ANALYSIS OF TEXT DOCUMENTS IN AUTOMATIC ABSTRACTING SYSTEM

Stanislav Lipnitsky, Denis Nasuro

Abstract: Mathematical model of syntactic and semantic analysis of text documents is offered. On the base of this model a procedure of the detection of informative sentences in text documents is realized. Results of modeling are used in the computer system of automatic abstracting of big text documents.

Keywords: formal grammar, syntagma, syntactic tree, text analysis, corpus of texts, informativity, text abstracting.

ACM Classification Keywords: I.2.7 Natural Language Processing - Text analysis

Introduction

The main purpose of the automatic abstracting system is the intellectual analysis of big text documents. The syntactic and semantic analysis of the text goes first regardless of analytical processing methods using in the system. It consists in building of syntactic structure and finding of semantic characteristic for each sentence.

Method of syntactic and semantic analysis of the text is described in the article. This method is based on modeling of syntagma detection in the text by means of special formal grammar and knowledge base of object domain. Knowledge base is represented as a situation-syntagmatic network that consists of informative syntagmatic structures and their situational bindings.

Text syntactic analysis

Stroke-grammar.

Let $F = \langle V, N, I, R \rangle$ – a formal generative grammar, where V – *a* nonempty set of terminal symbols (will name them *words*), $N = \{I, '\}$ – a set of nonterminal symbols, I – a start symbol, and R – a grammar schema, i. e. a set of derivation rules $\alpha \to \beta$ (α and β – different chains in the dictionary $V \cup N$). Schema R of grammar F is defined as follows:

1) For any word $a \in V$ there are derivation rules $I \to a'$ and $a' \to a$;

2) Remaining derivation rules are of the form $a' \to a'b'$ or a' $\to b'a'$, where $a, b \in V$.

Symbol «*'*» (stroke) is included in nonterminal symbols for convenience. That's why grammar F is named a *stroke-grammar* with dictionary V [Кравцов, 2005]. Generated by stroke-grammar language $L(F)$ is named a *source language*, chains of this language will be called *source language sentences* or *source sentences*. Dictionary V will be called a *source language dictionary* or a *source dictionary*. Any nonempty fully ordered subset of the language $L(F)$ will be defined as a *text* of this language or a *source text*.

Syntagms and syntagmatic structures.

We shall use a relation of syntactic subordination when modeling a syntactic structure of sentences of the language $L(F)$. This relation will be defined as follows.

Let $\pi = a_1 a_2 ... a_n$ – is a arbitrary sentence of the language $L(F)$, where $a_1, a_2, ..., a_n$ – are words of the sentence. Some nonempty non-overlapping (that have no common words) subchains of the sentence π are denote by μ

and ν. We shall name a binary relation Ω_π in a set of all such subchains of the sentence π as *a relation of syntactic subordination* in the sentence π of the language $L(F)$ if:

1) For any words a_i, a_j $(i, j = \overline{1, n};\ i \neq j)$ of the sentence π $(a_i, a_j) \in \Omega_\pi$ if and only if there are subchains $\alpha a_i' \beta$, $\gamma a_j' a_i' \delta$ (or $\gamma a_j' a_i' \delta$), and $\alpha a_i' \beta \Rightarrow^* \gamma a_i' a_j' \delta$ at that (or $\alpha a_i' \beta \Rightarrow^* \gamma a_j' a_i' \delta$) in derivation of sentence π from start symbol I. Here $\Rightarrow^*$ is a symbol of derivability in the grammar F, and α, β, γ, δ are chains of the dictionary $V \cup N$. Some of chains α, β, γ, δ could be empty (possibly all). If $i < j$ (or $j < i$), than chain $a_i a_j$ (or $a_j a_i$) will be called *a syntagma* of the sentence π of the language $L(F)$. When $j \neq i+1$ (or $i \neq j + 1$) syntagma $a_i a_j$ (or $a_j a_i$) will be called *separated*, and when $j = i + 1$ (or $i = j + 1$) – *unseparated*;

2) For arbitrary nonempty non-overlapping subchains μ and ν of the sentence π $(\mu, \nu) \in \Omega_\pi$ if and only if there is such syntagma $a_i a_j$ of the sentence π, that in derivation of the sentence π from the start symbol I the chain received from a_i', and the chain ν – from a_j'. Let us denote by $\prec$ full order in the set of all nonempty non-overlapping subchains of the sentence π, corresponding to words natural order, i.e. that for all $i, j = \overline{1, n-1}$, $r, s = \overline{1, n}$ $a_i a_{i+1} \ldots a_j \prec a_r a_{r+1} \ldots a_s$ if and only if $j < r$. If $\mu \prec \nu$ (or $\nu \prec \mu$), then the subchain $\mu\nu$ (or $\nu\mu$) we shall call *syntagmatic stucture* of the sentence π of the language $L(F)$. In this case let us say that μ – *determined*, and ν – *determining* members of syntagmatic structures $\mu\nu$ and $\nu\mu$.

Union $\Omega_{L(F)} = \bigcup_{\pi \in L(F)} \Omega_\pi$ of relations of syntactical subordination in all sentences of the language we shall call *relations of syntactical subordination* in the language $L(F)$. Syntagms and syntagmatic structures of sentences of this language we shall call syntagms and syntagmatic structures of the language $L(F)$.

Syntactic tree of sentence.

If ab – is a syntagma of a certain sentence of the language $L(F)$ and $(a, b) \in \Omega_\pi$, then let us say that syntactical binding is *directed* from word a to word b. If $(b, a) \in \Omega_\pi$, then such binding is oppositely directed. Let us denote direction of syntactical binding of words by arrow that starts over determined member of the syntagma and ends over determining syntagma member (for instance, $\alpha\overrightarrow{a\beta b}\gamma$, $\alpha\overleftarrow{a\beta b}\gamma$). If direction of syntactical binding is unknown or insignificant then we shall denote it by the line over a syntagma (for instance, $\alpha\overline{a\beta b}\gamma$).

Usually syntactical bindings of words in the sentence are represented as a directed graph, nodes of graph are words, links correspond with syntactical bindings. We shall define formal notion of syntactical graph as follows.

A directed graph of relation Ω_π on the set of all words of the sentence π we shall call a syntactical graph of the sentence π. A syntactical graph of the sentence that includes only one word a we shall consider graph $(\{a\}, \varnothing)$. A syntactical graph of any chain δ, derived from the sentence π by transposition of words in it we shall call syntactical graph of the sentence π.

Let us see what form has a syntactical graph of the sentence that belongs to the source language $L(F)$.

The statement 1. A syntactical graph of any sentence of the language $L(F)$ is a directed tree (let call it *a syntactical tree*).

The proof. Let us prove it by method of mathematical induction. If $n = 1$ and $n = 2$ then syntactical graphs of word and syntagma are directed trees. Let us assume that if $n = k$ then syntactical graph of the sentence with k words is a directed tree. Let us prove that after adding one more word to the sentence, i.e. if $n = k + 1$, then syntactical graph of the sentence is still a directed tree. We shall denote added word by b. Then according to the definition of grammar F there is a word a in the sentence that is an determined member of syntagma ab or ba. If

knot *a* of a directed tree with *k* knots connect with knot *b* by link (*a*, *b*) then it's obviously that we have a directed tree again. Q.E.D. The statement 1 is proved.

Marginal syntagms.

Let $\alpha a\beta b\gamma$ (or $\alpha b\beta a\gamma$) – is a arbitrary sentence of the language $L(F)$, where α, β, $\gamma \in V^*$ (V^* – is a set of all chains in the dictionary V of the grammar G), *ab* (or *ba*) – is a syntagma of this sentence with a determined member *a* and determining *b*.

Let us call syntagma *ab* (or *ba*) a *marginal syntagma* of the sentence $\alpha a\beta b\gamma$ (or $\alpha b\beta a\gamma$), if chains *bc* and *cb* are not syntagms for any word *c* ($c \neq b$) of the sentence [Липницкий, 2005]. Word *b* of the syntagma *ab* or *ba* we shall name a *marginal word* of syntagms αb and *ba*.

Properties of marginal syntagms.

Let δ – is a arbitrary chain of set V^+ of all nonempty chains in dictionary *V*, and certain sentence π of language $L(F)$ is its subchain. If $\mu\nu$ – is a syntagmatic structure of the sentence π, then we shall consider it also as a *syntagmatic structure of chain* δ. If *ab* (or *ba*) – is a marginal syntagma of sentence π with marginal word *b*, such that for any word *c* ($c \neq b$) of chain δ pairs *bc* and *cb* are not syntagms, then *ab* (or *ba*) we shall name a *marginal syntagma of chain* δ.

Lemma. If $\rho \in V^+$, *and ab* (*or ba*) *– the marginal syntagma of chain* ρ, *and in schema R of grammar F there is a derivation rule* $a' \to a'b'$ (*or* $a' \to b'a'$), *then chain* σ *received from* ρ *by removal determining member b of syntagma ab* (*or ba*) *is the sentence of language L*(*F*) *if and only if* $\rho \in L(F)$.

The proof. Necessity. Let chain σ is the sentence of the language $L(F)$. Then necessity, i.e. existence of relation $\rho \in L(F)$ follows from the fact of existence in schema *R* of grammar *F* of derivation rules $a' \to a'b'$ (or $a' \to b'a'$) and $a' \to a$, $b' \to b$.

Sufficiency. Let there is a syntagma *ab* with a determined member *a* and determinating member *b.* Then for chain ρ there is a derivation $W = (I, \alpha, \beta, \ldots, \gamma, \mu a'\nu, \mu a'b'\nu, \ldots, \mu ab\nu, \ldots, \rho)$ in grammar *F* where α, β, γ, μ, $\nu \in V^*$. As *ab* – is a marginal syntagma of sentence ρ then, as follows from marginal syntagma definition, for any word *c* of sentence ρ chain *bc* is not a syntagma, i.e. when derive a sentence ρ rules like $b' \to b'c'$ are not used and chain $\mu a'b'\nu$ in derivation *W* is received from chain $\mu a'\nu$ by applying derivation rule $a' \to a'b'$. If we shall exclude a chain $\mu a'b'\nu$ from the derivation *W* then we shall recieve a derivation of chain σ from start symbol *I*. The case when chain *ba* is a syntagma of sentence ρ could be considered similarly. The lemma is proved.

Using this lemma it is easy to prove the following

The statement 2. If $\mu a'b'\nu$ (*or* $\mu b'a'\nu$) *– some chain in dictionary V*, *where* μ, $\nu \in V^*$, *ab* (*or ba*) *– a marginal syntagma with determining member b*, *and in schema R of grammar F there is derivation rule* $a' \to a'b'$ (*or* $a' \to b'a'$), *then chain* $\mu a'\nu$ *could be raised to start symbol I of grammar F if and only if chain* $\mu a'b'\nu$ (*or* $\mu b'a'\nu$) *could be raised to symbol I*.

The proof. Necessity. Let chain $\mu a'\nu$ could be raised to start symbol *I*. Let us prove that chain $\mu a'b'\nu$ could be raised to symbol *I*. Indeed, applying a derivation rule $a' \to a$ to chain $\mu a'\nu$ we shall get a chain $\mu a\nu$, that is a sentence of language $L(F)$, this implies chain $\mu a'b'\nu$ could be raised to symbol *I* because there are derivation rules $a' \to a$, $b' \to b$. Necessity of chain $\mu b'a'\nu$ could be proved the same way.

Sufficiency. Let now chain $\mu a'b'\nu$ could be raised to start symbol *I*. The proof that chain $\mu a'\nu$ could be also raised to this symbol follows from sufficiency of lemma. Let us apply derivation rules $a' \to a$, $b' \to b$ to chain $\mu a'b'\nu$. We shall get a sentence $\mu ab\nu$ of language $L(F)$. By virtue of lemma, chain $\mu a\nu$ is also a sentence of the language,

this implies chain $\mu a'\nu$ could be raised to symbol *I*. The proof that chain $\mu a'\nu$ could be raised to symbol *I* (if chain $\mu b'a'\nu$ could be raised to this symbol) is similarly. The statement 2 is proved.

According to the statement 2 the algorithm of source chain syntactical analysis could be constructed in the form of cyclic reduction process to start symbol by a principle "from below-upwards". The derivation rules are applied differently than at derivation of sentences: right parts of rules are replaced with corresponding left parts. Process of analysis is realized as follows. On a first step all words of source chain are marked with strokes, i.e. replaced with corresponding chains with use of derivation rules of a kind $a' \to a$ (for instance, word *a* is replaced with chain *a'*). On the second step we look for subchains of a kind *a'b'* or *b'a'* in a chain, *ab* and *ba* – are marginal syntagms with a determined member *a*, and are replaced with chains of a kind *a'* with use of rules $a' \to a'b'$ (or $a' \to b'a'$). Then the second step repeats cyclically. Process of syntactical analysis is over when we receive start symbol *I* or chain that includes more than one symbol *I*. In the latter case the analyzed chain, by virtue of the statement 2, is not the sentence of language *L*(*F*).

From the statement 2 and necessity of a lemma follows

The statement 3. If $\rho \in L(F)$ *– the any sentence, and ab* (*or ba*) *– its marginal syntagma*, *then chain* σ *recieved from* ρ *by removal of determining member b of syntagma ab* (*or ba*) *is the sentence of language L*(*F*).

The statement 3 provides receiving of the sentence of language *L*(*F*) after elimination of determining members of all unseparated marginal syntagms. According to this statement raising of syntagms by derivation rules of grammar *F* could be replaced by more effective cyclic process. On a first step of this process we look for unseparated marginal syntagms in the analyzed sentence. On the second step determining members are excluded from these syntagms. Then process repeats the same way till we get absolutely determined member as its single word in each sentence of the text.

The semantic analysis of text

Informativity of syntagmatic structures.

Informativity of syntagmatic structures could be evaluated with use of results of syntactical and statistical processing of text thematic corpuses Th_i and the full corpus of texts *Fu* [Липницкий, 2006].

Let us denote with *Sint* a set of all syntagmatic structures of the full corpus of texts *Tu*.

Let us examine the following population of events:

- S_{Th} – some syntagmatic structure α is taken randomly from the thematic corpus of texts *Th*;
- V_{Th} – syntagmatic structure α belongs to thematic corpus of texts *Th*;
- H_{Th} – occurrence of the thematic corpus of texts *Th*;
- S_{Fu} – syntagmatic structure α is taken from the corpus of texts *Fu*.

Let $P_\alpha(S_{Th} / S_{Fu})$ – is conditional chance that a syntagmatic structure α is taken from the thematic corpus of texts *Th* with the assumption that it is already taken from the full corpus of texts *Fu*. This conditional chance, as known, equals

$$P_\alpha(S_{Th} / S_{Fu}) = \frac{P(S_{Th} \cdot S_{Fu})}{P(S_{Fu})} = \frac{P(S_{Th}) \cdot P(S_{Fu} / S_{Th})}{P(S_{Fu})}.$$

Conditional chance $P_\alpha(S_{Th} / S_{Fu})$ we shall name *informativity* of the syntagmatic structure α in the thematic corpus of texts *Th*. If *Th* ($Th \subset Fu$) – is a text document, then we shall name this conditional chance informativity of the chain α in the text document *Th*.

We shall name syntagmatic structure α informative in the thematic corpus of texts (or in the text document) *Th* with *informativity level* p_0 if informativity of chain α is not less than p_0, i.e. $P_\alpha(S_{Th} / S_{Fu}) \geq p_0$.

Conditional chance $P_\alpha(S_{Fu} / S_{Th}) = 1$ as the event that syntagmatic structure α is taken from the full corpus *Fu* with the assumption that it is taken already from thematic corpus *Th* is authentic, because *Th* – is a subset of set *Fu*. Then we shall receive after simple transformations:

$$P_\alpha(S_{Th} / S_{Fu}) = \frac{P(V_{Th} / H_{Th})}{P(S_{Fu})} \cdot P(H_{Th}).$$

If we have a big enough full corpus of texts *Fu* and thematic corpus (or text document) *Th* then it is possible to consider

$$P(V_{Th}/H_{Th}) \approx \frac{n_{Th}}{N_{Th}},\ P(S_{Fu}) \approx \frac{n_{Tu}}{N_{Tu}},\ P(H_{Th}) \approx \frac{N_{Th}}{N_{Fu}},$$

where n_{Th}, n_{Fu} – are absolute occurrence frequencies of syntagmatic structure α in thematic and full corpuses of texts, and N_{Th}, N_{Fu} – quantity of all syntagmatic structures from set *Sint* in corpus *Th* and *Fu* accordingly. Then the formula for evaluation of informativity I_{Th}^{α} of syntagmatic structure α in the thematic corpus of texts (or in text document) *Th* will look like

$$I_{Th}^{\alpha} = \frac{n_{Th}}{n_{Fu}}.$$

Pragmatically full syntagmatic structures.

A pragmatically full syntagmatic structure (PF-structure) – it is a syntagmatic structure in a form of expression set that is informative in some thematic section of a subject domain (i.e. at least in one thematic corpus of texts).

Let us formalize concept of PF-structure. Let us consider some sentence $\pi = a_1a_2...a_{i-1}a_ia_{i+1}...a_n$ of source language $L(F)$, where a_1, a_2,..., a_{i-1}, a_i, a_{i+1},..., a_n – are words of the sentence. Let a_i – is an informative word of the sentence. Consistently attaching to a word a_i at the left and on the right other words of the sentence π we shall form a set Ch_0 of all its 2-words, 3-words (and so on) subchains, wich syntactical graphs are oriented trees. Let us match each selected subchain α and probability $P(\alpha)$ of its occurrence in the full corpus of texts *Fu*. We shall choose limit value p_0 of this probability and we shall remove from set Ch_0 all chains with probability of occurrence in corpus *Fu* less than p_0. Let us denote with Ch_2 a set of all remaining 2-words chains in Ch_0, with Ch_3 – 3-words chains and so on. We shall denote with Ch_j ($j \geq 2$) nonempty set with maximal index and shall introduce next concept.

All subchains of chain π of set Ch_j we shall call *pragmatically full syntagmatic structures*.

Situational-syntagmatic network.

The semantic analysis of the text in abstracting system is realized by means of the subject domain knowledge base that is presented as situational-syntagmatic network i.e. the graph [Кравцов, 2006]. Nodes of the graph are syntagmatic structures, links are their situational bindings that are formalized as situational relation in a set of syntagmatic structures.

Let us denote with *Str* a set of all syntagmatic structures of the full corpus of texts *Fu*. Then tolerance relation Θ (reflexive and symmetric binary relation) on the set *Str* we shall name *situational relation* in the full corpus of texts *Fu* if any ordered couple of syntagmatic structures (μ, ν) of the set *Str* is an element of relation Θ if and only if probability of co-occurrence of structures μ and ν in the corpus *Fu* not less than a proper limit value (a *level* of

situational binding). Saying co-occurrence of two syntagmatic structures we mean a presence of these structures in the same sentence of corpus *Fu*.

A graph of situational relation S_Θ we shall name a *situational-syntagmatic network*.

Route and graph of informativity of the text.

Let there is a text *T* (i.e. a tuple of sentences). We shall find out a subject of the text *T* and shall choose the corresponding thematic corpus of texts *Th*. (Depending on a task the subject of a text document could be found automatically by abstracting system or manually by user under the rubricator.) Let us calculate informativity of syntagmatic structures of all sentences of the text *T.* We shall use the full corpus of texts *Fu* and a thematic corpus *Th* for this purpose. We shall exclude from *T* all not informative sentences (i.e. sentences that have no informative structures). We shall get a tuple of sentences $T_{inf} = \langle \pi_1, \pi_2, \ldots, \pi_n \rangle$ in occurrence order in *T*. A tuple T_{inf} we shall name a *route of informativity* of the text *T*.

Let's build an oriented graph G_{inf}, assuming that all sentences of the route of informativity T_{inf} are nodes. Any pair of nodes π_i, π_j $(i < j,\ 1 \le i \le n-1,\ 2 \le j \le n)$ – is a link (π_i, π_j) if and only if there is a pair of linked nodes (subchains of sentences π_i and π_j) in situational-syntagmatic network S_Θ. Link is showing that there is a situational binding between subchains.

An oriented graph G_{inf} with full oder on a set of nodes corresponding to sentences order in route of informativity T_{inf} we shall name a *graph of informativity* of the text *T*.

Semantic trace of the text.

A route of informativity T_{inf} is a basis for construction of the abstract of text *T* in a form of sequence of informative sentences. Let us build a semantic trace of the text to reduce quantity of sentences in the graph of informativity. Let us define a semantic trace as follows.

A *semantic trace Tr* of the text *T* is a subgraph of the graph of informativity G_{inf}, wich nodes are all nodes of the oriented graph G_{inf}, with quantity incidental links not less than n_0. Links of the oriented graph *Tr* are all links of the oriented graph G_{inf} that ties in *Tr* only adjacent nodes (figure 1).

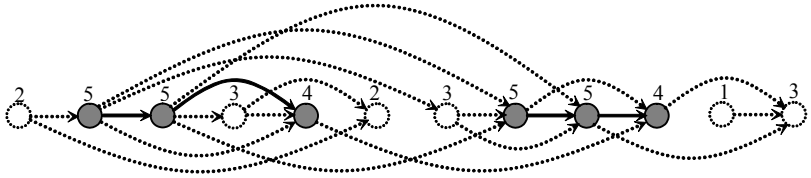

Fig. 1. An example of a semantic trace of the text in the graph of informativity

On figure 1 each node of the graph of informativity G_{inf} is marked with the number indicating quantity of incidental links. Nodes and links of the oriented graph G_{inf} that are not a part of the semantic trace *Tr* are shown as dashed lines.

A semantic trace of the text is a model of abstract that is constructed by abstract system.

Implementation in software

Visualization of informative sentences in the text.

On the basis of the offered model an experimental program of visualization of informative sentences in Russian text documents is developed [Hacypo, 2006]. The program can process source documents in the following

formats: html, txt, rtf, doc. These formats were chosen as they cover the majority of formats of available scientific-technical texts. The program can process pdf-documents by conversation to supported formats by third-party software such as paq pdf2txt [paqtol, 2006] or able2convert [able2convert, 2006] or others.

The experimental program is an executable application for Windows operating systems. The main window of the program is divided into three working areas. Top working area is for source file visualization. Left working area is for a list of found informative words. Working area at right side of the main window of the experimental program is for visualization of found set of informative sentences. Left and right working areas are empty by default and are filled with data when program processes a text document. The program has a number of toolbars. Service functions of the application are available from the program menu. These functions are as standard for windows applications then also specific for this program. They include file open, save and close operations, print results, help and settings of the program (figure 2).

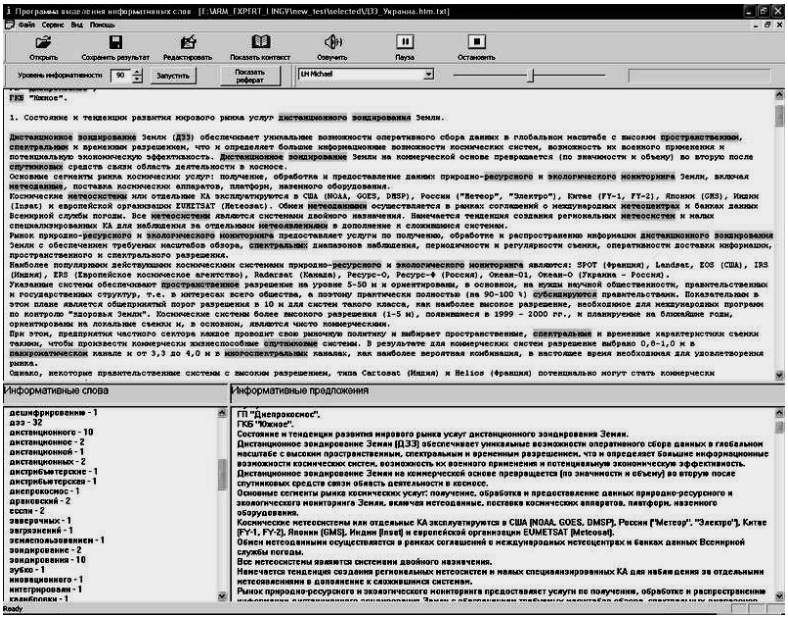

Fig. 2. The main window of the program of visualization of informative sentences

The developed application has a toolbar “informativity level” that it used to set up a necessary level of informativity. A set of informative sentences will be formed from source sentences that include informative words with informativity equal or more than a defined level.

The program analyses quantity of words in the source document to chose an algorithm of search of informative words in big documents or in small texts. When program processes a big document using algorithm for small texts it selects the useful thematic information. The thematic corpus of texts is a kind of thematic filter in this case.

The program uses inflexion paradigm. It is very important especially for Russian that is an inflexional language.

Found informative words are highlighted. The program forms a file with selected informative sentences that could be edited and saved.

The program allows to see a context of the informative sentence that is highlighted in edit mode (figure 3).

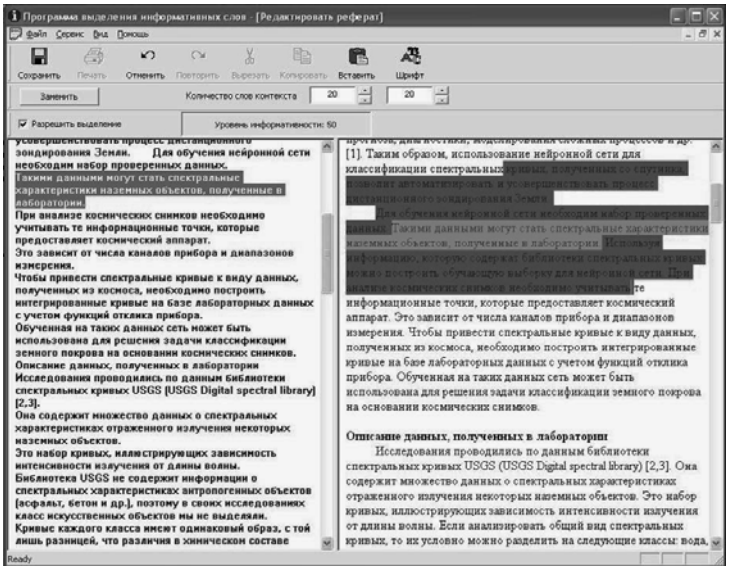

Fig. 3. The edit mode of the program

The user can adjust the length of context.

The program forms an information portrait of the document. The information portrait is a set of PF-structures most often used in the document and a set of informative words with occurrence frequencies. The program allows to adjust quantity of displayed PF-structures and informative words.

When process a document the program uses statistical data from an inflexion paradigm dictionary. This dictionary is formed from thematic corpuses of texts. Thematic corpuses of texts are formed with special software that is developed in the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. Data obtained as a result of the processing is a knowledgebase of the experimental program of informative words visualization.

The program is developed on C++ programming language.

Conclusion

The offered mathematical model of text documents analysis allows to realize detection of informative sentences in big texts.

Monothematic fragments of the text document could be detected in a set of informative sentences with use of situational-syntagmatic network.

A complex of special software is developed on the base of results of mathematical modeling. The software is for automatic abstracting of Russian texts. The size of the abstract can be adjusted by setting up a syntagma informativity level and text semantic trace characteristics.

Bibliography

[Кравцов, 2005] Кравцов А.А., Липницкий С.Ф., Насуро Д.Р., Прадун Д.В. Интеллектуализация процессов обработки текстовой информации // Информатика. – 2005. – № 1. – С. 41–51.

[Липницкий, 2005] Липницкий С.Ф. Семантический анализ текста на основе ситуативно-синтагматической сети // Информатика. – 2005. – № 2. – С. 102–110.

[Липницкий, 2006] Липницкий С.Ф. Математическая модель и алгоритмы формирования схемы грамматики, порождающей проективные предложения // Весці НАН Беларусі. Сер. фіз.-тэхн. навук. – 2006. – № 3. – С. 71–75.

[Кравцов, 2006] Кравцов А.А., Липницкий С.Ф., Насуро Д.Р. Синтез рефератов текстовых документов на основе ситуативно-синтагматической сети // Искусственный интеллект. – 2006. – № 2. – С. 172–175.

[Насуро, 2006] Насуро Д.Р. Алгоритмы и программы визуализации информативных предложений в системе автоматического реферирования текстовых документов // Искусственный интеллект. – 2006. – № 2. – С. 416–419.

[paqtol, 2006] A web site of pdf-conversion tool. – http://www.paqtool.com

[able2convert, 2006] A web site of pdf-conversion tool. – http:// www.investintech.com/prod_a2e_pro.htm

Authors' Information

Stanislav F. Lipnitsky – Doctor of Sciences, United Institute of Informatics Problems, National Academy of Sciences of Belarus, Surganova str. 6, Minsk, 220012, Belarus; e-mail: lipn@newman.bas-net.by

Denis R. Nasuro – Research fellow, United Institute of Informatics Problems, National Academy of Sciences of Belarus, Surganova str. 6, Minsk, 220012, Belarus; e-mail: nasuradr@newman.bas-net.by