A Contribution to Image Semantic Analysis

Jozef Stašák
University of Matej Bel - Faculty of Finance, Slovakia
jozef.stasak@umb.sk

INFORUM 2004: 10th Conference on Professional Information Resources
Prague, May 25-27, 2004

This contribution deals with semantic analysis of image. The image is divided into areas called *segments*. Each segment may have assigned one or more semantic networks. These semantic networks are applied when providing an image description or completing image based on segments and might be created based on a verbal description of image or based on verbal facts creating basis for completing image. However, this contribution deals with a life cycle of image, structure, features and creation of the above-mentioned semantic networks, as well.

1. Introduction

Recent research into integrating Image Processing techniques has produced a number of new methods for semantic analysis of image content, which plays a role of great importance, when considering creation and update of image databases and an appropriate delivery of information contained in these databases. Many different methods have been developed for semantic analysis of image utilizing various types of algorithms. However, important semantic information necessary to interpret image content is mostly not represented in single pixels but in meaningful image objects and their mutual relations. These objects may be closely related to fractals or segments, which represent the analyzed image structure units applied for multi-scale image analysis. A procedure based on Fractal Net Evolution is considered o be an efficient method to describe complex semantics within largely self-constructing and dynamic networks. It combines insights in the fractal structure of the world and of semantics with object orientation. The procedure first extracts image objects, which afterwards are classified by means of fuzzy-logic. Basic strategy is to build up a hierarchical network of image objects, which allows representing the image information in different resolutions simultaneously [1]. Subsequently, many other methods applied for semantic analysis of image content are based on existing of objects called segments; regardless the image to be analyzed is of static or dynamic nature [2, 3, 6, 7 and 8]. The approach described in subsequent sections of this contribution is also based on the theory of image segments, which carry their own semantic content. However, these segments are decomposed again in two types of subordinated structural units: clusters and gasps and their semantic content is quantified via partial semantic functions, while an information capability of the image to be analyzed from semantic point of view is formalized via its final semantic function. Any image, its segments, clusters or gasps may be described via text in natural language and any text in natural language in form of text fragments or semantic networks of facts and information closely related to the above-mentioned image structure units. Information and facts contained within image segments, units and clusters and text fragments are usually stored in appropriate records or documents and it is needed to have an adequate interconnection among these facts and the above-mentioned documents. This interconnection is assured via databases called: reference databases. This approach is being implemented via adequate application program, which consists of three principal modules: extraction, representation and retrieval module.

2. Image – Structure & Life Cycle (Processing)

2.1 Structure of Image

Let us consider an image, which has an appropriate information capability, represented by the set of facts or information. This image consists of (n) segments, while one set of facts or information (SFI) describes one image segment, from semantic point of view (see also Fig.1). A function defined on selected SFI is called*: partial semantic function* (S_{fp}) and its value is called: *partial semantic value* (S_{fpv}). A logical product of partial semantic functions and their semantic values creates a *final semantic function* S_{ff} with their appropriate *final semantic value* (S_{ffv}) The following formulas formalize the above-mentioned statements.

$$\text{Image} = \prod_{i=1}^{n} \text{Segment}(i) \qquad (1)$$

$$\text{Segment}(i) \longrightarrow \text{SFI}(i) \qquad (2)$$

$$S_{fv}{}^{i} = S_{fpv}{}^{i}(\text{SFI}) \qquad (3)$$

$$\text{Segment}(i) = S_{fpv}{}^{i} \qquad (4)$$

$$S_{ffv} = \prod_{I=1}^{n} S_{fpv}{}^{I} \qquad (5)$$

$$\text{Image} = S_{ffv} \qquad (6)$$

Each segment represents a piece of image and has its *header* and *body*. The segment's body represents a visible piece of image and consists of two types related subordinated areas. The first type closely related to subordinate areas is called *cluster*[1] and the second type of subordinated areas is called *gasp* (see also Fig.2). Clusters and gasps are considered to *be visible parts of an image segment*. A segment header contains pointers closely related to semantic networks and their components, which describe image and its structural units (clusters and gasps) from semantic point of view. A segment header is considered to be an *invisible part of the image segment*.

[1] The cluster may have any form and size

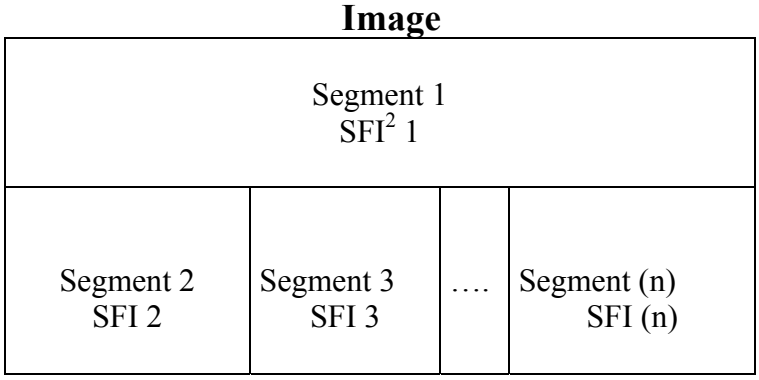


Fig. 1
Image divided into segments

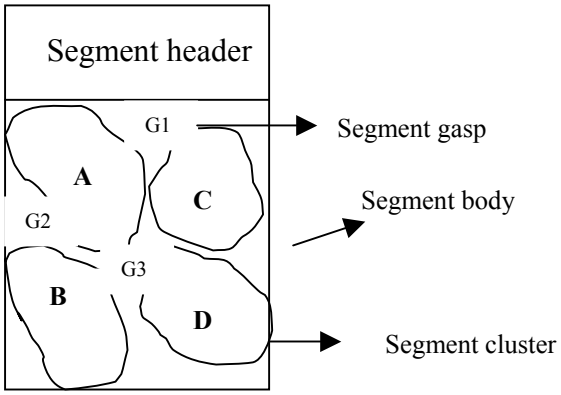


Fig. 2
A structure of segment

One central point (e.g. A, B, C, D or G1, G2, G3) is assigned to each of clusters Cl or gasps, while each cluster or gasp has an appropriate weight form semantic point of view. On the other hand, an appropriate semantic network of facts and information containing appropriate Tbe, Pet and Ret terms[3] represents each cluster or gasp. The image may be considered to be a universal class U_c and any cluster Cls or gasp Gsp closely related to this image is considered to be a member of the appropriate SFI set with an appropriate pertinence measure a whole, so that the following formulas may be postulated:

$$SFI^i\ Cls^i{}_j = L^i{}_{cl} \qquad (7)$$
$$SFI^i\ Gsp^i{}_k = L^i{}_{gsp} \qquad (8)$$

[2] SFI – **S**et of **F**acts and **I**nformation

[3] Tbe term – term to be explained, Pet term -Principal terms based on them the Tbe term is being explained, Ret term – representing relations among Pet terms. For more details see the material [4].

Where i=1, 2 ……n, j=1, 2…..m_1 and k=1,2 …..m_2 , while the following relations between m_1 and m_2 might be postulated:

$$m_1 = m_2 \text{ or } m_1 \neq m_1 \quad (8)$$

These considerations lead to application of fuzzy sets when interpreting and formalizing a semantic content of image and its structure units (segments, clusters and gasps)[4]. According to this fact, SFI sets might be considered to be fuzzy sets, based on which are defined the above-mentioned semantic functions: *partial and final* semantic *function together with their appropriate values. However, these functions have a fuzzy character as well.*

2.2 Semantic Aspects of Image Structure Units

In general, segments, clusters and gasps are considered to be principal units related to structure of image. They have an appropriate semantic content, which may be interpreted via semantic networks of facts and information and quantified via partial semantic functions and values. A text in natural language (hereinafter known as TNA) may describe a content of any image or its segments, while the following rules may be postulated:

- Any image, its segments, clusters or gasps may be described via text in natural language and any text in natural language [4]. One or more fragments may describe one image cluster, gasp from semantic point of view.

$$\forall \text{ Segment(i)}: \ \exists \ (\text{Fragm(i, 1), Fragm(i, 2)}\text{…….Fragm(i,m))} \quad (9)$$

- A semantic function is considered to be a facility for quantitative representation of semantic networks, which describe semantic content of any image cluster, gasp and segment.
- Principal terms and relating terms represent variables of semantic function and term to be explained represents the semantic function value (fsv) or fsv(i) [4]. For formal description, the following formulas may be postulated:

$$\text{fsv(i)} = \text{fsv}_i\,(\text{Pet (i ,j), Ret (i, j))} \quad (8)$$

and

$$\text{Tbe (i)} = \text{fsv(i)} \quad (9)$$

where

i=1…..n
j=1….m

[4] The problems related to Application of Fuzzy Sets Apparatus in Image Semantic Analysis are discussed in [5]

As mentioned above, any image or its segments may be described via text in natural language. They may be grouped to fragments. These fragments are treated according to principles described in [4]. On the other hand one sentence or more sentences in natural language describe an appropriate image segment. There may be two principal approaches:

- Semantic analysis of image content (description of image based appropriate semantic networks and reference database respectively.
- Completing of image based on selected sentences or fragments concerning to appropriate image segments.

2.3 A Life Cycle of Image

The term "A Life Cycle of Image is closely related processing of image, which passes through two types of phases closely related to:

- Semantic analysis of image content and creation of image database or knowledge base.
- Retrieval and presentation of image database or knowledge base for information or knowledge delivery purposes.

However, the phase called: "Semantic analysis of image content and creation of image database" contains subordinated phases defined as follows:

- Decomposition of image into segments (see also Fig.1)
- Determination of clusters and gasps within each of segments (see also Fig.2)
- Assignment of Tbe term to each cluster or gasp
- Assignment of appropriate text fragment or fragments to each Tbe term.
- Analysis of actual text fragments in order to get appropriate Pet and Ret terms for each Tbe term
- Assignment of appropriate Pet and Ret terms to each Tbe term
- Creation of appropriate semantic networks and partial semantic functions and values.
- Creation of a final semantic function and assignment its semantic value to the analyzed image.
- Retrieval and presentation of image database content

2.4 Semantic networks and Reference Databases

In general, the standard oriented graphs represent semantic network of facts and information, which describes image content. However, this semantic network consists of two subordinated semantic networks defined as follows:

- Semantic network, which represents the information capability of the actual image represented by its segments, clusters and gasps (Fig.3). Clusters, which create segments and the image as a whole, are denoted

via Tbe – terms, while each Tbe- term represents an identifier for cluster and fragment, as well. A special supplementary identifier may be applied in order to assure the resulting identifier to be unique. Clusters and segments may have adequate interlinks to a reference database.

- Semantic network, which consists of fragments, Tbe, Pet and Ret terms and enables describing the actual image, its segments, clusters and gasps via text in natural language. (Fig.4).
- Reference databases have the same structure as those, which are applied within semantic networks concerning to text in natural language.

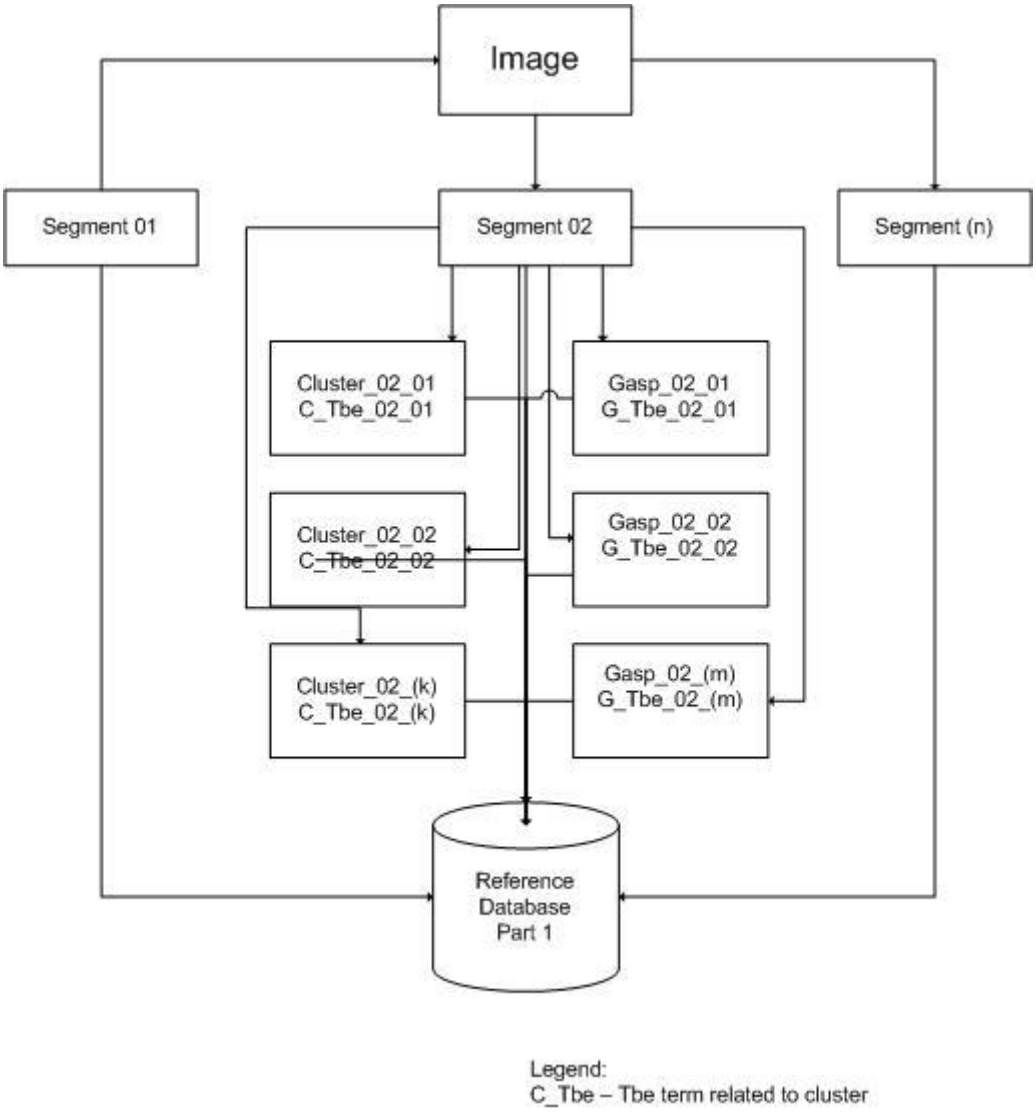
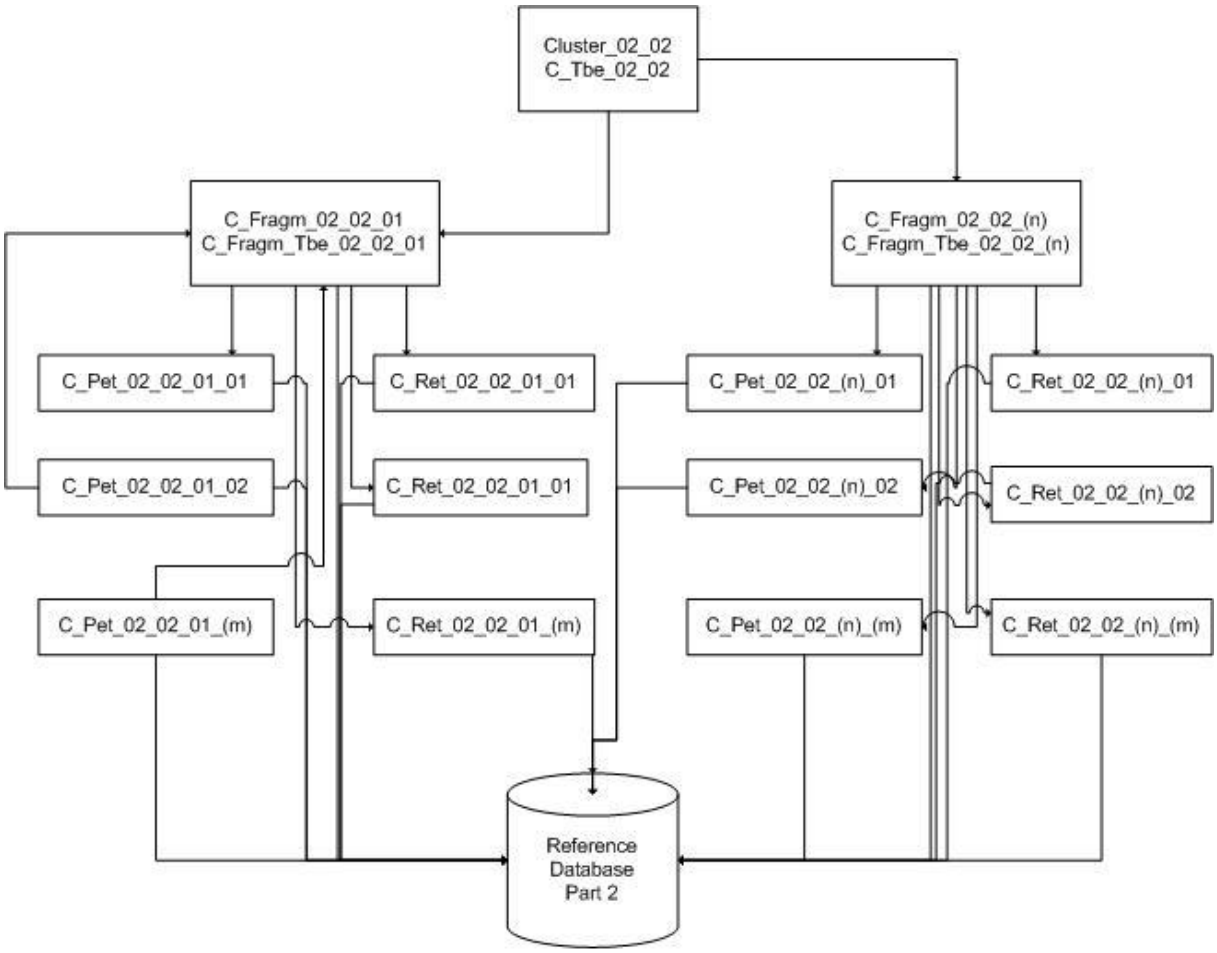


Fig.3
Semantic network concerning to information capability of the actual image which consists of clusters and segments (SNIC).



Legend:

C_Fragm –	text fragment explaining semantic content of the image cluster
C_Fragm_Tbe –	term to be explained via the text fragment C_Fragm and image cluster
C_Pet -	principal term contained within the C_Fragm
C_Ret	term representing relations among C_Pet terms

Fig.4

Semantic network, which consists of fragments, Tbe, Pet and Ret terms and enables describing the actual image, its segments, clusters and gasps via text in natural language (SNTE)

3. Systems for Semantic Analysis of Images

3.1 Design & Implementation of Semantic Networks and Reference Databases

Design, implementation and operation of semantic networks and reference databases closely related to semantic analysis of image is based on theory described in section 2. However, a creation, update, retrieval and presentation of semantic networks and reference databases are possible only with the use of the appropriate application program, which is called: *SNIC application program*[5]. The SNIC application program operates in several phases defined as follows:

- Determination of image areas, which represent segments and clusters and gasps and assignment of adequate Tbe terms to them.
- Assignment of text fragments to these Tbe terms.
- Semantic analysis of text fragments and generation appropriate relating Pet and Ret terms.
- Creation and update of semantic networks concerning to information capability of the actual image, which consists of clusters and segments.
- Creation and update of semantic networks describing the actual image, its segments, clusters and gasps, based on fragments, Tbe, Pet and Ret terms.
- Creation and update of records and items contained within reference databases together with their interlinks to image segments, clusters and gasps, as well as to fragments, Tbe, Pet and Ret terms within corresponding semantic networks.

This sequence enables creating the image database, a simplified structure of which is seen in Fig. 5.
A reference database consists of two principal parts:

- The first part (Part 1) represents references to documents and their fragments concerning to semantic networks related to image clusters, gasps, segments and the image as a whole.
- The second part (Part 2) represents references to documents and their fragments concerning to Tbe, Pet and Ret terms within the actual semantic network

There are three modules within SNIC application program:

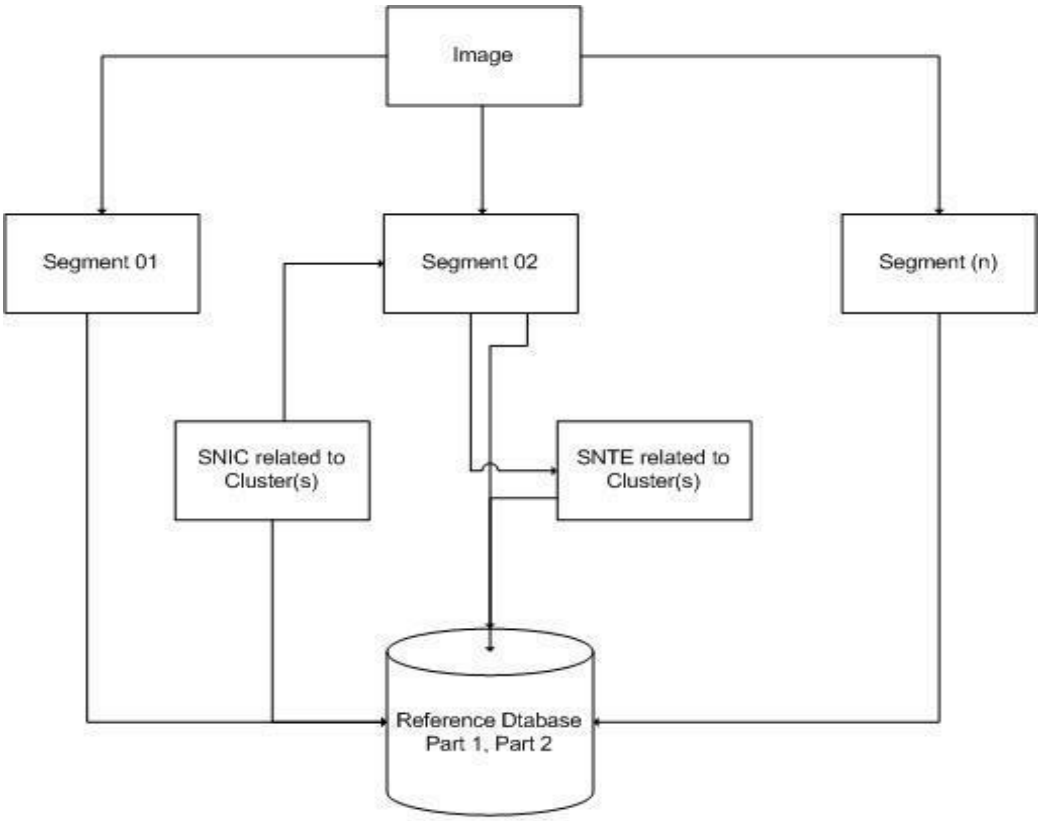
- *Extraction module* is responsible for image storage and for creation of image segments, clusters and gasps and assignment appropriate Tbe terms to them. However, this module is responsible for reference database content and creation of references to documents and their fragments concerning to semantic networks related to image clusters, gasps, segments and the image as a whole, as well. Its responsibility is extended to generation of adequate Pet and Ret terms and generation of corresponding semantic networks.

[5] SNIC – Semantic networks describing clusters and segments of image

Representation module is responsible for representation of the actual image content via appropriate semantic networks and reference databases.

Retrieval module provides information and knowledge delivery for authorized users. Its main task is to provide retrieval and presentation of semantic network and reference databases content.

A detailed description of functionality principles related extraction and representation module, in [5], while basic principles of functionality related to Retrieval module are described in the section 3.2.



Legend:

SNIC - Semantic network concerning to information capability of the actual image, which consists of clusters and segments.
SNTE - Semantic network, which consists of fragments, Tbe, Pet and Ret terms and enables describing the actual image, its segments, clusters and gasps via text in natural language

Fig. 5
Image database a simplified conceptual model

3.2 Retrieval and Presentation of Image Database Content

Two significant aspects determine the course of phase called: "Retrieval and presentation of image database or knowledge base for information or knowledge delivery purposes":

- Interpreting and explanation of image content via semantic networks and fragments
- Completing of image based on selected sentences or fragments closely related to appropriate image segments and images.

The phase called: "Interpreting and explanation of image content via semantic networks and fragments" contains subordinated phases defined as follows:

- *Entry of selection criteria,* which is being done based on a mouse cursor movement over the image surface. A procedure implemented for these purposes indicates planar or a spatial coordinate related to selected area of image and provides a presentation of corresponding Tbe terms. Each of these Tbe terms represents an actual semantic network of facts or information, which contains actual Pet and Ret terms and appropriate fragments, as well.
- *Selection and preview of actual semantic network,* while the corresponding image segment may be displayed simultaneously. Subsequently the complete image may be displayed and the corresponding segment is highlighted within this image. However, supplementary information contained in a reference database may be displayed, as well.
- Principal steps related to retrieval of image database for Image interpretation purposes based on facts and information contained within selected semantic networks might be seen via Fig. 6.

On the other hand, the phase called: " Completing of image based on selected sentences or fragments closely related to appropriate image segments and images" contains the following subordinated phases:

- *Entry of selection criteria* in form of the text string.
- Retrieval of information sets containing Tbe terms, Pet and Ret terms based on the above-mentioned text string.
- Presentation of list related to adequate Tbe terms
- Preview of corresponding semantic networks of set and information together with image segments assigned to them[6].
- Completing a resulting image.
- Principal steps related retrieval of image database for image completing purposes based on facts and information contained within selected semantic networks might be seen via Fig. 7.

[6] Corresponding semantic networks may be previewed individually or within appropriate semantic groups.

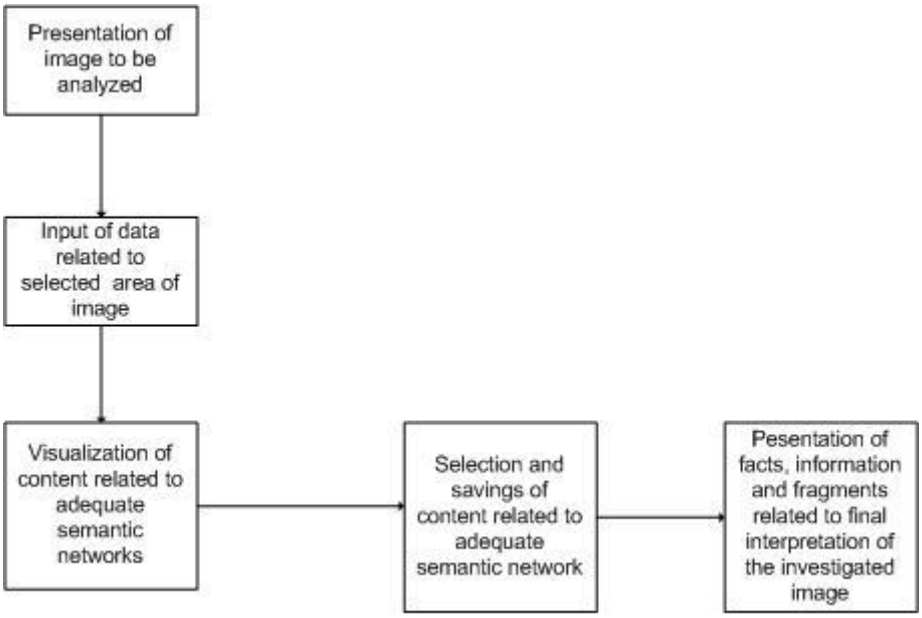


Fig.6

Retrieval of Image Database for Image interpretation purposes based on facts and information contained within selected semantic networks

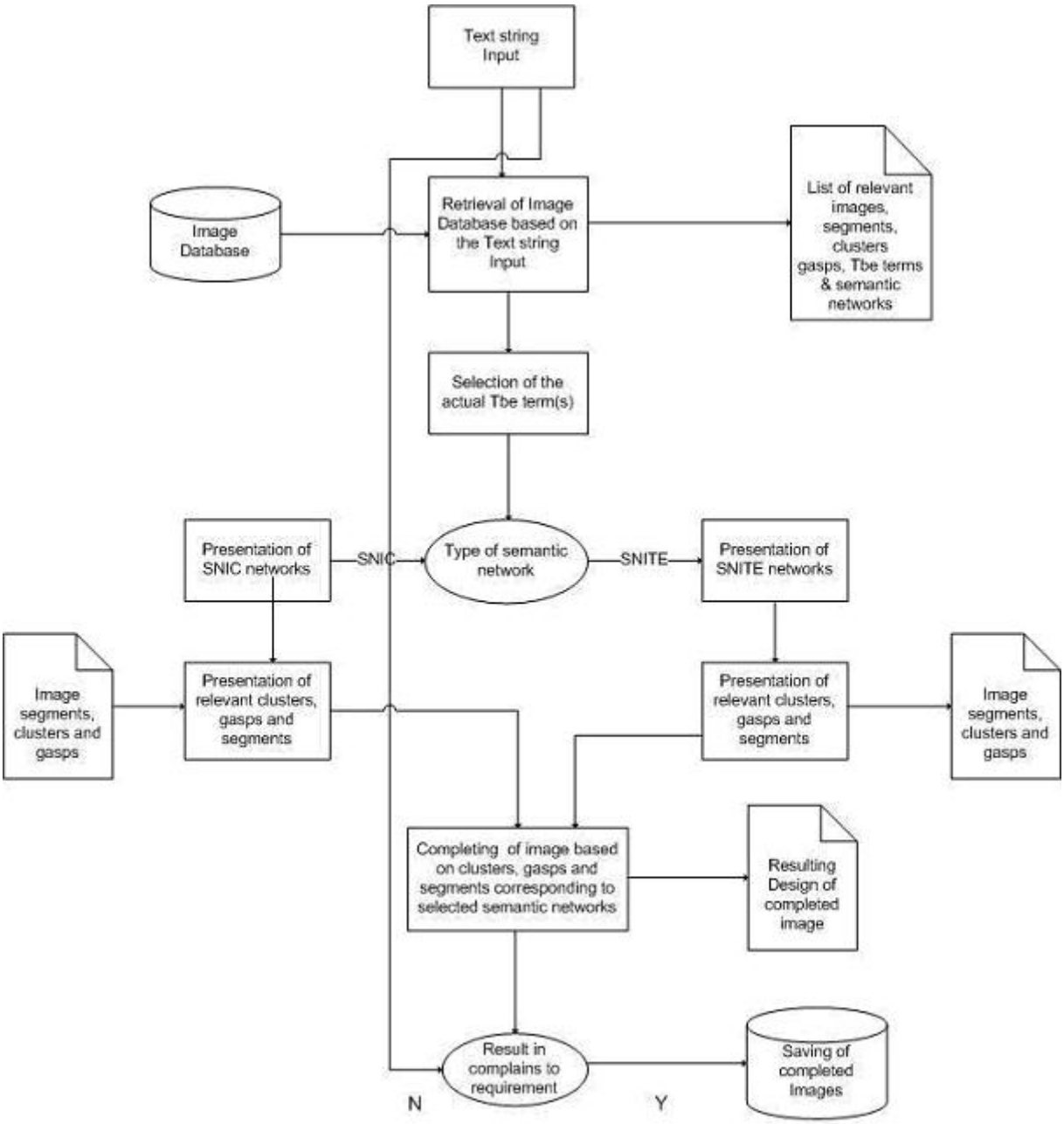


Fig.7

Retrieval of Image Database for Image completing purposes based on facts and information contained within selected semantic networks

4. Conclusion

The approach and methodology described in previous sections of this contribution concerning to semantic analysis of image creates a theoretical basis for design and implementation of an application program, which may be applied in two directions:

- Interpreting and explanation of image content via semantic networks and fragments
- Completing of image based on selected sentences or fragments closely related to appropriate image segments and images.

The described application program may be applied as part of records management and archival systems, when operating with image databases for interpretation and explanation of image content or as a stand-alone application, when completing the image based on facts or information contained in text fragments or in semantic networks of facts and information.

5. References

[1] Baatz, M.- Schäpe, A.:
Object-Oriented and Multi-Scale Image Analysis in Semantic Networks
http://www.definiens-imaging.com/documents/publications/itc1999.pdf

[2] Chang, S.F.- Sundaram, H.:
Structural and Semantic Analysis of Video
http://www.ctr.columbia.edu/~sundaram/pub/sfc-icme-ss-final.pdf

[3] Katharine E. Heintz-Knowles
Images of Youth: A Content Analysis of Adolescents In Prime-Time Entertainment Programming
http://www.frameworksinstitute.org/products/youth.pdf

[4] Stašák, J.: A Contribution to Semantic Text Analysis, printed.

[5] Stašák, J. Semantic Analysis of Text and Image
Internal Research Report, S.A.L. Banska Bystrica 2003

[6] Thong, J.V. -Blackwell, S. - Weikart, Ch. -Mandviwala, H.A.:
Multimedia Content Analysis and Indexing: Evaluation of a Distributed and Scalable Architecture
http://www.hpl.hp.com/techreports/2003/HPL-2003-182.pdf

[7] Viktor S. Wold Eide, V.S.- Eliassen, F.- Granmo, O.Ch. – Lysne, O.:
Scalable Independent Multi-level Distribution in
Multimedia Content Analysis
http://www.ifi.uio.no/~dmj/Publications/idms2002-presentation.pdf

[8] What is True High Content Analysis?
http://www.compucyte.com/util/highcontent.pdf

Author‘s address:
RNDr. Jozef Stašák, PhD.
Univerzita Mateja Bela v Banskej Bystrici
Fakulta financií – Katedra informačných systémov
Cesta k amfiteártu 1
974 01 Banská Bystrica – SR
E-mail: jozef.stasak@umb.sk