Data Bases in the Humanities and the Social Sciences
J. Raben and *G. Marks (eds.)*
North-Holland Publishing Company, 1980

SENECA: SEMANTIC NETWORKS FOR CONCEPTUAL ANALYSIS

E. García Camarero *(*,**)* , J. García Sanz (*)
M.F. Verdejo (*)
(*)Centro de Cálculo

Universidad Complutense de Madrid
(**)Facultad de Informática

Universidad Politécnica de Madrid
Spain

This paper outlines a representation based on semantic networks. It can be used in the construction of data bases, allowing the description of objects, its characteristics and relations, as well ma processes that occurring on these objects, modify the information contained in the bases. Some definitions require a—relations and semantic induction notions which are introduced. Finally, the ideas exposed in the present survey are illustrated in an example to a linguistic data base.

INTRODUCTION

The study of knowledge representation in analytic form, that is, divided in both parts and entailed relations, has been the idea that lead us in the present study.

The need of data bank arises when a particular information should be available to some users or to several applications of the same user. To represent such information seems pretty enough to mind its own structure and its independence from any application.

Initially, the storage and searching form have been considered only to define the data structure, later it has been made clear that data structuring cannot be determined by the physical structure of storage, by its forma of access or by its formal nature, but by a semantic component which is an essential part for an efficient data structuration.

In such a way, data bases become models of the external world, and this modelling is the representation of the world's knowledge on which our programs work. Thus, hierarchical bases, networks bases aud relational bases come op. All of them increases respectively, the use of the semantic component in their structure, however they continue (gradually less) tied to the ways of accessing and storing.

From other areas of research, such as language comprehension or memory study, a representation form called semantic networks, has been developed, (1, 2, 3)

Also , this methodology has been used for the construction of data bases (4). En our work we define formally a somatic graph, we include induction ideas that allow the development of minimals graphs and at the same time makes easy the idea of semantic association; the idea of semantic association; the notion of proximity simplifies searching processes. We detail a semantic graph of large application and will expose an example of implementation of these ideas.

Another element given is the idea active lot base in the sense that we include in the representation a description not only 01 static data bot procedural knowledge which allow us to consider a data base as a dynamic system.

1. SEMANTICS NEIWORKS

1.1. Definitions. We defined a semantic network as a directed graph with labels in nodes and edges. The labels of nodes means the type of knowledge that they represent and the labels of edges point out the kind of relationship connection the knowledge represented by the nodes of the edge.

Formally, a semantic network

$$G = (N , R , T_N , T_R ,\varphi_1 , \varphi_2)$$

consists of a non empty finite set N of elements called nodes, together with a set $R \subset M \times N$, of pairs of nodes which are called edges, with two sets T_N and T_R of elements called *node's labels* and *a—relations* respectively, and two functions ,φ_1 , φ_2 , such that

$$\varphi_1 : N \to T_N \text{ and } \varphi_2 : R \to T_R$$

If $\alpha \in T_R$, we put $R_\alpha = \varphi_2{}^{-1}(\alpha)$ that is, R_α is the extension of the a—relation α.

1.2. Inference rules. Given a semantic graph C, we will define some inference rules by which we can add or suppress labelled edges to G; in the first case we obtain a extended graph, and in the second we obtain a reduced graph. A graph that can not be extended is called complete, and a graph that can not be reduced is called Kernel. We put the following inference rules:

R.1 $(x_i, x_j) \in R_\alpha , (x_j, x_k) \in R_\beta \Rightarrow (x_i, x_k) \in R'_\gamma$,

R.2 $(x_i, x_j) \in R_\alpha , (x_i, x_k) \in R_\beta \Rightarrow (x_j, x_k) \in R'_\gamma$,

R.3 $(x_i, x_j) \in R_\alpha , (x_k, x_j) \in R_\beta \Rightarrow (x_i, x_k) \in R'_\gamma$,

R.1 is called *composition* rule; R.2 is called *outcomming* rule, and R.3 *incoming* rule. By $R'\gamma$ we mean an enlargement of $R\gamma$ that is, an edges set with label γ, which includes all the edges of $R\gamma$ and those obtaining by application of any of the rules above.

1.3. Neighbourhoods and boundaries. Let G be a semantic graph, we will define two functions ψ^+ , ψ^- called *affluence* and *confluence* functions respectively, as it follows: $\psi^+ = N \to P(N)$, $\psi^- = N\to P(N)$, it is

$$\psi^+ (n) = \{n_i, \mid (n, n_i,) \in R \} \text{and } \psi^- (n) = \{n_i, \mid (n_i, n) \in R \} \quad \text{where } n_i \in N.$$

We will call elementary neighbourhood of *n*, the set:

$$E(n) = \{n\} \cup \psi^+ (n) \cup \psi^- (n)$$

and boundary of *n*, the set

$$C(n) = \psi^+ (n) \cup \psi^- (n)$$

obviously $E(n) — \{n\}= C (n)$

In general, we will define neighbourhood of radius r as it follows:

$$E^r (n) = E^{r-1} (n) \cup (\cup E(n_i)) \qquad r\geq 2$$

such that $n_i \in C^{r-1} (n)$; and we will define boundary of radius r, as it follows:

$$C^r (n) = E^r (n) – E^{r-1} (n) \qquad r\geq 2$$

As well, we can define patterned affluence and confluence functions related to $\alpha \in T_R$, as following

$$\psi_\alpha^+ (n) = \{n_i \mid (n, n_i) \in R_\alpha\}$$

$$\psi_\alpha^- (n) = \{n_i \mid (n, n_i) \in R_\alpha\}$$

It is clear what we understand by patterned elementary neighbourhood related to α, and its boundary, represented by E_α (n) and C_α (n) respectively. Also is truth $E_\alpha (n) - C_\alpha (n) = \{n\}$

An element $X \in T^*_R$, that is, $X = \alpha_{i1} \alpha_{i2} \ldots \alpha_{ir}$, , with $\alpha_{ij} \in T_R$, j = 1,2, . . . r, is called a pattern. We will define patterned neighbourhood as

$$E^r_x (n) = E^{r-1}_{x'} (n) \cup (\cup E_{\alpha ir} (n_i)) \qquad r \geq 2$$

with $x' \alpha_{ir} = x$, and l(x) = r, considering $n_i \in C^{r-1}_{x'} (n)$.

Considering $C_x^r (n) = E^r_x (n) - E^{r-1}_{x'} (n)$ $r \geq 2$ and calling patterned boundary to $C_x^r (n)$ related to x.

The concept of neighbourhood applied on a kernel gives an idea about semantic proximity.

2. A SEMANTIC NETWORK SPECIFICATION

Assuming a universe of discourse, the semantic network representing the information in this area is constructed from basic elements called nodes. A label that is an element belonging T has been assigned to them, so we can classify the nodes depending on the type of knowledge it la required to express.

2.1. Nodes types. In our case, we consider a set T_N which seems general enough to us, for the universe used in data bases.

For each one of the types, two categories could be considered (abstract and particular occurrences) depending on, we either want to represent a generic concept or an instance of the same.

The set T_N is formed by the following elements:

{abstract object, instance object, abstract attribute, instance attribute, abstract relation, instance relation, abstract class, instance class, abstract action, instance action)

Graphically represented by:

2.2. A—relations types. Nodes represent a part of knowledge we have got about work object, the way to express the connection they have, between them, will be done by a basic set of binary relations composition, defined among those node classes detailed above.

We assign a label for each a—relation, the set is formed by:

$$T_R = \{ISA, HAP, CLAS, AP, EXT, ELEM, ARG, RES, ALC\}$$

2.3. Nodes types description. Object nodes: ore the nodes used to represent objects appearing in the topic being modelled. Attribute nodes: express properties. If they are abstract, a quality is represented, and if they are instances, they represent a specific value of this quality. Relation nodes: these are the nodes to express some characteristics that occurs among nodes classes. Class nodes: used to represent groups of elements. A class can be defined by enumeration of its elements, by determination of those elements verifying property or relation, or by a set of elements obtained as a result of an action. Action nodes: describe operations performed on objects. The number of actions are limited to a basic set which depends on the application domain. More complex processes will be expressed by structured composition of the primitive actions set.

2.4. A—relations types description. A—relations are specified by labels placed over directed edges between an initial and a destination node, as following: ISA relates an instance to its generic concept. CLAS express an inclusion relation of initial node meaning into destination node. HAP the destination node is one of the parts which compound initial node. AP express initial node attribute is a characteristic which apply to the destination node. EXT express that destination node (class type) is formed by n—uples which verify an n—order relation, or that it is a group of elements which verify a particular characteristic. RES express the entail between an action and its result. ELEM express the link between a class an the elements that compound it. ARG specifies classes among them a relation is defined or object on which an action acts. AG express the link between a relation and an action that supply the elements which verify it. ALC express the idea of repetition for an action set.

2.5. Inferences for the specific network. As it has been defined, inferences are operations between pairs of elements belonging R.

In our network the following ones can be done:

— Composition Rule (c) R_α c R_β = $\{(c, y) \mid (x, z) \in R_\alpha \wedge (y, z) \in R_\beta\}$, the new pair produced is included into R_γ defined for the following cases:

If $\alpha = \beta$	then	$\gamma = \alpha = \beta$	and	$R'_\gamma = R_\gamma \cup (x,y)$
If α = CLAS, β = HAP	then	γ = HAP	and	$R'_\gamma = R_\gamma \cup (x,y)$
If α =ISA, β = CLAS	then	γ = ISA	and	$R'_\gamma = R_\gamma \cup (x,y)$

— Incoming rules (A) R_α A R_β = $\{(x, y) \mid (x, z) \in R_\alpha \wedge (y, z) \in R_\beta\}$, the new pair produced is included into R'_γ defined the following cases:

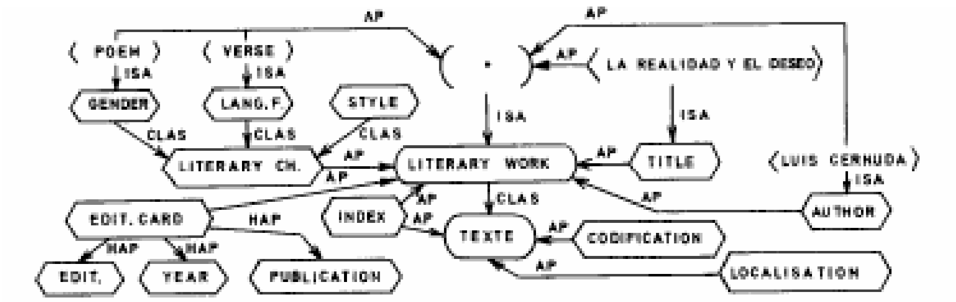
α=AP, β = ISA implies γ = AP and $R'_{AP} = R_{AP} \cup (x,y)$
α=EXT, β = ELEM implies γ = AP and $R'_{AP} = R_{AP} \cup (x,y)$
R_α A R_β = $\{(y,x) \mid (x,z) \in R_\alpha \wedge (y,z) \in R_\beta\}$

The new pair is included into R'_γ defined for the case
α = CLAS β = AP implies γ = AP and $R'_{AP} = R_{AP} \cup (y,x)$
Inferences are an expansion mechanism used in searching processes. The inferred new pairs can be added to the base graph producing an extended graph.

3. A CASE STUDY

An implementation case of this representation has bean builded—up for the linguistic computational area. From several literary works, divers statistic information can be obtained, an words frequency, grammar categories occurrences and semantic elements, words context, sentences context, concordances.

lo get this results, usually some specific program are constructed for each type of product required. With our approach, a representation of the knowledge relevant to the subject is done by means of a semantic network, where the object are the literary work. At a conceptual level, we have the following object: literary work text, capital letter, divider, accent, letter. Attributes applied to this elements are: for literary work: literary characteristic, editorial card, title, author and index; for text: codification, localisation, and structure; fur capital letters, divider and accent: length la characters; for letter: length in characters and order number. As conceptual object classes we have got: capital symbols, divider symbols, accent symbols and letters belonging to Latin alphabet. Among the relations can be mentioned: comparation, frequency an context. Soma attributes are formed by others attributes, for example editorial card is itself divided into editorial name, year and publication place. A partial diagram is show in figure 1 where various a—relations occurring between elements mentioned above can be observed.



The integration of an instance element into the network, for example “La realidad y el deseo” by Luis Cernuda, Spanish poet belonging the group “Generación del 27”, can be observed by the instances nodes which would describe the object. this initial network can be increased with all the knowledge needed by the inclusion of other literary works or the result produced by the performance of an action. From a network node, using the mechanism describes above: inferences, neighbourhood and contours, several informations can be derived from the graph. the same representation formalism is used to express the meaning of sentences in natural language, that allow as to resolve questions by graphs comparation processes. (5, 6, 7).

REFERENCES

[11 Schank, Conceptual Information Processing (North—Holland, 1975)
[2] Myloupulos, TORUS, Proceeding 4IJCAI (1975) 414—421
[3] Simmons, Semantic Networks, in: Computers Models of thought and language (Freeman, 1973)
[4] E. García Camarero, J. García Sanz, M. F. Verdejo, Red semántica para la gestión de una base da datos lingüísticos, CIL 79,467—478
[5] E. García Camarero,, J. García Sanz, M. F. Verdejo, Una basa activa da datos lingüísticos. Memoria Fundesco (1978)
[6] E. García Camarero,, J. García Sanz, M. F. Verdejo, Un método de respuesta a preguntas formuladas a un grafo semántico, AEIA 79.
[7] E. García Camarero, J. García Sanz, M. F. Verdejo M. F. Verdejo, Representación del conocimiento mediante redes semánticas, CII 79, 457—466
[8] E. García Camarero, J. García Sanz, M. F. Verdejo, Resolución de sistemas de ecuaciones booleanas con variables en el conjunto de los contornos de un grafo. AEIA 79.