

UDC 681.5:004.82(045)

¹Eugen B. Artamonov, Assist Prof.
²Oleksandr O. Zholdakov, Assoc. Prof.

CONCEPT OF CREATING A SOFTWARE ENVIRONMENT FOR AUTOMATED TEXT MANIPULATION

National Aviation University

¹E-mail: eart@ukr.net

²E-mail: olexazholdak@mail.ru

The basic concepts which have been used in the new program environment of texts editing "Intelligent book" were considered. The Primary goal of program environment is creation of electronic books with the content which automatically changes according to the level of information presentation depth. The basic mechanism of program work is settled on creation certain connections between text blocks, which allow to correct view of questions representation, which are revealed in the given text, automatically.

Розглянуто концепції, які були використані в новому програмному середовищі редагування текстів «Кмітлива книга», основною задачею якої є створення електронних книг з рівнем подання інформації, який автоматично змінюється. Механізм роботи програми базується на створенні певних зв'язків між блоками тексту, які дозволяють автоматично коригувати зміст питань.

Рассмотрены концепции, использованные в новой программной среде редактирования текстов «Умная книга», основной задачей которой является создание электронных книг с автоматически изменяющимся уровнем представления информации. Механизм работы программы базируется на создании определенных связей между блоками текста, позволяющими автоматически корректировать содержание вопросов.

Statement of purpose

The problem of automated detection the text content is not new.

For over than 30 years both separate scientists (N.M. Leont'eva) and scientific departments of number of institutes, primarily in Russia (The Institute of Linguistics of the Russian State Liberal University, The Institute of informatics problems of RAS, Russian SRI of Artificial Intelligence, The Kazan's state University [1]) are engaged by this problem.

The results of scientific researches became separate programmatic developments – Dialing, Polixtext, Galaxy-ZOOM, and even open-source projects (one of the most known is the Automatic text processing [2]).

But there is a task which has no decision until now – it is creation of adequate to the necessity-ties of concrete user electronic variant of book product by the help of computer programs in the automatic mode.

Even approaches to the decision of this scientific task are not enough formalized or take contradictory character until nowadays.

That, as soon as work reaches to recognition of text sense or even logical construction of connections between heterogeneous information in the book text, one have to use knowledge and ability of experts from the proper field of knowledge and bibliography.

The necessity of personnel training, mediocre speed of information treatment, subjectivism, possibility of errors, and on the whole – a high cost of treatment the large volumes of text, does expedient development and use of software for automation of this process [3].

Exactly on crossing of these two directions – automatic text processing and development of bibliography-robot, – lies the idea of creation the software environment «Intelligent book».

The basic task of this software environment is creation of electronic books with content which automatically changes according to the level of information presentation depth. This product on nowadays design time needs «hand» tuning (what in course of time possibly will be transferable on automatic) and provides the performance of the followings objectives:

- creation of the book itself;
- determination of the complete plural of connections between the book's text and it' content;
- filtration of the given information on the basis of the personal interest of user of questions which are lighted up in a book.

The basic mechanism of the program work is based on creation of certain connections between the blocks of text in the electronic book environment, which allow to correct the presentation of subject of this text in the automated mode.

Conception of the “Intelligence book” software product construction

The given problem needs the decision of the followings two questions:

- correct fastening of text elements, what will allow to represent information both in a complete form and as an abstract; thus an abstract form must maximally pass the common content and, at a necessity, expose the separate fragments of text, which can interest a reader;
- adjusting the flexible mechanism of watching the personal interest of reader, this will allow him to give the timely access to necessary information.

If the decision of the first question can be laid on the textbook author or group of experts which determine the richness of separate text elements content, the second question lies on the crossing of psychology and programming, where the basic task of programming is watching signs of the personal interest in text, among which are direct and sides signs.

The personal “interest” concept does not have a clear determination in psychology until now.

There is a point of view, that attention is determined as an arbitrary or involuntary orientation and concentration of psychical activity on some object of perception. And this concentration does not appear in a “clean” kind.

Unlike cognitive processes (perception, memory, thought, etc.) attention does not have the special sense (determination); it appears as being inside of these processes and inseparably from them.

Basic properties of attention in this case are volume, concentration, firmness, oscillation, switching. The volume of attention is measured by the amount of objects which are perceived simultaneously. The objects incorporated according to their sense are perceived in a greater amount, than not incorporated. For the grown man the volume of attention is equal to 6-8 objects. Separately here distinguish involuntary and arbitrary attention [4].

First who made an attempt to expose the concept of attention from the mechanical point of view was Ribo in the motive theory of attention, and later this approach was developed in behaviorists and reflexologists. In this case attention is taken to the reflex options. The second attempt was related to the theory of Gestalt psychology and reduced the phenomenon of attention to the structures of the sensory field [5].

There can be no doubt that reflex options act substantial part in initial, most primitive forms of attention. It is well known that influence on the organism of some irritant usually cause organism to reflexly adapt to its best perception. Thus, for example, influence of sound on the ear-drum irritant cause a reflexive turn toward the source of sound.

Obviously, that using this principle it is possible to watch and general signs of attention to the text for a reader. As basic elements for attention control were select.

1. Time of being on a page (only a middle area gets out, because a very rapid transition means absolute incuriosity, and large time of

revision can mean the loss of the personal interest) is a basic parameter of control which relates to direct properties of attention. In subsequent modifications of the system this parameter must change automatically, depending on the features of concrete user;

2. Motions of manipulator (for every reader there is a scenario of the interested transition – transition without the rapid returning on the page of content – in which the set of motions enters by a manipulator during the revision of text). This parameter needs the special tuning on a user and relates to the second properties of attention.

Technology of the correct text elements fastening

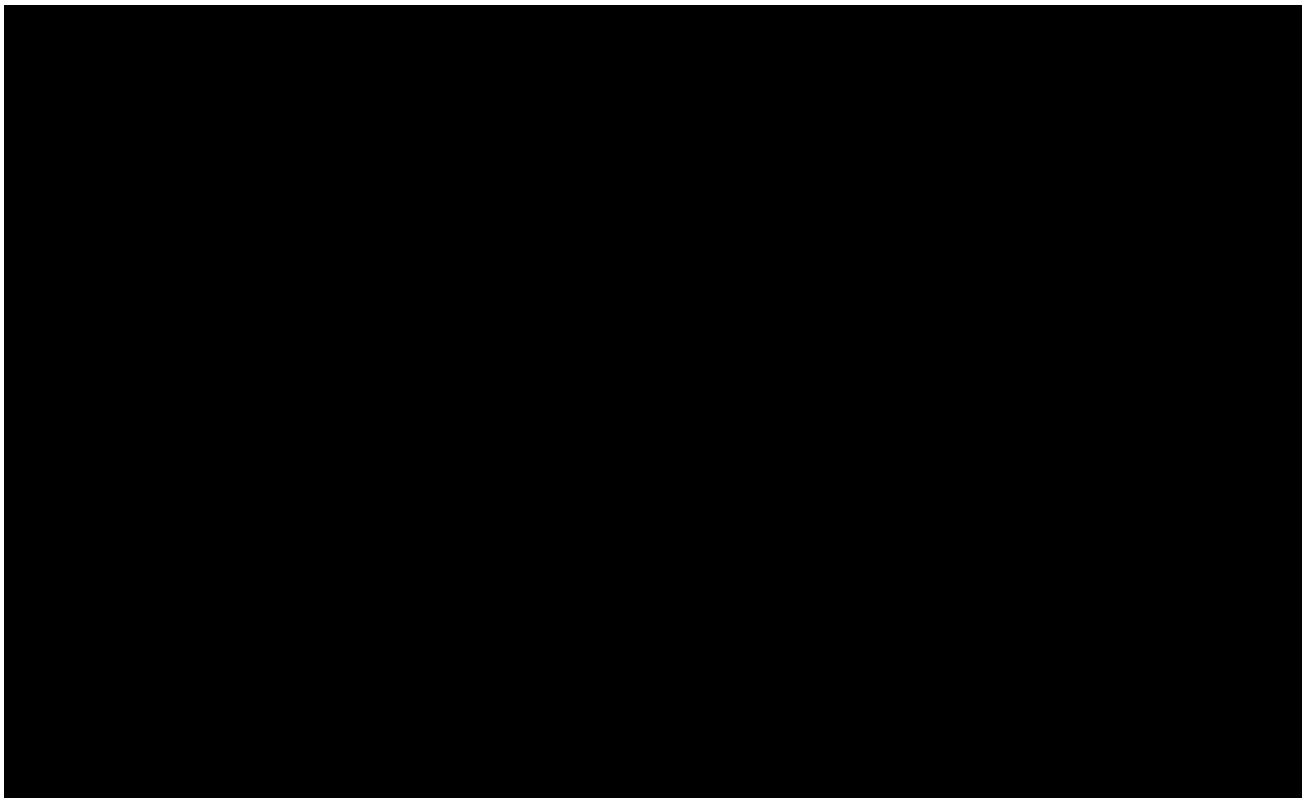
For the correct fastening of text elements, which will allow representing information both in a complete form and as an abstract, it was decided to present content of book text after a chart, represented on figure.

From this chart it is evidently, that there are three levels information presentations in the program. They are intended to grant the information to concrete user in the volume in which it is needed on the first stage of the personal interest.

The first level: it is a text without any reductions and treatments. This level is intended for use of readers who are only introduced the thematic of a book.

The second level: at this level the text of book passes certain processing (more detailed it is described below) which will allow the user with a shallow or middle knowledge of book subject to obtain most exact information, without a necessity to read what he already know.

The third level: information at this level passes the detailed treatment. User at this level has only a common picture of the book subject and requires just separate exact theorems and formulas without any explanations.



Schema of book content division on 3 levels

On each of these three levels it is necessary to give user the possibility of access to the searching system, which allows finding all necessary connections in a book, which could not get to the authors needed connections.

Inserting own semantic connections can be presented as a separate branch of the program development. But this mechanism will require not only grant permission for user to the management of semantic connections, but also requires creation of separate instrument for verification contradictory connections of absence.

Programmatic facilities of watching the personal interest by text contents

The question of watching personal interest of reader needs programmatic realization of such functions:

- control of time being on the page which automatically translates all elements of this page on the proper level of representation (also all elements of text, which are richly in content related to the elements of this page, get greater status);
- control of transitions between elements, which have rich in content copulas (this control requires the flexible extra charge of opening the separate text elements status marks);
- control of indirect user actions, such as manipulator moves, keys pressing, copying and insertion of text elements into other programs.

Obviously time have to be a basic control parameter, what reader outlays on an acquaintance with the text element. But even the book author cannot exactly present the expected time of reading the select group of text fragments.

That is why the next formula of nominal presentation time calculation of the group of selected for showing elements was selected as a standard measure:

$$T_N = \sum_{i=1}^A (K_i \cdot T_T) + T_F \cdot F + T_P \cdot P,$$

where T_N is normative time for consideration of selected text fragment which consists of A indentions, F of formulas and P of pictures;

K_i is amount of letters in an i indention;

T_T is mean time on reading of single character;

T_F is mean time on understanding of one formula;

T_P is mean time for consideration of one picture.

For determination of mean values of T_T , T_F , T_P , which can be different for different users, is used program unit, at reading of text from screen, requires to push the button upon termination of reading, and at consideration of formulas and pictures – to answer question to them (time is taken into account only in the case of positive answers).

After determination of time, necessary for mastering of this text element, all elements related to him are opened for access and the additional opening marks begin to be counted for all connections of the first and second levels (after achieving the proper amount of total marks they also will open).

Thus user does not manipulate transitions from one level to other. But he gets possibilities to preliminary define the level of opening the text contents or specify on erroneous, according to his opinion, copulas.

As was said higher, user can get possibility of creation his own connections for adjusting of material opening procedure in obedience to own desires.

The use of network mechanisms for the improvement of the program quality work

On the initial stage of using any new book in the described environment the basic problem is determination of possible time being on a page.

The mechanism of determination of mean time is described in a previous section.

However, it is almost impossible to define limit values (maximum and minimum) according to this scheme. But exactly these values with greater authenticity specify on the personal interest in the selected fragment of text.

For the decision of this task a network version, which foresees work with the electronic book of not one user, but whole group, was developed.

Thus working with an electronic book is possible not only in the conditions of network usage. Connection to the server can be made only for sending of data about own statistics and receiving of updates about statistics of other users.

Every user passes registration and gets unique type which will be saved not only for current electronic book, but also for all next books with which this user will work in this software environment.

Having possibility of collecting the statistical information, it is possible to move from the evaluation of normal time of reading a fragment after a middle formula which is used for all fragments, to determination of this time for every text element.

Conclusion

Potential of the use of similar software environment of books with internal semantic copulas is very high; in fact everyone would like to get necessary data quickly, without the considerable charges of time on an information retrieval, as in an ordinary book, and what is main – consistently.

Also a question of common acquaintance and receiving of the initial understanding of textbook or book content is important. Taking into account the large volume of information exactly the mode of acquaintance can be most used for work with materials of books in this software environment.

It is obviously expedient to consider possibility of application of this software product in the controlled from distance education at creation of electronic courses and books.

Certain application this program will also be able to get in works of bibliographies of scientific editions in matters of automation.

References

1. Биряльцев Е.В. Особенности лексико-семантической структуры наименований артефактов реляционных баз данных / Е.В. Биряльцев, А.М. Гусенков, М.Р. Галимов // Тр. Казанской школы по компьютерной и когнитивной лингвистике TEL-2005. – Казань: Казан. гос. ун-т, 2006. – Вып.9. – С. 4–12.
2. *Project* of automated text process. – <http://www.aot.ru>.
3. *Лингвистический* процессор для сложных информационных систем / Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др. – М.: Наука, 1992. – 256 с.
4. *Вопросы* развития познавательных интересов учащихся в процессе обучения / под общ. ред. Д.К. Гилева. – Свердловск: Свердл. кн. изд-во, 1970. – 143 с.
5. *Світ* психології. – <http://www.psychology.net.ru>.