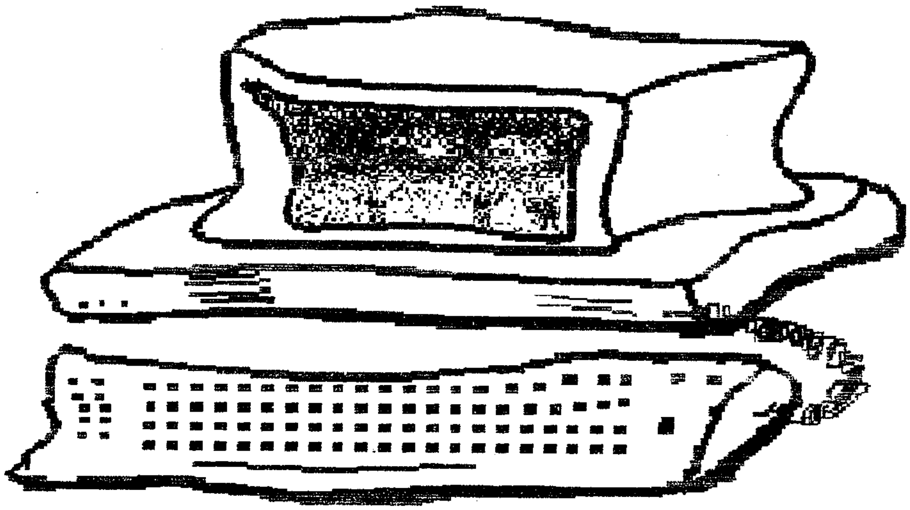


Fuzzyfication of Conversation Theory
First approach

Luis Jorge Mateus Rocha
Instituto Superior Tecnico
Lisbon

This Paper is supported by the NATO special Programme Pannel on advanced educational technology. Nonetheless, the responsability for the views and opinios contained herein pertains, entirely to the author.

@CybeR New Power Generation Software

"The Fuzzyfication of Conversation Theory - Introduction"

ABSTRACT:

The work here presented, intends to contribute to the extension of **Gordon Pask's Conversation Theory**, by introducing the following subjects:

- The concept of resemblance;
- Fuzzyfication of Conversation Theory;
- Creation of a new operation - **S**preading **S**elective **P**rune;
- Control the spreading of the Conversation Domain;
- Moving from a non-hierarchical structure to a hierarchical structure.

To achieve the objectives presented above, some aspects of the work presented by **Kiyohiko Nakamura** and **Sosuke Iwai** on the paper **"A representation of analogical inference by Fuzzy Sets and its application to information retrieval systems"** were used and extended.

The first part of this work, briefly presents a knowledge structure (Conceptual network) constructed on **resemblance** between concepts; also presenting the operations defined over this structure, "inference by resemblance Fuzzy functions", and an algorithm that uses this structure and its operations to construct an interactive associative database.

The second part, develops the first part's knowledge structure, to a structure capable of representing and extending Conversation Theory's mesh of clusters structure. The four remaining subjects described above are then discussed.

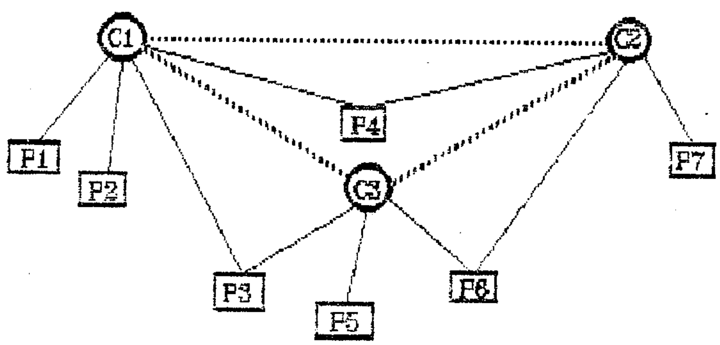
The third part presents two computer programs that implement the first and second part knowledge structures, **CyBase** and **CyBer L**.

PART I - INTRODUCTION

On this first part, the construction of an interactive associative database, is presented. The work developed on this first part uses as a basis the paper, referred above, presented by Nakamura and Iwai, developing it according to the author objectives.

I-1 Knowledge structure as a Conceptual network constructed according to RESEMBLANCE between concepts.

There are two different kinds of nodes and links, on the structure presented on this part: **Concepts (c_i)** and **Properties (p_k)**. The links Concept-Property (c_i-p_k), represent the attribution of property p_k to concept c_i, and the link concept-concept represents de (Un)**Resemblance** between concepts.



Resemblance between concepts is defined as:

$$r(x_i, x_j) = N(x_i \cap x_j)/N(x_i \cup x_j) =$$

$$= N(x_i \cap x_j)/(N(x_i) + N(x_j) - N(x_i \cap x_j))$$

where:

- $N(x_i)$ represents the number of properties that qualify concept x_i, in a direct manner;
- $N(x_i \cap x_j)$ represents the number of properties that qualify both x_i and x_j;
- $N(x_i \cup x_j)$ represents the number of properties that qualify x_i or x_j.

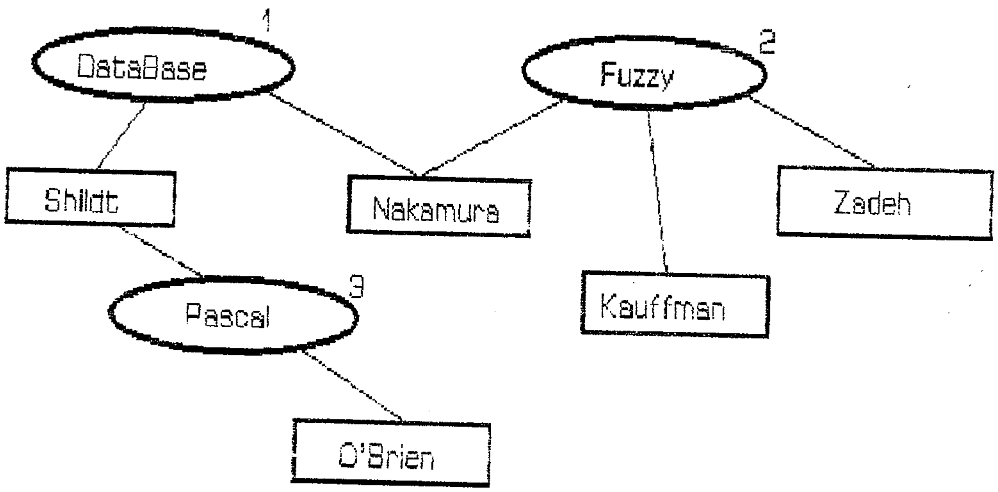
The diresemblance or a measure of the distance between concepts is assumed to be the inverse of the resemblance between them:

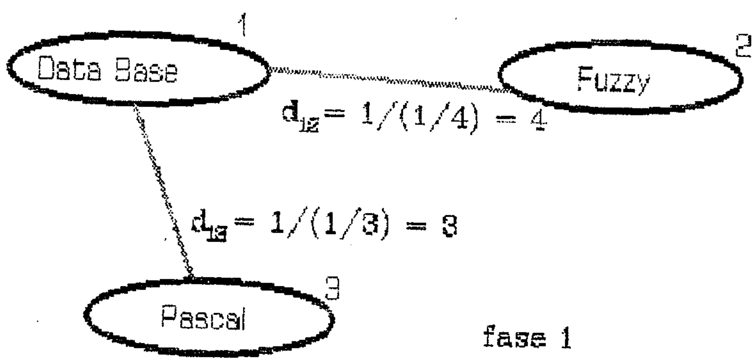
$$l(x_i, x_j) = 1/s(x_i, x_j)$$

On this network, this distance is only calculated for those concepts that have properties in common, since that for concepts that have no common properties the resemblance is nil and the distance infinite. After these (direct) distances are calculated, the distances between concepts without common properties, are calculated using an algorithm for searching the shortest route from one concept to another in a conex graph.

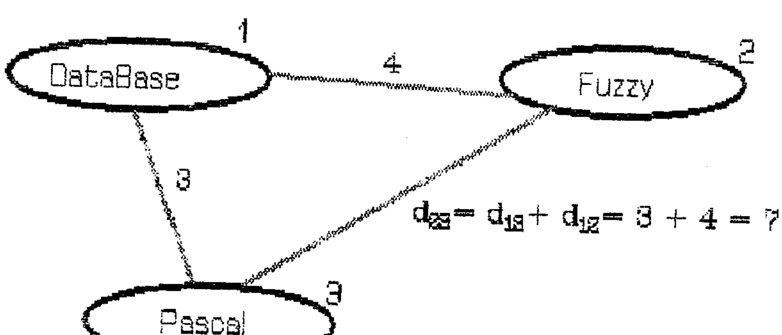
After the distances between all concepts are calculated, a new graph is obtained. This graph has only one kind of nodes and links,. The nodes are the concepts and the links are the distances between them. This graph is a topological space called **Knowledge Space**.

The construction of a small knowledge space, is presented next. Considering 3 concepts: Data-Bases, Fuzzy Sets and Computer language Pascal; and 5 properties which are names of authors of papers and books on the subjects presented as concepts.





fase 1

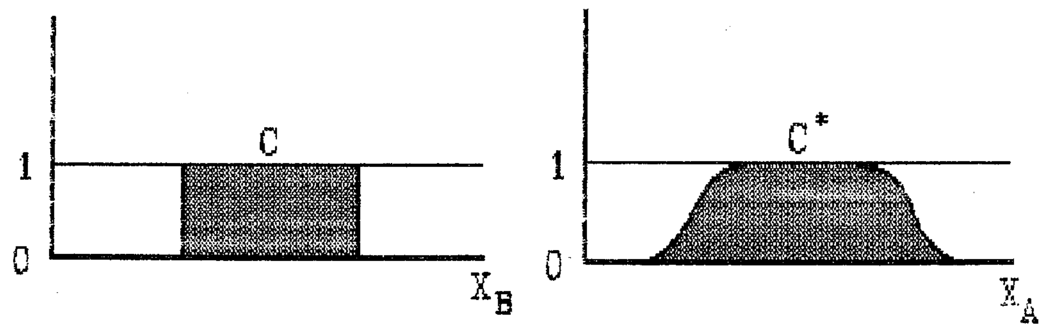


fase 2

I-2 - Transmission of User's interest to the DataBase's Knowledge Space.

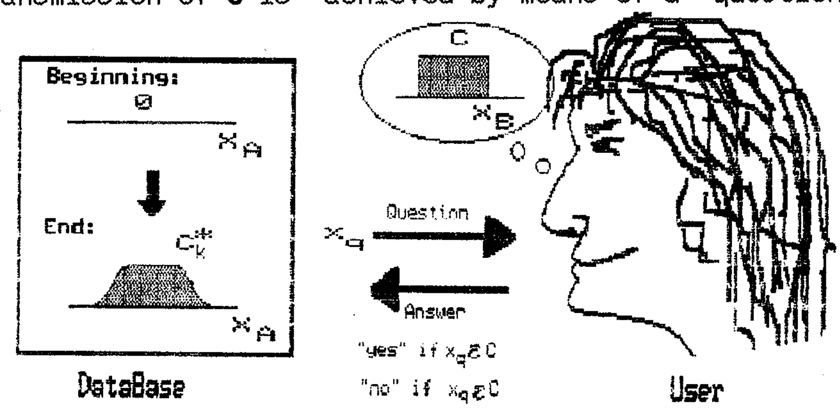
Considering then, the Database's knowledge space **X_A**, the prime objective is the transmission of a subset of the user's own knowledge space **X_B**, on which the latter is interested, the **Subset of Interest** **C**, to **X_A**. The transmission of **C** from **X_B** to **X_A**, results on a fuzzy subset **C^*** of **X_A**, called the **Learned Subset**.

The subset **C** of **X_B** is an ordinary subset, since the user knows exactly which concepts interest him.

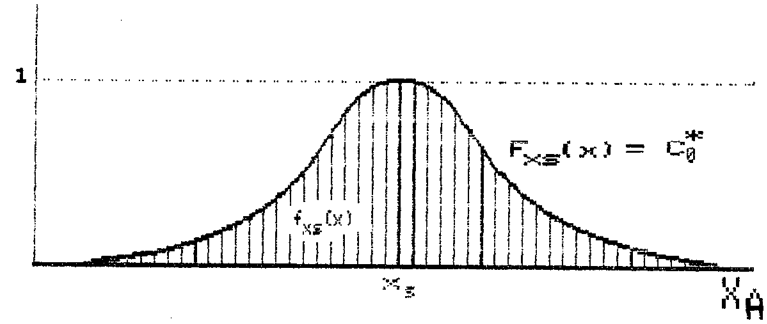


On the other hand, the Learned Subset **C^*** of **X_A**, is a fuzzy subset, because it is the way the Database "apprehends" the subset of interest **C**. This fuzzy subset should be as close as possible to an ordinary subset, and this "closeness" is an indication of the quality of the process of transmission.

The transmission of **C** is achieved by means of a question-answering system:

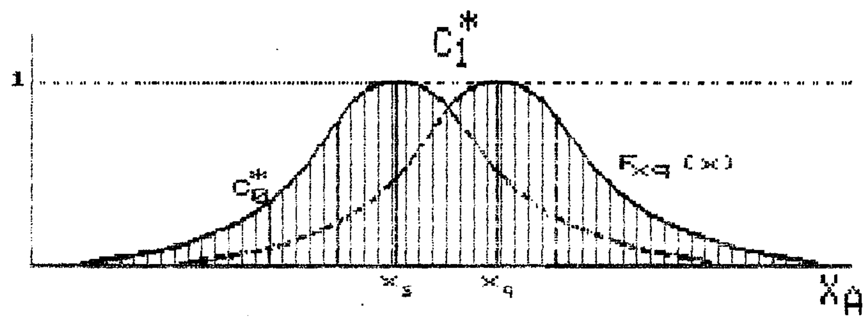


First the user is asked to introduce a concept, **x_s**, from his subset of interest. At this point, the system accepts that this concept is contained on **C** , creating the first fuzzy subset of **X_A** representing the Learned Subset **C^*_0**. To create this fuzzy subset a fuzzyfication function, **$f_{xs}(x)$**, is generated. This function, spreads the information $x_s \in C$, to the other concepts of **X_A**, according to their distance or (di)resemblance to **x_s**, that is, the degree of membership, of the **X_A** concepts to **C^*** is inferred by their resemblance to **x_s**.

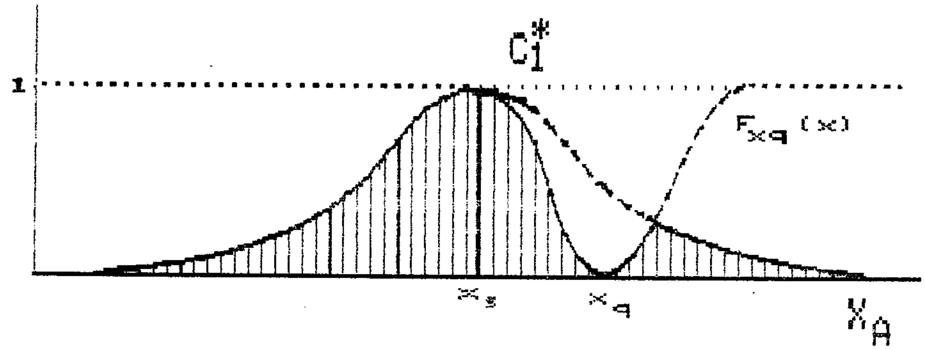


Next, the system chooses a concept, x_q, from its own knowledge space X_A, according to a selection criterion outlined later on. The user is then asked, whether or not this concept is contained in **C**. If the answer is "Yes", the system recognises that x_q is contained on **C**, creating the fuzzyfication function described above, this time centred on x_q. The new Learned subset, is now given by the fuzzy union of the previous Learned subset with the fuzzy subset defined by the fuzzyfication function:

$$C^*_1 = C^*_0 \cup^* F_{xq}$$



If the answer is "No", the system recognises that X_q is not contained in **C**. A different fuzzyfication is generated this time, $\underline{f}_{xq}(x)$, since the information that must now be spread, is that concept x_q is <u>not</u> contained in **C**:



The new learned subset is now the fuzzy intersection of the previous learned subset with the fuzzyfication function presented above:

$$C^*_1 = C^*_0 \cap^* \underline{F}_{xq}$$

This process is then repeated, until the stopping criterion is accomplished. After k steps, the learned subset, C^*_k is given by:

$$C^*_k = (\cup^*_{xy} F_{xy}) \cap^* (\cap^*_{xn} F_{xn})$$

where x_y represents the first concept and all those to which the user replied "Yes", and x_n all the "No" answered concepts.

I-3 Selection and Stopping Criterion.

The selection criterion, chooses the concept to be presented to the user, in such a way as to reduce the fuzzyness of the learned subset C^*_k.

The fuzzyness of C^*_k can be measured by the Kauffman Index of Fuzzyness:

$$I(C^*_k) = 2/(X^\#_A)^{1/2} * D(C^*_k, C^{*\wedge}_k)$$

where $X^\#_A$ is the total number of concepts in X_A and $D(C^*_k, C^{*\wedge}_k)$ is the Euclidean distance between the fuzzy subset C^*_k and the ordinary subset $C^{*\wedge}_k$ which makes this distance minimum.

Another index, the **Expectation of $I(C^*_k)$**, is created to predict the fuzzyness of C^*_{k+1} when the concept presented to the user is the concept x_q:

$$E(I(C^*_{k+1}))_q = 0.5 * (I(C^*_k \cup^* F_{xq}) + I(C^*_k \cap^* \underline{F}_{xq}))$$

this index considers as having equal probability both a positive or a negative reply.

This way, the selected concept is the one that presents the lowest $E(I(C^*_{k+1}))$.

The question-answering process stops when the fuzzyness index is reduced to a previously stated level, in this case, his level is established, as half of the maximum value that the fuzzyness index took on previous steps of the process:

$$I(C^*_{kq}) < 0.5 * \max(I(C^*_k)) \qquad k = 0,1,\ldots,k_q$$

$I(C^*_k)$ converges to zero, if the stopping condition is not activated, all the concepts in the knowledge space will be presented to the user, allowing the value of the membership function of C^*_k to be 0 or 1, that is, it will be an ordinary subset, it can be said, in this case, that the subset of interest **C** was totally received or learned.

I-4 Adaptive aspect of information spreading.

It was said that, when information is received regarding one concept (x_i is/is not contained in C), it is spread to other concepts according to their resemblance to this initial concept - inference by resemblance.

This spreading of information, is controlled by the shape of the Fuzzyfication functions presented on section I-2. But, this shape is not constant, in fact it is adaptively changed during the transmission of C to C*. Their shape is defined as:

$$f_{x_i}(x) = \exp(-\alpha|x-x_i|^2)$$

and,

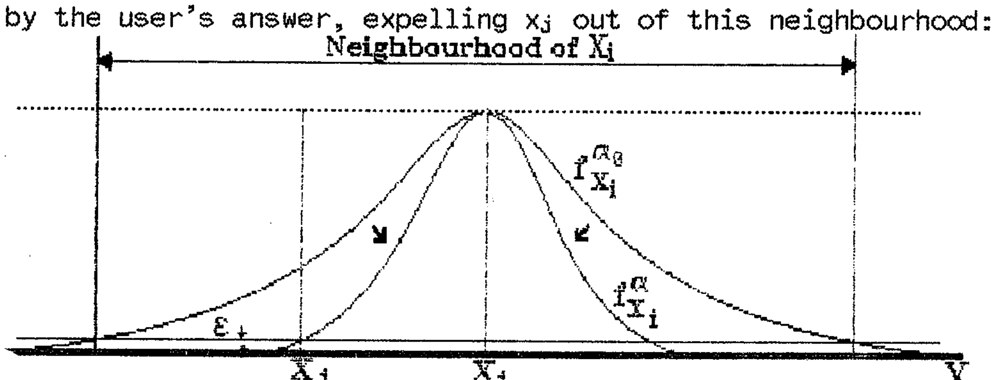
$$\underline{f}_{x_i}(x) = 1 - f_{x_i}(x).$$

The positive parameter α, controls therefore the spreading of these functions and consequently the spreading of information. When α is small, the functions spread more, and the transmission of of C is rougher, when α is large, the functions are thinner making the transmission more careful.

The initial value of α is α_0, which is established considering the number of concepts and properties on the database, and an average inference.

The **neighbourhood** of a concept x_i, is the region of the knowledge space for which the value of $f_{x_i}(x)$, with α_0, is above a small prescribed value ϵ.

Consider x_i the concept of question k, and x_j a concept used on a previous question. If x_j exists in the neighbourhood of x_i on the knowledge space X_A, and the user's answer to x_i is opposed to the one given to x_j, the, boundary of C* must be between these two concepts. This way, the spread of information should be shortened, allowing a more careful recognition of C*, and assuring that the answer given to x_i will not affect the answer already given to x_j. To achieve this, the value of α is increased to make $f_{x_i}(x_j)$ less than ϵ, shortening the neighbourhood of x_i affected by the user's answer, expelling x_j out of this neighbourhood:

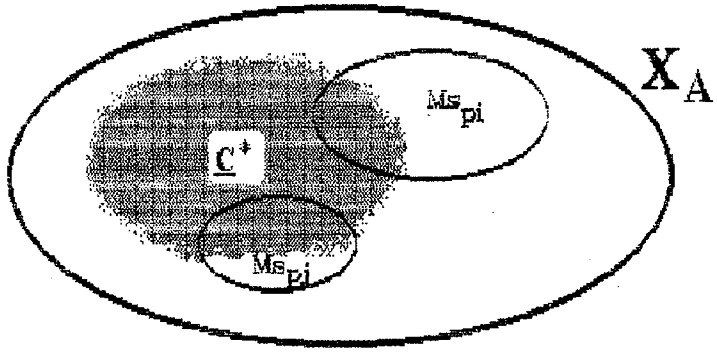


After this, α returns to its initial value α_0.

I-5 - Retrieval

After the question-answering process of transmission of the subset of interest, the system must return the properties that are, in some way, relevant to the subset of concepts of X_A, C^*.

In general, each property qualifies several concepts, defining an ordinary subset of X_A, $\mathbf{M_{pi}}$. So, a property $\mathbf{p_i}$, is relevant if $\mathbf{M_{pi}}$ largely identifies with C^*:

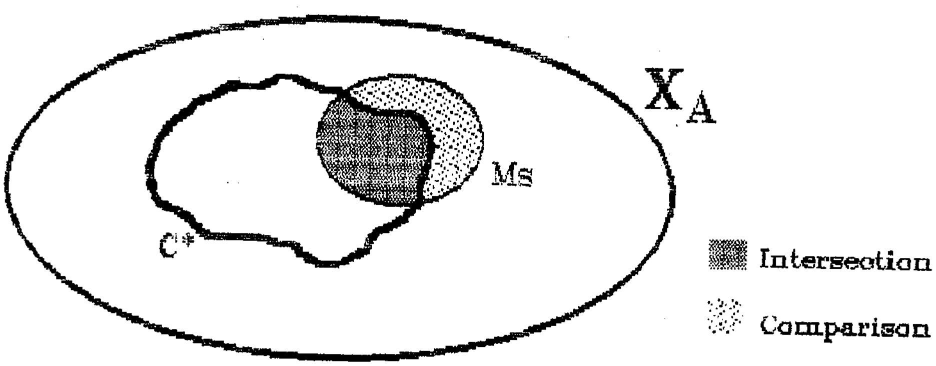


Nonetheless, there can be several ways of measuring this identification. The index $\mathbf{R_{s1}}$, is a ratio of the intersection between M_{pi} and C^*, and the subset M_{pi}:

$$\mathbf{R_{s1} = (\Sigma_{x_i \in X_A}\ \mu_{M_{pi} \cap * C*}(x_i)) \ /\ (\Sigma_{x_i \in X_A}\ \mu_{M_{pi}}(x_i))}$$

where $\mu_{M_{pi}}$ is the membership function of the ordinary subset M_{pi}, which can only assume the values 0 and 1.

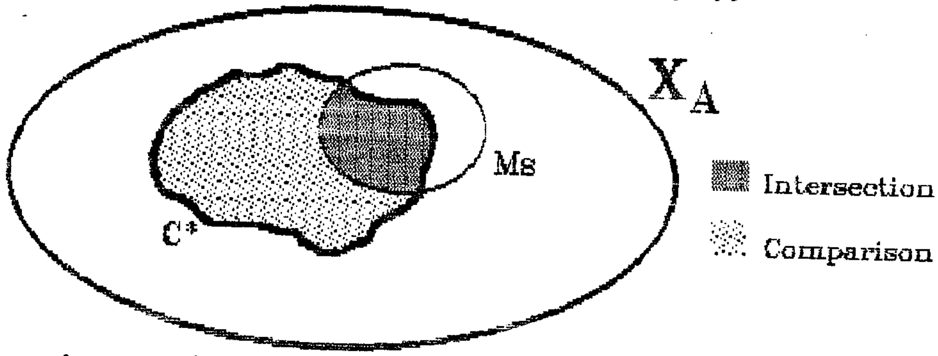
This index can sometimes give strange responses, since if the concepts of M_{pi} are also, to a large extent, contained in C^*, the value of R_{s1} will be large, even if C^* contains much more concepts than those of M_{pi}. This way, this index, is better used when we want the returned properties not only to qualify the learned subset as a whole, but also, and specially, its subsets. It is therefore, an index that emphasises the independence of the concepts of C^*.



When it is more suitable to emphasise the dependence of the concepts contained in C^*, this is, to consider that the learned subset represents a concept, qualified by the contained concepts, then the returned properties, should be those that qualify this aggregate of concepts as a whole.

To achieve this, another index is created. This index, **R_{e2}**, compares the intersection not with the property set of concepts M_{pi}, but with C*:

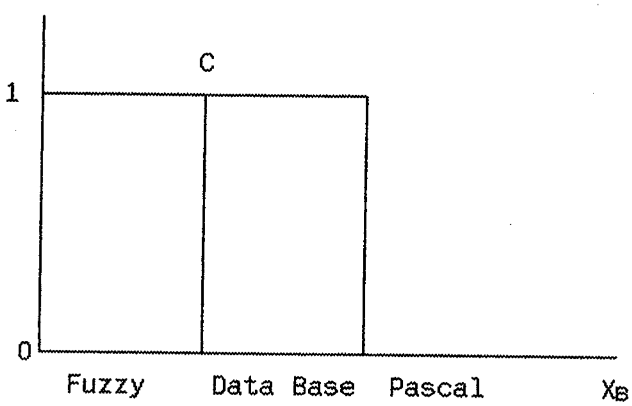
$$R_{e2} = (\Sigma_{x_i \in X_A} \mu_{M_{pi}} \cap * c*(x_i)) / (\Sigma_{x_i \in X_A} C^*(x_i))$$



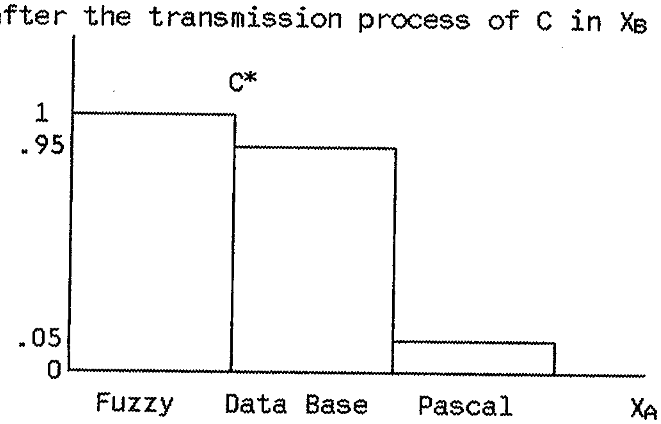
However, in practice, on the majority of the cases, the user is looking for properties that qualify the learned subset as a whole and also those that qualify its subsets. What we are looking for here, is a compromise between the two indexes presented, this way, indexes that weight its desirable characteristic among the two cases, can be formed. The most obvious, is the one that averages the two previous indexes:

$$R_{e3} = (R_{e1} + R_{e2})/2$$

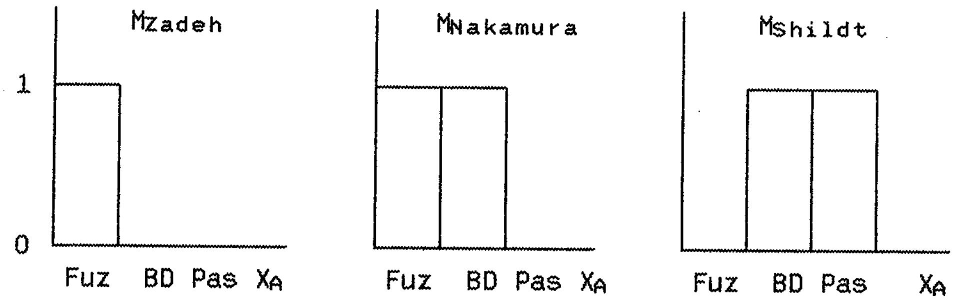
Consider the database presented on I-1, if the subset of interest is **"Data Base"** and **"Fuzzy"** :



After the transmission process of C in X_B to C* in X_A :



The subsets of concepts qualified by the properties **"Zadeh"**, **"Nakamura"** e **"Shildt"**, are now presented :



$R_{s1} = 1/1 = 1$

$R_{s2} = 1/(1+.95+.05) = .5$

$R_{s3} = (1+.5)/2 = .75$

$R_{s1} = (1 + .95)/2 = .975 = .98$

$R_{s2} = (1+.95)/(1+.95+.05) = .975 = .98$

$R_{s3} = .98$

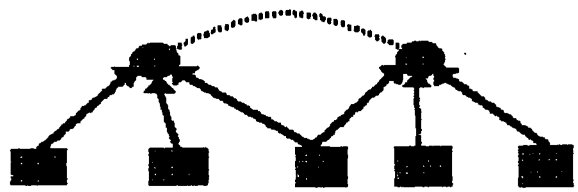
$R_{s1} = (.95+.05)/2 = .5$

$R_{s2} = (.95+.05)/(1+.95+.05) = .5$

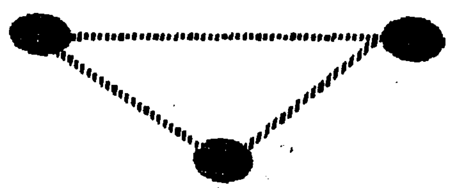
$R_{s3} = .5$

PART II - INTRODUCTION.

The basic structure presented on the first part, is a hierarchical structure, where there are two different levels, the properties, on a lower level, and the concepts, on a higher level:



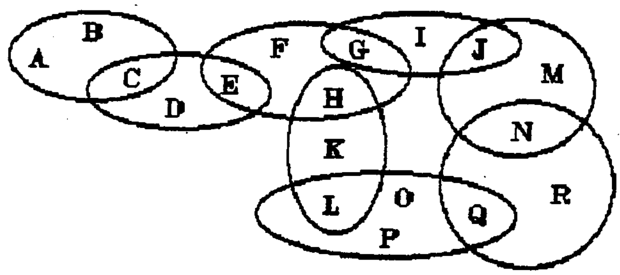
The knowledge space, is of course, a non-hierarchical structure created with information from the lower level properties, but it is only a part of the structure, the higher level part.



This second part, develops the first part's knowledge structure, to an associative non-hierarchical structure capable of representing and extending Conversation Theory's mesh of clusters structure.

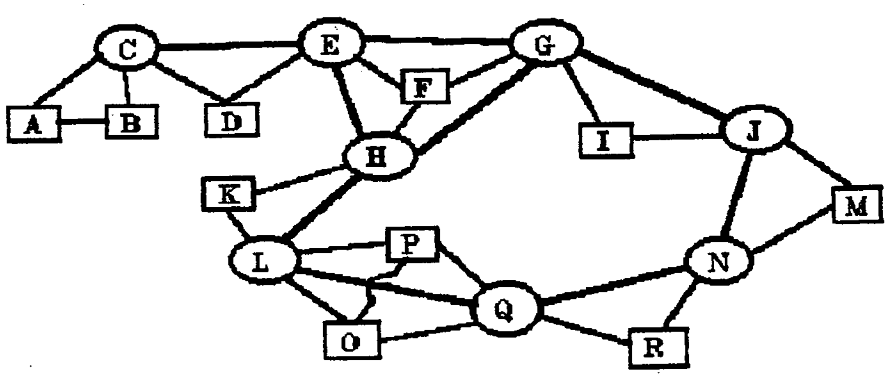
II-1 Structure development.

Gordon Pask's entailment mesh structure is formed of clusters of concepts. One Cluster, represents the existance of a relationship between the concepts within it. A generic mesh of clusters can be seen on the next figure:

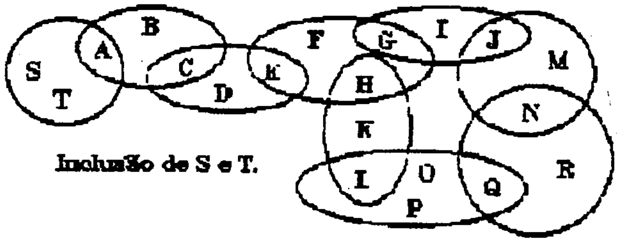


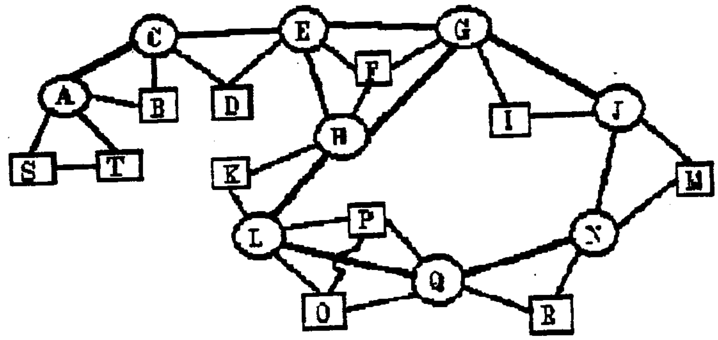
In this structure, there is only one entity, the concepts, which are all on the same level, that is, associated in a non-hierarchical manner; there is no division between concepts and properties, since concepts are qualified by other concepts. But, there are concepts which have the particularity of being included on more than one cluster, the **liaison concepts** .

A different representation of the previous structure, can be established. This representation is a graph, where the liaison concepts are shown as a circle, and the other concepts called **terminal concepts** as a square:



The resulting structure, is similar to the structure presented on the first part, but with the important difference of, even though two types of nodes are present, it is still a non-hierarchical structure, because regardless of its status, any concept can qualify or be qualified by another concept. Also, the status of a concept is not constant. When the structure grows, terminal concepts may become liaison concepts. Consider the case of the inclusion of a new cluster with concepts A,S and T; A will become a liaison concept.

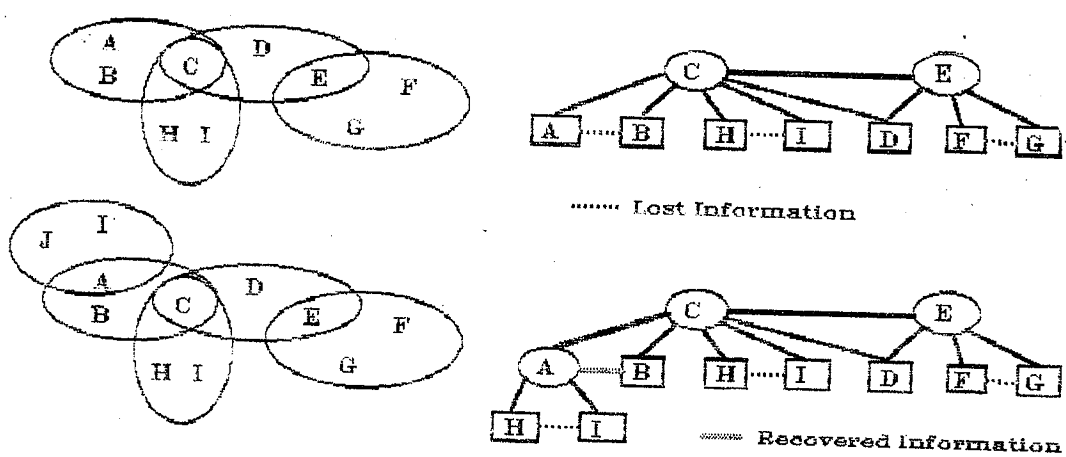




In order to bring this structure closer to that of the first part, the connections between terminal concepts are now scraped. Only connections involving at least one liaison concept remain. This is an important simplification, because the resulting structure is no longer totally equivalent to the originating entailment mesh. On this new structure, connections between terminal concepts don't really exist, nonetheless, these connections are implied by their connection to the same liaison concept. it can be said that the liaison concepts, represent the clusters, and the distance between liaison concepts is the distance between clusters. The terminal concepts connected to the same liaison concept(s) have connections between themselves implied.

The information lost with this simplification, can nonetheless be recovered when the structure grows, because when a terminal concept changes to a liaison concept its connection will be clarified. This way, it can be said, that the information loss is only a momentary phenomena, and that with time as the structure grows, this information tends to be recovered.

This process can be seen on the next figure:

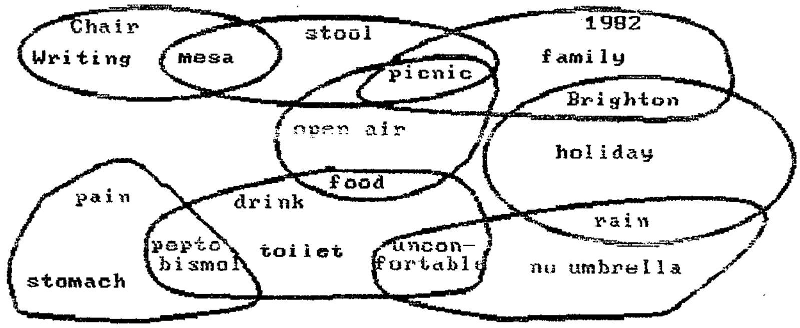


On this new structure - which, as it was shown, is different from that of the first part because it is non-hierarchical since there are no definitive differences between its objects - the process show on the first part, with some modifications, can be used revealing some interesting results. The distances will be established between the liaison concepts (representing distances between clusters), but the important difference is that, the resemblance is not only based on the connection to terminal concepts but also to other liaison concepts.

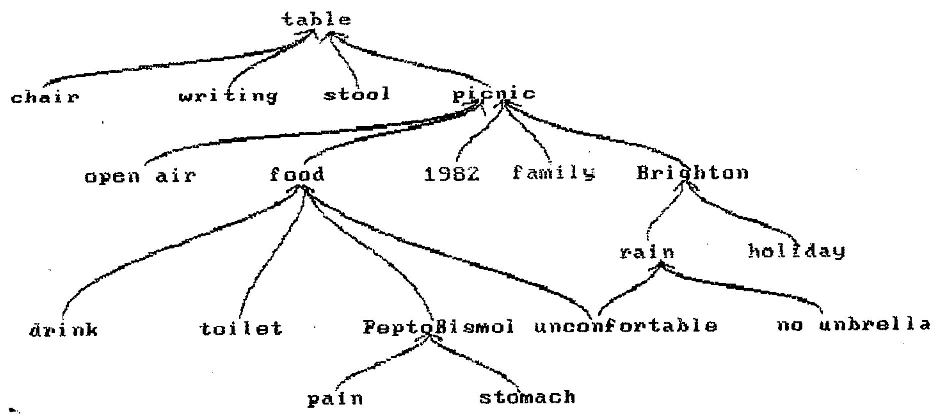
The results of this process on the structure now introduced lead to the fuzzyfication of some aspects of Conversation theory.

II-2 Control of the spreading of the conversation domain - Spreading Selective Prune.

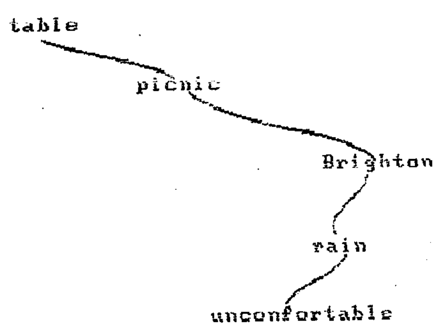
Consider the following entailment mesh:



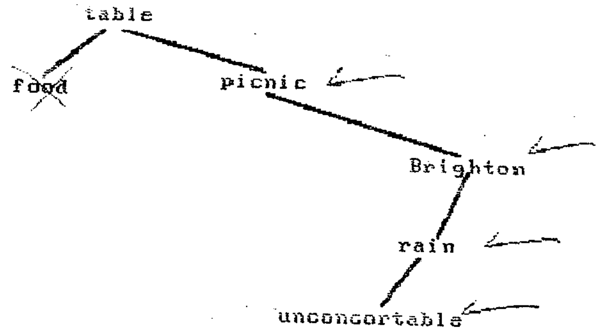
The result of the operation **Prune** with the initial concept **table**, that is, all the associative chains departing from table, can be seen now,



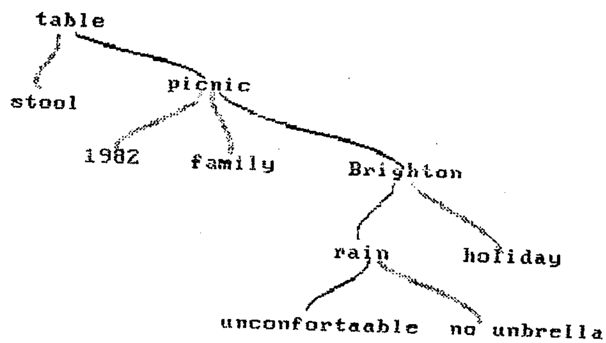
A possible perspective of the concept table, is given by the following associative chain, which is the result of a **Selective Prune**, is:



By choosing the appropriate value α(see first part), this new system, will follow the associative chains, just like a selective prune would. The value of α for which this is true is called the **Set Point α.** The following figure, represents the behaviour, of the system, the crossed concept (food), is a concept presented to the user, which does not want to follow that chain or perspective of table:



After this, choosing the appropriate value of R_{s3}(see first part) will return the terminal concepts closer to the associative chain. It can be said that the associative chain resulting from this process is an enlarged Selective Prune:



Because this chain varies with the values of α and R_{s3}, this process can be called the **S**preading **S**elective **P**rune.

The variation of α, poses another question, for smaller values of α, the associative chain will be followed on a rougher manner, it would for example follow from table to Brighton to Uncomfortable. That is, it would give larger steps. It can be said that, it aggregates clusters, close clusters melt into a larger cluster. In the example presented, table and picnic would be too close for the system to differentiate, just like Brighton and rain.

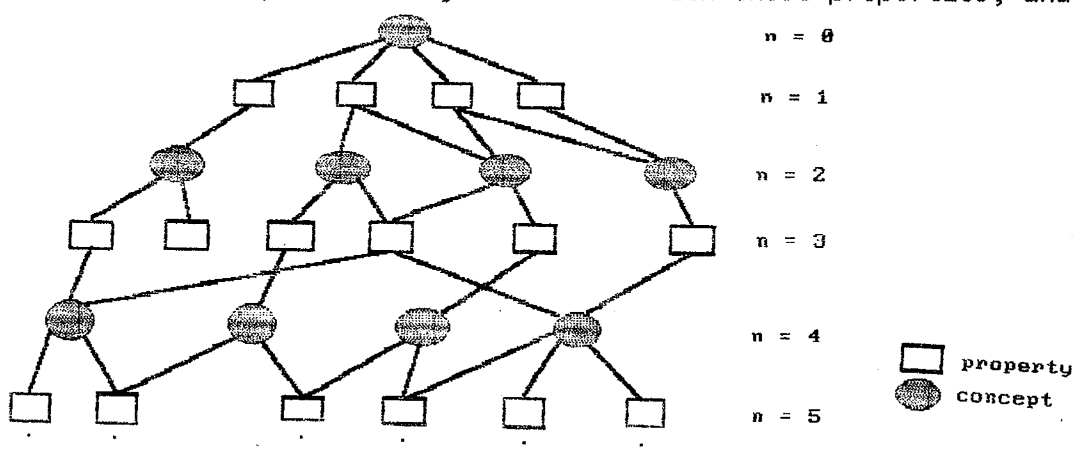
Then, the control of this parameter α, controls the spreading of the conversational domain.

II-3 Moving from hierarchical to non-hierarchical structures.

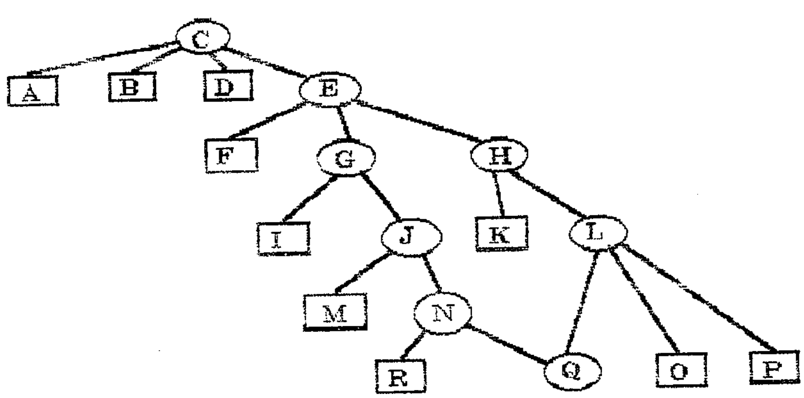
A system designed to simulate mental processes must be capable of simulating both associative and deductive processes. The first ones are implemented on non-hierarchical structures, and the second ones are best implemented on hierarchical structures. Since, the two processes, must exist and even Co-exist, a way of moving from one structure to another must be found.

It is possible to define hierarchical structures from the associative structures presented on the first and second part of this work.

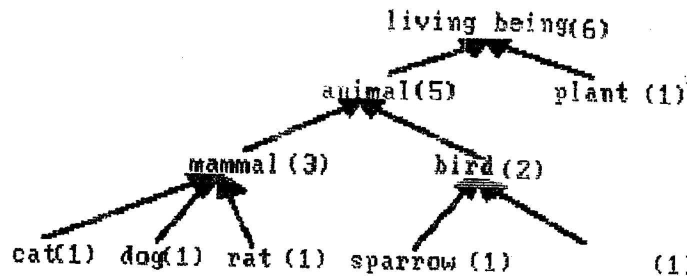
The first part structure, can be organised hierarchically on the following manner. The first level is composed of a concept, the second level is composed of the properties directly connected to this concept, then come the concepts directly connected with these properties, and so on:



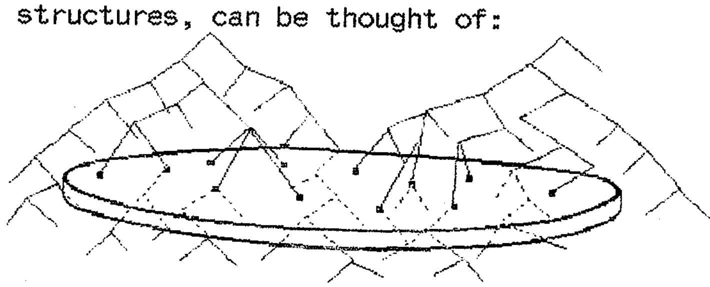
Something similar can be done to the structure on the second part, departing from one concept, the second level is formed by the concepts directly connected and so on. This is in effect the operation prune of the departing associative structure:



But, these structures are hierarchically organised, by distance or resemblance between concepts, not by concept generality, as it is desirable to simulate deductive processes. An example of a hierarchical structure can be seen on the following figure:

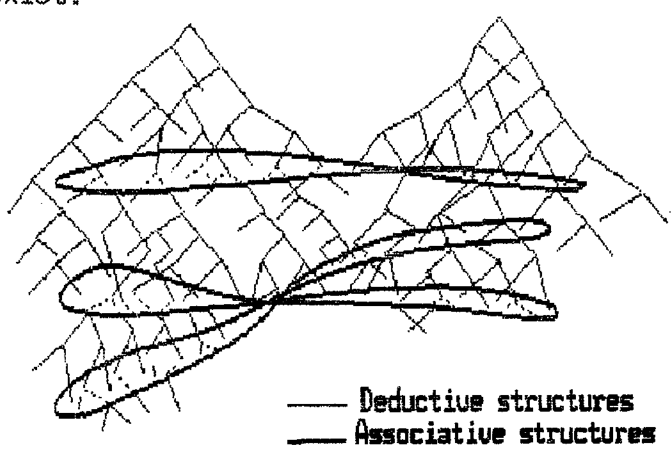


To achieve the objectives proposed, more complex structures must be constructed. A tri-dimensional structure, where the associative structures are organised horizontally, with its concepts involved in hierarchical vertical structures, can be thought of:



—— Deductive structure (hierarchical)
—— Associative structure (non-hierarchical)

This structure can be further expanded to one where multiple associative planes exist:



The construction of such a data-base, is still a far reaching goal, but the work here presented may be a first step to achieve it.