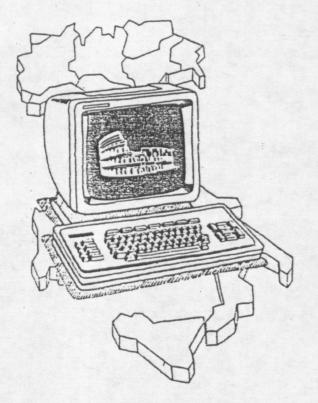


ASSOCIAZIONE ITALIANA PER L'INFORMATICA ED IL CALCOLO AUTOMATICO



ATTI DEL CONGRESSO ANNUALE ANNUAL CONFERENCE PROCEEDINGS

Volume 2



and another the second can be problem whether course on the property and another an extension of

WHAT CAN WE DO WITH IMAGES IN A MULTI-MEDIA DOCUMENT FILING SYSTEM?

F. Rabitti (*), P. Stanchev (**)

- (*) Istituto di Elaborazione della Informazione Via S. Maria 46, 56100 Pisa
- (**) Institute of Mathematics Computing Center Bulgarian Academy of Sciences, Sofia

Classification (ACM Comp. Rev.): H3.3 I4

Keywords: Office information systems, user interface, multimedi information, information retrieval, image processing, imag retrieval.

Abstract

The problem of retrieving multimedia documents by image conten specification is here addressed. The fuzzy set approach for image de scription, allowing the measure of image similarity, has been adopted Moreover, the association of different types of information to image descriptions is exploited for more efficient retrieval.

1. Introduction

The office information system of the future should be able to effectively handle multimedia documents. Multimedia documents are composed of attribute information (often called formatted data), text information, voice information, image information or any mix of the above types (Tsichritzis83). A key component of this office information system will be a subsystem for the filing and retrieval of these multimedia documents.

We can envisage the office information system of the future as a set of workstations connected through a Local Area Network (LAN). These workstations will support sophisticated user interfaces capable of displaying in all their aspects these multimedia documents (i.e. good resolution for images, audio equipment for voice, etc.). The workstation will also support, in its user interface, capabilities for entering, modifying and querying these documents.

Connected to these workstations through the LAN there will be a specialized file server for efficient storing and retrieval of sultimedia documents. Performance and cost effectiveness will be the sharacteristics of the Multimedia File Server, given the large amount of document we can expect in a office information system and the space equired for storing image and voice items. For this purpose the sultimedia File Server will probably use technologically advanced evices as optical disks together with magnetic disks for document toring, Optical Character Readers (OCR) and facsimile scanners for ocument entering, laser printer for document output.

In essence we can think of the Multimedia File Server plus ne workstations, where the user interface is implemented, connected brough the local area network, as an entire subsystem for the anagement of multimedia documents. The enphasis of this subsystem will; in the retrieval of the multimedia documents based on some adication of their content and structure. It is the same principle of the database systems where data items are retrieved on the ground of their content rather than their location. This aspect is still glected in actual office information systems (for example, Xerox ar). This problem has been extensively studied for attributes, in tabase management systems, and for text, in information retrieval stems.

An innovative approach which integrates attributes and text uses e signature techniques, as superimposed coding (Christodoulakis84). this case the document retrieval activity is performed on an abstractor of the document. This process is faster even if a limited error is troduced.

A decisive challange of the office information systems of the future will be to allow the retrieval of the multimedia documents through their image and voice parts. However the general pattern recognition problem is still too difficult to solve given the technological limitations of today. The use of the abstraction level for the retrieval of image and voice items is difficult to exploit.

This paper will focus on the problem of accessing multimedia documents through the content and structural specification of the contained images.

2. Images in office multimedia documents

The present work is mainly addressed to the particularity of image analysis and retrieval in the context of a multimedia document system. We are not trying to address the problem of the general image processing which is still far from effective solutions at the today state of the art (Ballard82).

In our case, we want to combine the limited capabilities in the system for recognition of images with other well known techniques that can be applied to other types of information. This should result in a much more powerful and efficient possibility of describing and retrieving images in office documents.

We suppose that an image in a multimedia document is composed of different types of information: an image representation, such as raster form, an image explanation, such a caption and eventual annotations (text strings), and image attributes (formatted data). In this manner, what it is difficult to perform for content recognition on the raster form only, can be made easier if combined with the other textual and formatted parts of the image.

Moreover an image can be structured in objects, that is regions of the image which can be seen as logical constituents of the image itself. This partitioning process can be recursively applied, subdividing regions into subregions and so defining a hierarchical logical structure of objects in the image. We suppose to associate the graphical characteristics of these objects, as position in the image and boundary of the region, with other properties described by different types of information, such as object caption, object annotations and object attributes.

Thus, for the activity of searching the content of an image we bring the advantage of coupling image pattern matching and text and attribute search, down to the internal structure of the image, at the internal object level.

We intend for a caption a text string explaining the content of

an image or an object. This is an objective description in the sense that is a part of the document, and reflects the point of view of the document itself (that is, its author). Instead an annotation is a text string containing the comment of a particular person, who puts this annotation about an image or an objects. This is a subjective description or explanation where the indication of the author's name is essential. The annotation is not part of the original document.

For searching captions and annotations, information retrieval techniques can be used. They allow substring search, also with meta-characters, in an efficient manner using appropriate indexing techniques (Salton83). They also help in overcoming the problem of different naming for the same thing or concept. In fact a thesaurus allows to lefine synonyms chains.

For searching attributes, database management techniques can be used. They make available powerful languages for query specification, several indexing techniques for fast retrieval etc. (Ullman82). However, unlike in usual databases, it is not possible to define a fixed schema for image and image object attributes. Each object must be independently defined giving the name, the type and the value. Attributes ogically referring to the same property can be named in a different manner and the user himself may not know how to refer to them. In this case some synonym check mechanism for attributes names can help the user in attribute definition and query formulation. This mechanism should ask for the on-line user confirmation in the decision about quivalent attribute names since semantic is involved in this process.

Another aspect stressed about multimedia documents is the ossibility of logical relationships between elements of different ocument components with different types. Usually in a document the omponents have hierarchical structures, i.e. a text component can be ecomposed into sections, subsections, paragraphs. However lateral elationships representing some logical association could be stablished, for example, between a text section and an image. The roposed structuring of images allows a more precise definition of ateral relationships between image logical elements and other document omponents. For example, an image object can be associated to a text aragraph.

The main activities concerning image processing in a multimedia ocument management system are the image analysis, storing and etrieval.

.1 Image Analysis

In the image analysis phase the image is structured according to he knowledge of the system. The rules specifying how an image can be tructured are described in a grammar. The adoption of a grammar allows to describe an image structure with a formal language, which is more easily understandable by the system.

Unfortunately, the precision of formal languages does not allow to describe concepts as similarity or approximate type recognition, which are extremely useful for images. This characteristic of formal languages has led to their failure in the application to natural language description. In order to overcome this problem we adopted the approach of fuzzy grammars (Lee69). They have already been used in the area of natural languages (Zadeh72) and picture recognition (Pavlidis68), (Lee83).

They allow to define a class of image structures (i.e. an image type) with unsharp boundaries, that is, a class in which the transition from membership to non-membership may be gradual rather than abrupt. This means that an image can be recognized in a fuzzy grammar at a certain degree, which is not only the answer yes or no (two-valued logic). Moreover, using fuzzy productions instead of usual productions we enlarge the set of productions and we obtain a much more flexible object description.

In the image analysis, the additional information consituted by attributes and captions (annotations are usually added at a later stage, when users retrieve and comment the image) is associated to the image and/or its constituent objects.

The image analysis is a partially automatic and partially interactive process. Many steps can be fully automatized depending on the level of sophistication of the system (i.e. facsimile scanner for entering images, OCR for reading captions, etc.). However in many activities the system cannot completely perform its tasks (i.e. in scanning as image for recognizing elementary objects many ambiguities can arise) so the help of an on-line user may be essential. The use of a format specification language, such as a fuzzy grammar, can greatly simplify this process since can serve as a guide for both the system and the user in their decisions.

2.2. Image Filing

Image filing involves the storing of both the image presentation (i.e. raster form) and the image internal information which result from the analysis phase. This internal information is constituted by the image internal structure described in a formal language and the added information, that is attributes, captions and, later on, annotations associated to the various components of the structure.

Special techniques (i.e. compression) may be used in this proces to save storage space and other techniques (i.e. something equivalen to clustering) may be used for faster access.

2.3 Image Retrieval

Image retrieval is the key point of this approach. In fact a the effort described for the image analysis phase is justified by the advantage which derives in querying the image.

The image retrieval based on the image content is part of the broader process of multimedia document retrieval based on document content. The query language in this environment consists in a filter specification by the user. This idea is an extention of the use interface proposed in the Query-by-Example (Zloof75) relational database management system.

The user specifies in a filter all the information he recall about the document content and structure. He builds a filter as he thinks the target document looks alike. Thus the filter will contain the minimal requirements that the stored documents should have in order to pass the filter and to be displayed to the user for final confirmation. The same approach is used for querying images: the user specifies a filter containing the essential characteristics of the image to be searched. In fact it should be possible to combine the specifications on images as well as on the other document components (of different types) in the complete filter.

The filter on images should combine some specifications about the image content and structure, according to be grammar defined for the application, and some specifications about the other types of information associated: attributes indicating properties, captions containing descriptions, annotations containing personal comments.

Moreover, query reformulation should also be enforced. In fact, the user should also be allowed to change the filter specification as he begins to sense the result of his query, that is, when the first stored images matching the initial filter specification are being displayed. This allows an interactive searching process converging towards the target image.

Filtering images is not an exact process, unlike the query process on data objects in a database management system. In fact it is extremely difficult to describe an image in very precise terms. Thus the possibility of introducing the approximate matching in the image retrieval process is essential. For this reason the use of fuzzy grammars has proven very effective, since it makes possible to define a degree of similarity for images (Lee72). This also allows to rank the retrieved images in decreasing order of similarity with the filter specification.

3. Image Analysis

In the following we are going to describe more in detail the three fundamental activities of image analysis, filing and retrieval. We will only mention the technical aspects pertaining to the area of general image processing, giving mainly the references, while we will focus on the innovative aspects which can be exploited in a multimedia document management subsystem contained in a office information system.

The purpose of image analysis is to determine the complete structure of the image: the objects contained, their attributes and captions.

In order to obtain the complete image description, a strong interaction between the user and the system, helping each other, is supported.

For the image analysis we propose a methodology composed by four steps (see Figure 1):

STEP 1 element recognition;

STEP 2 object recognition;

STEP 3 attribute and caption definition;

STEP 4 complete image description.

3.1 Element recognition

The purpose of this step is to partition the image space into meaningful regions, called elements. The elements should be the basic components which will constitute the building blocks of the image structure.

One approach is to locate the boundaries as edges of regions. Another approach is to group image points into similar regions, so determining the boundaries. Another possible technique exploits the concept of texture. A texture can be considered as a repetition of a basic pattern defined over a local region. Differences in texture over such a region can be used in order to define the edges (Gonzalez77).

In our particular environment we can couple one of the previous techniques for partitioning the image space into basic elements with another technique. It consists in trying to match the possible elements in the image with a catalog of stored elements constituting the pictorial representation of the terminal nodes in the grammar adopted for the application. In this process, different variations as changes in size, rotation, translation (i.e. using discrete Fourier transform) can be attempted. Two main problems should be solved in this approach for element recognition. Firstly, elements can partially overlap in the image shading part of each other shape and making difficult their recognition. Secondly, these elements can be portrayed from different perspectives in the image, so it is necessary to store in the catalog several different projections of prototype elements (Foith81).

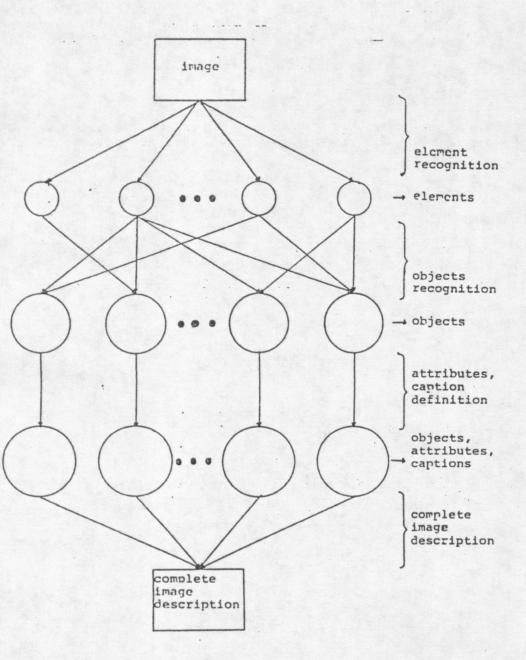


Fig. 1 Methodology for image analysis.

If this automatic process fails an interactive dialog with a user, expert of the image database content, can be invoked.

3.2 Object recognition

The purpose of the object description step is to organize the image elements into meaningful relational structures, called objects.

A promising technique for the description of these relational structures is to use fuzzy grammars (Lee69). In fact using fuzzy grammars instead of usual grammars allows a much more flexible object description. Fuzzy productions represent more closely the real world than "black and white" productions, since they can indicate any intermediate grade of membership, ranging from 0 to 1, and not only the values 0 and 1.

So far the main applications of fuzzy grammars lay in the area of modelling fuzzy automata (Dubois79). The fuzzy approach has been used also in systems for medical diagnosis and questionnaire (Stanchev83), (Barnev74).

$$A \xrightarrow{\varrho} \beta$$
, $A \in V_N$ and $\beta \in (V_T \cup V_N)^*$

where 0=0≤1 is the grade of membership of β given A.

If $\alpha_1, \alpha_2, \ldots, \alpha_n$ are strings in $(V_T \cup V_N)^*$ and $S \xrightarrow{\varrho_1} \alpha_1 \ldots \alpha_{m-1} \alpha_m$ $(0 \leqslant \varrho_1 \leqslant 1)$ the grade of membership of $x \in L(G)$, where L(G) is the fuzzy language generated from the fuzzy grammar G, is given by:

$$\mu_{L(G)}(x) = \sup \min (\varrho_1, \varrho_2, \dots, \varrho_m)$$

where the supremum is taken over all the chains from S to x.

For the application of fuzzy grammar for our environment we have slightly changed the definition. Now a fuzzy grammar is a five-tuple G = (V_N, T_T, P, V_S, S) , where V_N, V_T, P, S are as usual and $V_S \subseteq V_N$.

We define V_S as the set of admissible non-terminals which allows the recognition of partial sentences: x is a partial sentence if $A \in V_S$ and $V_S = \dots = x$.

Every A \in V is a label for a syntactic category of type "partial sentence". In our approach we associate a partial sentence with an identified image object, so we can call them "object sentences". Let us have a context free fuzzy grammar G = (V_N, V_T, P, V_S, S). We adopt this grammar for the image description.

Positional relational operators (like, near, right, left, above, selow, contained or a combination of them) are also included in the fuzzy grammar productions to describe the relative position of elements in the image.

In order to perform the object recognition, we apply several times the fuzzy grammar G to the image, trying to arrange all the elements (identified in step 1) into languages of G.

During this process the following situations can happen:

- .) An object, or a part of it, belongs to several languages with different grades of membership. In this case we choose the language with the highest grade of membership;
- !) Some elements cannot be connected into a language, so some objects cannot be correctly described in G. In this case either the user interactively helps the system trying another segmentation (back to step 1) or an expert user is asked to make some changes in the fuzzy productions P in G in order to solve the problem.

This method for object recognition can be seen as a procedure which derives a fuzzy set from the processed image. This fuzzy set is composed by couples, whose first element is a label in V_S , and the econd element is the grade of membership of this label to the image. ach object in the image is thus associated to a couple in this fuzzy set.

1.3 Attribute and captions definition

The purpose of attribute and captions definition is to add more nformation to the object description.

In this step of image analysis only attributes and captions are intered while the annotations are added later. In fact they reflects the observations about the objects in the image by some user that has reviously retrieved the document. Attribute and caption definition is ually requires some explicit intervention of the user assisting the locument input.

We suppose that a set Ω of attributes is defined in the system reflecting a wide range of possible properties that can be associated with the kind of images expected in the database. In this step, for every object X (where X \in V) identified in step 2, a membership function is defined for every attributes in Ω . The membership value of A

for X expresses the degree according which the property A characterizes the object X in the present image. This membership function takes the values 1 or 0 if the attribute is valid or not for the object in the given image. However the membership function can take intermediate values if the corresponding property characterizes only partially the object in the image (i.e. the membership value quantifies A for X). In this manner the attribute membership definition is nicely adapted to the fuzzy set definition for image objects.

The caption definition for an image object consists in entering a text string describing in words the object itself. Its only purpose is to help in looking for an image containg an object with a certain description, using an information retrieval technique for text searching.

3.4 Complete image description

After the entire image has been partitioned into objects and the attributes and captions have been defined for them, the system continues his analysis process trying to find a language in G which completely describes the image.

We suppose that for the fuzzy grammar adopted, a set V_I can be defined containing the labels for all possible image types in the database. We define: $V_I \subset V_S$, and $V_I = \{I | S - I \text{ is in P}\}$.

Thus every I \in V is a label for a syntactic category of type "image sentence". The purpose of this step is to associate to the complete image a language starting from some I \in V I.

If several such languages are found, the language with the highest degree of membership is chosen.

4. Image storing

In our approach, together with the image representation (i.e. in raster form) we need to store the image description as obtained in the four steps of the image analysis.

This image description includes:

- 1) the image name, which is a label in $\boldsymbol{v}_{\boldsymbol{I}}$, and the membership grade.
- 2) The description of all the objects in the image. Actually we need to store not only the objects (labels in V_S) found in step 2 of image analysis, but also all their contained objects (labels in V_S) found as intermediate results in that process.

- Each object description includes:

- a) the object name, which is a label in V_{S} , and the membership grade;
- b) the attributes with their membership value;

- c) the object concion;
- d) any eventual object annotation.

For physical organization of the image description either linearly linked lists or quintary trees (Lee80) can be used.

Image retrieval

In our approach the user queries the image database specifying a filter. In this filter he enters all the essential characteristics that he thinks the sought image should have . We can also say that user describes to the system a prototype of the wanted image.

Using fuzzy techniques for image description allows us to define the distance between two images. In this manner, as a result of an image search we can find an ordered set of retrieved images. In this set, the retrieved images satisfying the query conditions are ranked by decreasing similarity with the prototype image, that is by increasing distance with the prototype image (for this purpose we can use relative Hamming distance between fuzzy sets (Kaufmann75)).

We propose three steps for image retrieval:

STEP 1 Prototype creation

STEP 2 Searching function definition

STEP 3 Browsing through the retrieved images.

Since we want to emphasize the query reformulation, at each moment during STEP 3 the user if not satisfied by the answers he is getting, should be able to go back to STEP 1 and change the prototype. This constitutes an interactive process that (hopefully) will converge to the sought answer.

5.1 Prototype Creation

In the prototype image creation the user should put all his knowledge of the wanted image. During this process the user creates the image objects that he thinks are contained in the target image.

In order to create each object in the prototype image the user has the following choices:

- a) the user can ask the system, using some menu driven interface, for the image icons stored for the image element corresponding to the terminals (in $V_{_{\rm T}}$) of the fuzzy grammar.
- b) using some interactive graphic package, the user can compose the previously selected image elements, make some changes or draw some part of the object. During this graphical composition of the object in the prototype, the system computes the corresponding fuzzy set.

or complains if it is not able to do so (that is, the object cann be associated to a language in G).

After each object definition in the prototype, the user specifithe values for the associated attributes, captions and annotatio (partial string matching and synonym search is used for captions a annotations).

The prototype image is completely defined when all its objec are defined. At this point the system can determine the fuzzy set corresponding to the prototype.

5.2 Searching function definition

In this step the user must define the degree of similarity I wants between the prototype image and the retrieved images.

For this purpose linguistic predicates, such as "exact", "more α less similar", "very, very similar", "similar within a tolerance of α can be specified (Lee83). The system then translates them in terms α fuzzy set dinstance.

5.3 Browsing through retrieved images

In this step the user is presented with the images retrieved from the image database, ranked according to the degree of similarity (if the fuzzy set terms) with the prototype specified for the query. If the topmost images in this ranked set are completely unsatisfactory, if means that the prototype definition was not enough accurate. In this case the user must go back to step 1 and then modify the prototype (using the same procedure).

If after several iterations he does not find the wanted image either that image is not in the database or the image specification to a fuzzy set of the fuzzy grammar G is not satisfactory, and G should to modified by an expert.

6. An Example

Let us have the context-free grammar $G = (V_N, V_T, P, V_S, S)$.

.....

Let
$$V_T = \{a_1, a_2, ..., a_{12}\},\$$

where a_1 , a_2 , ..., a_{12} are given in Figure 2.

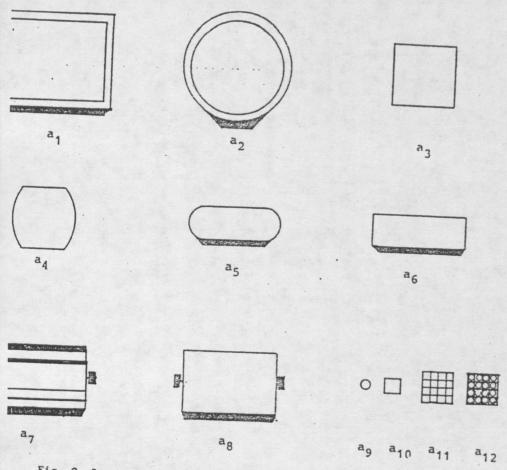


Fig. 2. Icons for elements corresponding to terminal symbols.

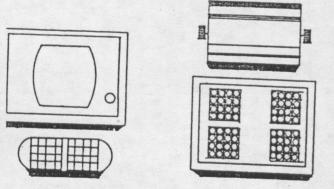


Fig. 3 Stored image.

S - personal computer

S - typewriter

S - television set

S - keyboard for personal computer

S - keyboard for typewriter

where A - television set box

B - screen

C - television set buttons

D - keyboard box for personal computer

E - keyboard buttons for personal computer

F - typewriter part

G - keyboard box for typewriter

H - keyboard buttons for typewriter

- Basic

Q₂ - Fortran

- Office Software

- Educational Software

- Latin alphabet

Q - Cyrillic alphabet.

and
$$V_I = \{S_1, S_2\}$$

..... Some elements of P are given in Table 1,

where A.B means A is near to B,

ADB means A includes B.

Let us have these sets of attributes:

for
$$S_3 = \{Q_1, Q_2, Q_3, Q_4\}$$

" $S_4 = \{Q_5, Q_6\}$
" $S_5 = \emptyset$
" $S_6 = \{Q_5, Q_6\}$
" $S_7 = \{Q_5, Q_6\}$

Let us have the following image (Figure 3). The following attributes, caption and annotation are added (Table 2).

In this case our image description is given in figure 4.

A possible dialog between user and system for image retrieval is given in Appendix 1.

7. Conclusions

Storing and retrieval of documents containing images will be a key problem of the office information system of the future. The exposed approach tries to explore the potentiality of the access through the images. Although this is a very difficult problem in general, it seems more feasible in this environment. More research is needed for a better specification and testing of this approach in order to effectively exploit image search, as it is being done for attributes and text.

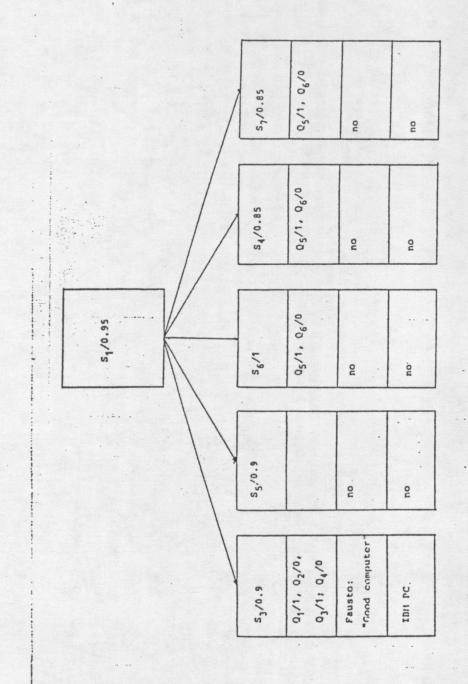


Fig. 4. Image description.

$s \xrightarrow{1} s_1$	$s \xrightarrow{1} s_2$	s ₁ :95 s ₄ .s ₃ =0 ₃	s ₃ 1 s ₅ .s ₆
S ₅ 1 A-B.C	$\lambda \xrightarrow{1} a_1$	∧ ,9 a ₂	A -2 ag
B .9 → a3	B .9 a4.	$c \xrightarrow{1} a_9$	$c \xrightarrow{1} a_{10}$
c 15 a11	c .5 > a ₁₂	$c \xrightarrow{1} c.a_9$	c 1 -> c.a10
c 1 -> c.a11	$c \xrightarrow{1} c.a_{12}$	S ₆ 1 D=E	$ \begin{array}{c} n & \xrightarrow{1} & a_5 \end{array} $
D .9 → a ₆	E 1 > a11	E 1 a12	E 1→ E.a ₁₁
E 1 > E.a ₁₂	$s_4 \xrightarrow{1} F.s_7$	F 1 37	F .8 a8
5 ₇ 1 → G-H	G -85 a1	G .15> a2	$G \xrightarrow{.95} a_5$
$G \xrightarrow{1} a_6$	н .5 > в	н .5 > a ₁₀	H 1 -> a11
H 1 a12	н 1→ н.а9	н 1 → н.а 10	H 1→ H.a ₁₁
н 1 > н.а 12	$s_2 \xrightarrow{.5} s_4$	$s_2 \xrightarrow{1} s_2.s_4$	

Table 1

Object	Attributes	Annotation	Caption
s ₃	Q ₁ /1, Q ₂ /0, Q ₃ /1, Q ₄ /0	Fausto: "Good computer"	IBM PC
s ₄	05/1, 06/0	no	no
s ₅		no	ro .
s ₆	05/1, 06/0	no	כת
s ₇	05/1.06/0	no	no

Table 2

	Appendix 1
THE SYSTEM	THE USER
a) object names; b) continue; c) drow; d) stop objects	С
drow	
a)keyboard for personal computer; b)?	ь
a) keyboard for type-writer; b)?	a
keyboard for type-writer. Attributes:	
Latin alphabet	1
Cyrilic alphabet	0
Caption:	no
Annotation:	ne
a) object names; b) continue; c) drow; d) stop objects	Ь
type-writer part	
- The state of the	
the state of the s	
a) b) c)?	a
type-writer	
Active 19 of types of the	
B	
The second secon	

type-writer. Attributes:	
Latin alphabet	1
Cyric alphabet	10
Caption:	no
Arnotation:	ne
a) object names; b) continue; c) drow; d) stop objects	a
a)personal computer; b)type-writer; c)television set;	
d)keyboard for personal computer; e)keyboard for type-	
writer	a
television set box	国际型型排头 发
a) b) c) d)?	unknown
screen .	
a) b) c)?	ь
television set buttons	
a) b) c) d) e)?	unknown
elevision set.	no
nnotation:	no
a) h) c)?	a

keyboard buttons a) b) c)7	unknown
Keyboard for personal computer	
a) b)?	a
keyboard for personal computer. Attributes: Latin alphabet	1
Cyrilic alphabet Caption:	no .
Annotation:	no
personal computer	
a) b)? personal computer. Attributes:	a
Basic	1 0
Fortran Office software	1
Educational software	0
Caption: Annotation:	IBM PC
a) object names; b) continue; c) drow; d) stop objects	6
search function not object	exact
search function	very simular

(77.) Rabitti F., Stanchev P., "What Can We Do with Images in a Multimedia Document Filing System", *Proc. Annual Conf. of the Italian Computer Society (AICA)*, Rome 1984, 439-461.