwegian University of Science and Technology, 2002.

[40] D. Trandabăţ, "Using semantic roles to improve summaries", *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 2011 pp. 164-169.

[41] R.E. Vlas and W.N. Robinson, "Two rule-based natural language strategies for requirements discovery and classification in open source software development projects", *Journal of Management Information Systems*, 28(4), 2012.

[42] A. Wawer," Mining opinion attributes from texts using multiple kernel learning", *IEEE 11th International Conference on Data Mining Workshops*, 2011.

[43] D. Weiss, "A Hybrid Stemmer for the Polish Language", *Technical Report RA-002/05,* Institute of Computing Science, Poznan University of Technology, Poland, 2005.
[44] T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis", *Computational linguistics,* 35(3), 2009.
[45] C. Zhang, X. Zhang, W. Jiang, Q. Shen and S. Zhang, "Rule-based extraction of spatial relations in natural language text", *International Conference on Computational Intelligence and Software Engineering,* 2009.
[46] E. Branny, M. Gajecki: "Text Summarizing in Polish", *Computer Science, Annual of AGH University Of Science and Technology,* 2005, pp. 31–46.
[47] J. Korczak, M. Bac, K. Drelczuk and A. Fafuła, "A-Trader - Consulting Agent Platform for Stock Exchange Gamblers", in *Proceedings of Federated Conference Computer Science and Information Systems (FedCSIS),* Wrocław, 2012, pp. 963-968.
[48] L. Sliwko, N. T. Nguyen, "Using Multi-agent Systems and Consensus Methods for Information Retrieval in Internet", *International Journal of Intelligent Information and Database Systems* 1(2), 2007, pp. 181-198. doi:10.1504/IJIIDS.2007.014949
[49] N. T. Nguyen, "Using consensus methods for solving conflicts of data in distributed systems", in: *Proceedings of SOFSEM 2000, Lecture Notes in Computer Science* 1963, 2000, pp. 411-419. doi: 10.1007/3-540-44411-4_30
[50] M. Hernes M. and J. Sobieska-Karpińska , "Application of the consensus method in a multiagent financial decision support system", *Information Systems and e-Business Management,* Springer Berlin Heidelberg 2015, doi: 10.1007/s10257-015-0280-9.