Hypertext : Data Model To Reduce Disorientation

R. N. Jugele* and Dr. V. N. Chavan
*Department of Computer Science, Science College, Congress Nagar, Nagpur. Maharashtra,
Head, Department of Computer Science, S. K. Porwal College, Kamptee, Dist : Nagpur. Maharashtra.

Abstract - ***A mapping from text units to representation structures is computed based on linguistic and background knowledge about the domain of discourse. The link context defines which part of the currently displayed document is replaced or affected by following the link. Any visible components which are not affected remain on the screen unaltered. In fact the type of a structure can be considered such a high level specification. Each of these should be compiled into a set of low level presentation specifications and link contexts on the items constituting the structure.***

Keywords – **Linear, microstructure, model, hypertext, hierarchical, index, mapping, orientation, frames.**

I. INTRODUCTION

While retrieved a set of related texts the user has to face the following problems:

- Based on the best match paradigm[21] most information retrieval systems measure the relevance of a document according to the degree of similarity between document representations and query. A user being interested in pieces of information scattered over several texts may therefore have to read couples of redundant texts before reaches a text which contains a new informational item.
- The presentation techniques employed in conventional retrieval systems are poor. In the worst case the user is confronted with a list of references.

Macrostructures :

The oldest science which deals with structuring and formulation of text is rhetoric. The special fields of rhetoric traditionally are:

- Inventio: The discovering of the ideas which shall be expressed in a text
- Dispositio: The ordering of ideas
- Elucutio: The finding of adequate formulations

Writing a text is interlocking process of inventio, dispositio and elucutio which leads stepwise refinement. The hypertext author provides a network structure (dispositio) which contains pieces of information (inventio) verbally expressed in text units (elucutio). The final ordering of these text units are linear hypertext paths.

Macrostructures are sequences of propositions expressed in first order calculus and derived from the microstructure of a text by the application of macro-rules. These macro-rules are semantic transformation rules mapping tuples of propositions to more general propositions and produce general descriptions of the text. Thus a hierarchy can be built which reaches from sentence topics over paragraph and chapter topics to the topic of the whole text.

Following are four macro-rules:

- Information which is not needed to understand the subsequent text is deleted
- Information which can be inferred by presupposition is deleted
- Special information is replaced by more general information
- A proposition is constructed which comprises a set of propositions from the text or macro-propositions from macrostructures

Macrostructures in hypertext :

Linear texts macrostructures give a hierarchical representation of the text's topical structure on a meta level which reflect the process of generalization in text analysis or stepwise refinement in text generation. Both of these processes are ruled by a special notion of relevance. Where text generation is driven by contextual relevance, text analysis is guided by textual relevance[28]. It is a meta level description of the deep structure of the complete hypertext network. Hypertext navigation is a construction of texts called hypertext paths, built up from a set of given pieces. Elements of these path-macrostructure are macro-text-units, derived from the original text units by the application of macro-operation. These macrostructures in hypertext comprise the relevant information of a set of text units in a condensed form. These text units are from different linear texts, therefore macrostructures in hypertext reflect special aspect of intertextuality of document fragments[1]. A hypertext containing text fragments taken from any magazines can be traversed as per a path:

- It gives a description of one special device show in figure 1
- It compares two or more devices with respect to their properties shown in figure 2

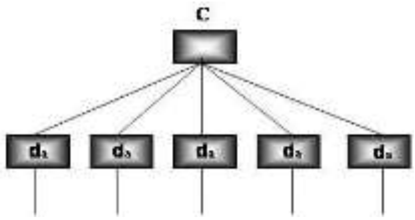


Fig. 1. Various text units shows aspects of a single concept

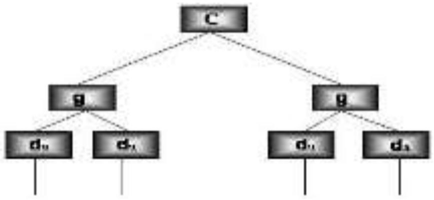


Fig. 2. Generic concept are compared with respect to their properties

Hypertext is a methodology for reading or writing in a non-linear environment and have two primary components

Nodes : These are receptacles for information in the form of text, video, etc.

Links : It represent connections between nodes, creating paths the user travels to access the desired information.

The hypertext data model is characterized by a hierarchical schema mechanism that allows a predetermined, open-ended schema embedded in the hyperdocument.

It uses two types of nodes:

- Concept nodes: It provide organizational structure and contain the ideas, subideas and provide organizational structure.
- Information nodes: It contain text and other data

Existing hypertext systems have following problems:

- Users can feel disoriented i.e. not knowing where they are in the network[15].
- Users may suffer from information myopia in which the large amount of available data makes it difficult to decide which information is important [19].
- It is often difficult to customize hypertext systems to meet the needs of specific applications [14].
- The non-linear representation of hypertext information can present problems for the system designer and user. When designing a hypertext system, for example, one such difficulty comes from the fact that it is unnatural to break thoughts into discrete units.

Users access information like in an outline, progressing through the main ideas until the required concept is found. The hierarchical structure is maintained through a system of implicit links between concept nodes.

II. DOCUMENT MODELS

Models for creating paper documents are the Open Document Architecture (ODA)[14][20] and Standard Generalized Markup Language (SGML)[3]. Without the loss of generality, the terminology from ODA is used. In [14] the layout or geometric structure is defined as the hierarchy of objects resulting from decomposing the document, based on the way it is presented on paper.

The Dexter model[7] is in use for creating hypertext structure. The atomic components and composite components are like in the ODA structures, the unbreakable data items and their hierarchical composition. They form the nodes in the hypertext. Anchors are used to address locations within an atomic component. They are considered at this level to assure independence from the physical representation of the components and finally the model defines links to capture relations.

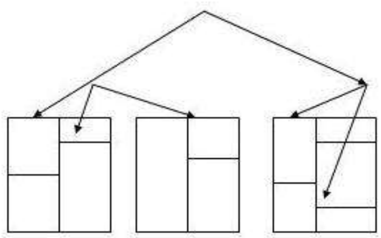
Unidirectional link l consists of specifications of its starting point $s(l)$ and its ending point $e(l)$. For a set of links L, $s(L)$ and $e(L)$ denote the set of all starting and ending points respectively of the links in L. $d+(n)$ and $d-(n)$ denote the number of incoming and outgoing links of a node n.

III. STRUCTURED HYPERDOCUMENT

$H = \{N, S\}$

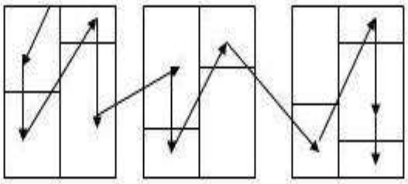
Here N is the set of nodes in the hyperdocument. These are basic components in the hypertext and composite components can be treated in a different way. The set S is the structure set[6]. It can also be viewed as an extension of the hypertext structures in [19]. An element $s \in S$ can be written as $\{Ns,Ls\}$. The set Ls is a structure specific set of links providing some relational structure on the set of nodes Ns, their grouping within a structure and/or their anchors. Each s captures a specific structure in the hyperdocument. Following are different types of structure in S each impose different constraints on Ns and Ls as shown below.

a. **hierarchical structure:**



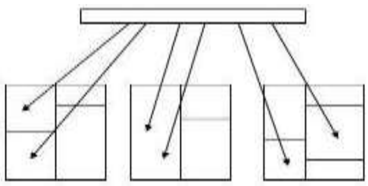
Ns is a set with a tree structure. *Ls* is the set of links required for accessing the nodes pointing to the different children. In a hyperdocument the prime hierarchical structure of importance is the logical structure. The geometric structure is only relevant in the document analysis phase.

b. **linear structure:**



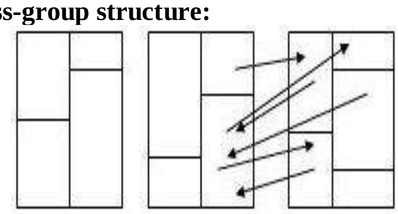
The elements of *Ls* lead to each elements in *Ns* without reaching the same node twice. The most important linear structure is the reading order which corresponds to a depth-first ordering of the text blocks in the logical structure[25]. Here the navigation channels provide means to go forward and backward in the structure relative to the current position. The only node one can jump to directly is the first node in the structure. When a hierarchical structure is placed in the navigation channel some visual representation for the different levels is required. Hence they can be used as the source of a link which one can follow to the related node.

c. **Index structure:**



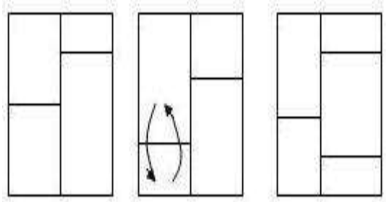
Here $s(Ls)$ contains the single element n with $d+(n)>=1$ and $d-(n) = 0$. *Ns* is equivalent to $e(Ls)$. It is the common index to a document. For example, an index based on all labels in figures or important keywords in the text. Here the visualization of the node have a direct connection to the content of the node. Content and navigation are clearly separated and link context defines that the currently displayed node in the content channel should only be replaced when it is not the component the activated link points to.

d. **Cross-group structure:**



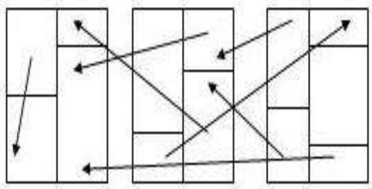
Here $s(Ls) \cap e(Ls) = \varnothing$. The set *Ns* is formed by the union of $s(Ls)$ and $e(Ls)$. This class of links is important in the application as these connect labels in a figure with the text. One other cross-group structure set relates the collection of figures and the whole text. It identifies the parts in the text where certain figures are described. The sets $s(Ls)$ and $e(Ls)$ are called each other's scopes thus the pieces of text related to one specific figure form the scope of that figure. Here the displayed component should only be replaced when it is not the component the activated link is pointing to. When separate channels for the two components forming the cross-group are used, subsequent activation of links in either direction assures that both components are displayed and remain displayed as long as links within the cross-group are used. The navigation channels should display the different ways of leaving the cross-group structure.

e. **Side-loop structure:**



$Ns=\{n1,n2\}$ and $Ls=\{l1,l2\}$ with $s(l1)=e(l2)=n1$, $s(l2)=e(l1)=n2$ and $d+(n2)=d-(n2)=1$. Examples of side-loops are footnotes, explanations, asides and appendices. They should be readable upon request but should not influence the reading order in the document. Here context plays an important role. When it is activated the content of the displayed node should be replaced by the content of the endpoint of the link. Navigation control should be adjusted such that the only navigational possibility is to go back to the place where the node was activated. A pop-up window is a typical example implementation of this behavior.

f. **Cross-reference structure:**



Here *Ls* is the set of relations it is given by $Ns = s(Ls) \cup e(Ls)$.

Presentation models:

The link context defines which part of the currently displayed document is replaced or affected by following the link. Hypertext structures needs two channels:

- For displaying the content of the nodes
- Presents the navigation controls to the user

Navigation controls provide access to the nodes and links in the structure. For paper manuals two content channels are used:

- for text
- for figures

The channel for a figure is composed of two sub-channels for displaying the figure itself and its caption.

IV. HYPERTEXT MODEL

linking motivates the choice of semantic networks for hypertext in

- TEXTNET[27]: It employs a semantic net-work which directly connects text by semantic and rhetoric relations.

- Thoth-II[5]: It is better suited for automatic integration of new text segments, it is provided with a semantic network modeling.

The connection between text and conceptual knowledge is established by a string-oriented matching procedure. If the handling of hypertext nodes is emphasized, object oriented approaches to hypertext are preferred[4][29]. The planning of hypertext dialogs requires additional features like agent models for the support of an extrinsic task by an agent/task driven hypertext[10][9] or constraints a task of planning argumentation[23] is supported by constraints on graphical objects.

The hypertext model has to fit the following context:

- Automatic decomposition and analysis of texts : A mapping from text units to representation structures is computed based on linguistic and background knowledge about the domain of discourse. Text is fragmented to text units and mapped to representation structures by automatic text analysis as shown in figure 3.
- Based on these representations semantic relations between text units and macrostructures of sets of text-units may be established as shown in figure 4.

Presentation of hypertext paths depends on a user formulated query, text plans and prototypical informational objects which control the mapping from semantic objects to graphical objects[26] and thus form the elements of a graphical text presentation language[16].

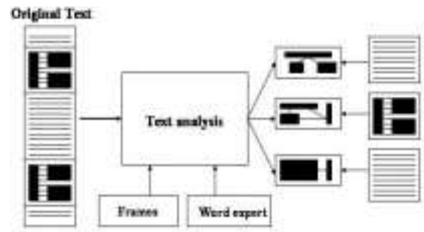


Fig. 3. Mapping to represent structure by automatic text analysis[11][12]

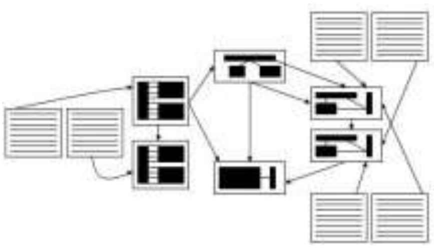


Fig. 4. The Structure of the hypertext depends on semantic properties of text unit representations

A uniform representation formalism can be used for conceptual knowledge, text plans and graphical structures is the frame construct[18]. Frame like structures like case frames[8] or scripts[22] are widely used in linguistics and text understanding. The relation between the constructs of frame and object-oriented languages. The programming paradigm prevailing in computer graphics[13][2][24].

V. HYPERTEXT DATA MODEL

It consists of three entities

Hypertext = < *Nc*,*Ni*,*L* >

Nc a set of concept nodes representing the structure, *Ni* a set of information nodes containing information for that application and *L* a set of implicit and explicit links.

Concept nodes contain a key phrase that identifies the information stored in the structure, it also contain explicit links to other nodes. Data is contained in linear linked lists of information nodes and it is owned by a specific concept node. Information nodes hold the data in an application specific format. A concept node may or may not have information nodes associated with it. Information nodes associated with concept nodes are expected to contain data relating to the topic implied by the concept node's key phrase.

Implicit links are derived from the structure and order of the nodes. It connect concept nodes to other concept nodes or to information nodes. Explicit links connect any two nodes and used to represent non-hierarchical relationships[17] as shown in Figure 5 and specific example is shown in Figure 7.

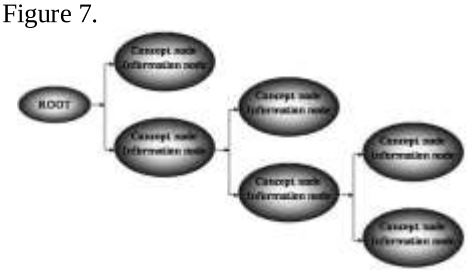


Fig. 5. Form of Hyperdocument

Nodes are embedded in a hyperdocument in two ways:

1. nodes and links can be simulated with concept nodes or added to the model as a new type of object.
2. nodes and links can also be installed in more elaborate models like Dexter or Trellis, applying the schema on top of the existing model.

The primary goal are:

- To determine if the use of a schema in a hypertext system improves the use of that system
- To determine the user's orientation within a hypertext system in order to assess orientation capabilities
- To evaluate how hypertext schema mechanism system assist users in maintaining orientation
- To evaluate how hypertext assists users in determining information seeking is located in relation to their current location

- To assist the user in choosing the correct path to access the desired information
- To evaluate ease of use to improve the user interface
- To check the acceptability, effectiveness and reliability of hypertext schema from the viewpoint of user

VI. THE BASIC STRUCTURE

The text units contained in the hypertext are mapped to sets of frames which is built up by a set of slots, which is associated with a set of permitted and a set of actual entries. Additionally an activation weight is assigned to frames slots and actual entries. The structure is formalised as cascade of partial mappings as shown in figure 6.

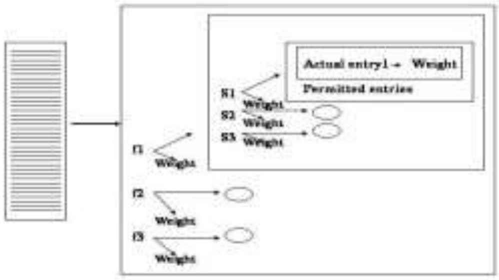


Fig. 6. Cascade mapping of hypertext representation

$$HTREP := \{f \mid f : Tu \cup \{w,q\} \rightarrow FRAMES\}$$

$$FRAMES := \{f \mid f : Names \rightarrow \{<s,w> \mid s \in SLOTS, \; w \in \{rejecting, irrelevant, relevant, dominant\}\}\}$$

$$SLOTS := \{f \mid f : Names \rightarrow \{<act, perm, w> \mid act \in ENTRIES, perm \in 2^{Names}, \; w \in \{rejecting, irrelevant, relevant, dominant\}\}\}$$

$$ENTRIES := \{f \mid f : Names \rightarrow \{irrelevant, relevant,\}\}$$

Following functions are defined, which allow access to:

- The frames of a text unit
- The weight of a frame
- The slots of a frame
- The weight of a slot
- The permitted entries of a slot
- The actual entries of a slot
- The weight of an entry

VII. EXPERIMENTAL TEST

This is a hypertext system built using schema mechanism and the process is using new model on the specified data set. In this experiment the object for the users to locate a specific word and return the related fact. Each system contained the same words and facts. All users searched for the same words.

Navigation used the implicit linking of the schema. No explicit linking is used. The words are organized into a hierarchical structure with four levels. The system contained concept nodes and information nodes. The words are grouped using primarily intuitive concepts and user could not rely on prior experience with word searching. Level is grouped according to:

1st Level : Function (Movie, TVSerial)
2nd Level: Grouping associated on 1st Level :
1(Movie): Bollywood, Hollywood, Troliwood
2(TVSerial): Hindi, Marathi, English
3rd Level : First letter of the desired word
(a, b, c,……, z)
4th Level : listing of words

First and second level node explaining the concept. Fourth level have associated information node that contains the fact which user is searching for.

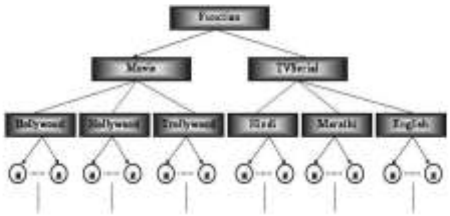


Fig. 7. Concept nodes in the experimental test

The desired information is accessed by reading the displayed sentence then keying on a word (anchor) to move to another sentence. There is no suggested paths, user has to determine where to go to access the information. The text is structured so that the user has to move through at least four sentences before the desired fact could be located. This minimized the differences in access times and keystrokes between the two test situations.

The 30 metrics used for the evaluation of the hypertext systems came from several sources. The first set of metrics (1–26) was from a questionnaire answered by the user. Four additional metrics were also used for evaluation. The accuracy of the information retrieved by the user was metric 27. The questions asked by the user were evaluated as metric 28. The total time of session was metric 29. The logged data of the user's keystrokes was metric 30. This metric was later broken into several individual components. This log of keystrokes is created automatically by the system while the users were working without their knowledge.

VIII. CONCLUSION

The concepts of text oriented linguistics like macrostructures and paragraphs as text segmentation units are fruitful for hypertext. A frame-oriented hypertext model formalize macro-operations for hypertext.

Schema consists of concept nodes that organize the document and information nodes containing text and other material. It allows to modified the structure of the schema. This testing showed that a hypertext system designed is easier to use and allowed users to find the desired information faster. These results indicate that hypertext data model is a viable and valuable tool for hypertext system design.

Model for hypertext could serve as the basis for making the overall structure of the document clear to the user. The mental structure of the information in the user's head may be very different from the node and link structure provided by the system. The model designed to serve as a basis for a hypertext system into which data can easily be placed, located or removed without the user becoming disoriented using integrated hypertext schema.

REFERENCES

[1] Begthol, C. 86: Bibliography Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and Cognitive Act of Classifying Documents. In: Journal of Documentation, Vol. 42, No. 2, 1986, pp. 80-113.

[2] Bobrow, D.G. / Winograd, T. 77: An Overview of KRL-0, a Knowledge Representation Language. In: Cognitive Science, Vol. l, No. l, 1977, pp. 3-46.

[3] B.C. Watson and R.J. Davis. ODA and SGML: an assessment of co-existence possibilities. Computer Standards & Interfaces, 11:169{176, 1991.

[4] Christodoulakis, S. / Ho, F. / Theodoridou, M. 86: The Multimedia Object Presentation Manager of Minos: A Symmetry Approach. In: Sigmod Record, Vol. 15, No. 2, 1986, pp. 295-310.

[5] Collier, G.H. 87: Thoth-II Hypertext with Explicit Semantics. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 269-289.

[6] F. Garzotto, L. Mainetti, and P. Paolini. Adding multimedia collections to the Dexter model. In Proceedings of the European Conference on HyperText, pages 70{80, 1994.

[7] F. Halasz and M. Schwartz. The Dexter hypertext reference model. Communications of the ACM, 37(2):30{39, 1994.

[8] Fillmore, CJ. 68 The Case for Case. In: Bach, E. / Harms, R.T. (eds): Universals in Linguistic Theory. New York, 1968, pp. 1-88.

[9] Garg, P.K. 87: Abstraction Mechanisms in Hypertext. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 375-395

[10] Garg, P.K. / Scacchi, W. 87: On Designing Intelligent Hypertext Systems for Information Management in Software Engineering. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 409-432.

[11] Hahn, U. / Reimer, U. 86: TOPIC-Essentials. In: COLING-86. Proceedings of the llth International Conference on Computational Linguistics, 1986, pp. 497-503.

[12] Hahn, U. / Reimer, U. 88: Knowledge-Based Text Analysis in Office Environments: The Text Condensation System TOPIC. In: Lamersdorf, W. (ed): Office Knowledge: Represen-tation, Management and Utilization. Amsterdam, 1988, pp. 197-215.

[13] Hollan, J.D. 84: Intelligent Object-Based Graphical Interfaces. In: Salvendi, G. (ed): Hu-man—Computer Interaction. Amsterdam, 1984, pp. 293-296.

[14] I.R. Campbell-Grant. Introducing ODA. Computer Standards & Interfaces, 11:149{157, 1991.

[15] Jeff Conklin, 'A survey of hypertext', Technical Report MCC Technical Report Number STP- 356-86, Rev. 2, Software Technology Program, (December 3, 1987).

[16] Lakin, F. 87: Visual Grammars for Visual Languages. In: AAAI87 — Proceedings 6th Nat. Conf. on Art. Int., Vol. II, Los Altos, 1987, pp. 683-688.

[17] L. Hardman, D.C.A. Bulterman, and G. v. Rossum. The Amsterdam hypermedia model: Adding time and context to the Dexter model. Communications of the ACM, 37(2):50{62, 1994.

[18] Minsky, M. 75: A Framework for Representing Knowledge. In: Winston, P. (ed.): The Psychology of Computer Vision, New York: McGraw Hill, 1975, pp. 211-277.

[19] R. A. Botafogo, E. Rivlin, and B. Shneiderman. Structural analysis of hypertext: Identifying hierarchies and useful metrics. ACM Transactions on Information Systems, 10(2):142{180, 1992.

[20] R.N. Jugele and V.N. Chavan: ODA: A Study of Document Design**,** International Journal of Emerging Trends & Technology in Computer Science**,** ISSN:2278-6856,Vol 2 , issue 1, Jan-Feb - 2013, pp. 194-198.

[21] Robertson, S.E. 80: Some Recent Theories and Models in Information Retrieval. In: Harbo, O. / Kajberg, C. (Hg.): Theory and Applications of Information Research, London, 1980, pp. 131-136.

[22] Schank, K. / Abelson, R. 77: Scripts, Plans, Goals, and Understanding. Hillsdale, 1977.

[23] Smolensky, P. / Bell, B. / Fox, B. / King, R. / Lewis, C. 87: Constraint-Based Hypertext for Argumentation. In: Hypertext '87 Papers, Chapel Hill, NC, University of North Carolina, 1987, pp. 215-245.

[24] Stefik, M. / Bobrow, D.G. 86: Object Oriented Programming: Themes and Variations. In: The AI-Magazine, Vol. 6, No. 4, 1986, pp. 40-62.

[25] S. Tsujimoto and H. Asada. Major components of a complete text reading system. Proceedings of the IEEE, 80(7):1133{1149, 1992.

[26] Thiel, U. / Hammwöhner, R. 89: Interaktion mit Textwissensbasen — Ein objektorientierter Ansatz. In: Tagungsband der Jahrestagung der Gesellschaft für Informatik 1989, to appear.

[27] Trigg, R.H. / Weiser, M. 86: TEXTNET: A Network-Based Approach to Text Handling. In:ACM Transactions on Office Information Systems, Vol. 4., No. l, 1986, pp 1-23.

[28] Van Dijk, T. 79: Relevance Assignment in Discourse Comprehension. In: Discourse Pro-cesses, Vol. 2, No.2, 1979, pp. 113-126.

[29] Woelk, D. / Kim, W. / Luther, W. 86: A Object Oriented Approach to Multimedia Databases. In SIGMOD Record, Vol. 15, No.2, 1986, pp. 311-325.