An Associative Information Retrieval Algorithm for a Kanerva-like Memory Model

Slobodan Ribarić, Darijan Marčetić
Faculty of Electrical Engineering and Computing
University of Zagreb
Unska 3, 10000 Zagreb, Croatia
{slobodan.ribaric, darijan.marcetic}@fer.hr

Abstract-This paper presents an associative information retrieval algorithm for a Kanerva-like sparse distributed memory (SDM) model. This memory model is used to implement the associative level of a hierarchical heterogeneous knowledge-base model consisting of multi-levels, starting from an associative level, through to the semantic, rule-based and description-generator level as the top level in the hierarchy. The architecture of knowledge-base was inspired by biological and psychological models. The proposed algorithm retrieves concepts from the associative level based on the similarity between a concept of interest and already stored concepts. The similarity is expressed by a value of the linguistic variable. With this approach it is possible to solve a problem when the inference processes at the semantic level encounter an unknown concept of interest. The algorithm is demonstrated by retrieving concepts that were stored based on the results of psychological experiment.

I. INTRODUCTION

The goal of our research is to develop hierarchical heterogeneous knowledge-base model consisting of multi-levels, starting from an associative level, through to a semantic, rule-based and description-generator level as the top level in the hierarchy. The approach of this work was inspired by biological and psychological models obtained by analyzing how animal and human brains abstract, process and store knowledge from the interaction with the environment [1-4]. Biological support for the hierarchical organization of a brain is presented in the neuroanatomical studies [3, 4]. In [1] authors have concluded: "All the cortical systems we studied displayed a significant degree of hierarchical organization". They have also shown that the visual and somato-motor systems have the organizations that are "surprisingly strictly hierarchical". The heterogeneous organization of the brain is supported with the fact that levels of the cortical systems are displaced at different brain regions with the size, shape and internal structures of neurons dependant on the particular task that is performed [2-4].

On the other side, the psychological support for hierarchical multilevel organization can be found in study by Piaget and Inhelder who have studied the intellectual development of children [5]. Their observations confirm that the acquisition of knowledge about specific objects starts at lower, and proceeds to higher levels of abstraction. Hierarchical learning systems have been demonstrated by the phenomena discovered and studied by the Gestalt psychology [6, 7].

The main functions supported by our hierarchical heterogeneous knowledge-base model are: associative information storing, retrieval and reasoning, as well as, robust inference procedures at the higher levels. The main reason for use of the hierarchical heterogeneous knowledge-base model, with the associative level at the bottom, lies in some of the limitations and drawbacks of the inference processes of the "pure" crisp or fuzzy network-, logical-, procedural- or frame-based knowledge representation schemas (KRS). In models, based on above "pure" KRS, inference procedures are unable to process unknown concepts at a single level of representation. For example, belief networks (Bayesian or probabilistic networks) [8-10], first order logic- and higher order logic-based schemes, qualitative probabilistic networks, rule-based schemes for uncertain reasoning [11], schemes based on the Dempster-Shafer theory [12], fuzzy logic and the fuzzy Petri net theory (FPN) based schemes [10] don't support handling with unknown concepts.

In our model of the hierarchical heterogeneous knowledge-base design it is possible to resolve a problem of handling with unknown concepts at the semantic level in the inference processes if these concepts are stored at the associative level. Note that the storage capacity of the associative level is by multiple orders of magnitude higher then the storage capacity of the semantic level, correspondingly, the number of stored concepts at the associative level is much higher than one at the semantic level. The lowest level of our model - the associative level is implemented with the Kanerva-like Sparse Distributed Memory (SDM) [13].

The related works are as follows. In [14] Ribarić and Lastrić described a heterogeneous hierarchical knowledge-base model called HETHI. It consists of one level of the Kanerva-like SDM that performs the associative retrieval process and supports the initialization of the inheritance process at higher levels – the semantic and rule-based levels. The HETHI supports different reasoning procedures: "pure" associative inference which is performed only by means of the first associative level, the inference procedure defined at the semantic level and the inference procedure based on the cooperation of the associative and semantic level. HETHI was starting point for the system presented in this paper.

In [15], an enhanced version of the SDM, augmented with the use of genetic algorithms, as an associative memory in a 'conscious' software agent CMattie is described. CMattie, as an intelligent agent, interacts with seminar organizers via email in natural language and is responsible for emailing seminar announcements in an academic department. The SDM is a key ingredient in the complex agent architecture that implements global workspace theory, a psychological theory of consciousness and cognition. In this architecture, the SDM, as the primary memory for the agent, provides associations with incoming precepts.

In [16] the authors describe in detail the IDA (Intelligent Distribution Agent) architecture of autonomous software agents as a cognitive model of human cognition that employs the SDM as a working and associative memory. The IDA heterogeneous, hierarchical architecture is composed of a number of different levels, each devoted to a particular cognitive process. An example of a system based on the IDA, used for gathering logistical and medical information from a patient for later use by the triage nurse, is given in [17].

In [18] issues in developing cognitive architectures as generic computational models of cognition are discussed in detail. This paper presents a set of essential desiderata for developing cognitive architectures. It argues for the importance of taking into full consideration these desiderata in developing future architectures that are more cognitively and ecologically realistic. A two level heterogeneous hierarchical design as most appropriate following listed criteria is proposed. Architecture CLARION is described that consists of two levels: the top level that captures explicit processes and the bottom level that handles implicit processes. CLARION provides a conceptual reasoning capability.

In [19] a system architecture for cooperation of heterogeneous information systems is discussed. Cooperation of heterogeneous information systems requires advanced architectures able to solve conflicts coming from data heterogeneity. The project ACSIS (Agents for Cooperation of Secure Information Systems) presents a way to resolve semantic conflicts coming from databases heterogeneity. A multi-level architecture is proposed for the cooperation of heterogeneous information systems. This solution enables semantic conflict resolution by using agents and ontology.

II. THE ASSOCIATIVE LEVEL OF THE HETEROGENEOUS HIERARCHICAL KNOWLEDGE-BASE MODEL

In order to explain the associative information retrieval algorithm, here we briefly described the main characteristic of the associative level organization [20].

Main task of the associative level is to enable and support inference processes for an unknown concept at the higher semantic level. The associative level of the knowledge base [20] is implemented with the use of some concepts related to the addressing mechanism of the Kanerva's SDM [13]. The SDM model is defined in the space of $\{0, 1\}^n$, $n \in \mathbf{N}$, where **N** is a set of the natural numbers, elements of which are n-dimensional vectors with binary components. These vectors are represented as points in an n-dimensional space. The number of points in an n-dimensional space is N = 2^n. N is also used for naming the space itself, i.e., N-space.

The main feature of N-space is its distribution, defined on the basis of the distances among the points. The distance $d(x, y)$ between two points x and y in N-space is defined as the number of corresponding vector components at which they differ, known as the Hamming distance. The distance is, by definition, an integer number in the range from 0 to n. The distance $d(x, y)$ can be used to express the similarity of the points x and y. Two points in N-space that are close to each other are more similar. The concept of the address region is used at the associative level. The address region of an arbitrary address location x in N-space is defined as a circle O with the radius r and the centre x. The address region of x contains a set of points that are at most r bits from x: $O(r,x) = \{y \mid d(x,y) \le r\}$.

The basic characteristics of the above-described SDM model are: the *similarity* and *sparseness* of the memory. *Similarity*, as mentioned above, is based on the distance between points. *Sparseness* is derived from the fact that the actual number of storage locations used is very few compared with 2^n, $n \gg 1$. The storage locations are distributed randomly in N-space and a unique address is assigned to each storage location. Even for a relatively small dimensionality of N-space (for example, n = 100) an exorbitant number of possible locations (N=2^{100}) exists. Let us suppose that only a fraction of the possible address space (for example, N' = 1,000,000) is available and points are randomly distributed over the entire address space. Such a type of space is called a sparse memory. Correspondingly, addressing N-space is now reduced to addressing the locations of the subspace N', N' $\subset$ N called N'-space. The address location in N'-space is represented by an n-dimensional address vector in the same way as in N-space.

A) Relation of a concept to a point in N'-space and a concept group organization

Let C be a set of abstract objects or concepts from the real world that are stored at the associative level. An element from C can be represented with one or more points in N'-space. A point in N'-space is called a location. The concepts are clustered into groups of concepts (locations) based on a user's knowledge and/or intuition. A concept can belong to one or more groups of concepts. Groups of concepts are used because it is natural for humans to organize similar concepts based on their characteristics.

Let us denote a group of points in N'-space by G_i, i =1, …, g (indicated by the name of the group NGC_i) where g is the total number of groups. Each group of points is represented by a uniform, random sample of N'-space. Note that $G_i \cap G_j = \varnothing$ for $i \neq j$, and that each point in N'-space is indicated by its unique address. For simplicity reasons, let us suppose that the number of points in every group is equal to N'/g.

The basic assumption for the proposed model is that the number of points in N'-space is much larger than the number of stored concepts. For mapping a concept $c_i \in C$ into a set of points in N'-space, the concept-address list is used.

A concept-address list for a concept $c_i \in C$, which is stored in the group with the name NGC_j, is represented by the 3-tuple: $\left(c_i, NGC_j, \left(a_1^{i,j}, \ldots, a_s^{i,j}\right)\right)$, where the s-tuple $\left(a_1^{i,j}, \ldots, a_s^{i,j}\right)$ is an ordered list that contains the addresses of points in N'-space; the first super index of a denotes the corresponding concept, the second super index denotes the group G_j and the s denotes the total number of copies of the concept c_i in the group G_j. When the same concept c_i belongs to different groups NGC_k, NGC_l, …, then for each group there is a corresponding concept-address list: $\left(c_i, NGC_k, \left(a_1^{i,k}, ..., a_t^{i,k}\right)\right)$, $\left(c_i, NGC_l, \left(a_1^{i,l}, ..., a_v^{i,l}\right)\right)$, … . Note that the concept-address lists are a rudimentary implementation of Kanerva's basic idea of *the word to focus transformation* defined by mapping of the physical signals to sensory features [13].

B) Similarity between concepts

A linguistic variable L is used to express the similarity between the concepts. The values of the linguistic variable L are from the following set {*minimally, minorly, more-or-less, moderately, considerably, very, extremely*}. The values of the linguistic variable L can be transformed to the Hamming distance between two points in N'-space. For example, in terms of the description of similarity with a value of the linguistic variable L, two concepts described as "*very*" similar will have a smaller distance between the corresponding points in N'-space than the concepts described as "*minorly*" similar.

III. THE CONCEPT STORING ALGORITHM AT THE ASSOCIATIVE LEVEL

The process of storing a concept in the associative level is described in detail in [20]. Brief description of the storage algorithm can be given as follows. A user specifies a concept $c_i \in C$, where C is a set of concepts, with the name of its group NGC_k, where the concept will be stored. Optionally, in the second case, the user can specify a measure of similarity, expressed by a value of the linguistic variable L, to another already-stored concept $c_j \in C$, and the name of its group NGC_l, where $l \neq k$ or $l = k$.

In the first case, when $c_i \in C$ and the name of its group NGC_k are given, the concept will be stored in the randomly chosen free location from its group NGC_k.

In the second case, a concept $c_i \in C$ will be stored in a location or locations from the group with the name NGC_k, based on the measure of similarity between c_i and c_j expressed with the value of the linguistic variable L. The measure of similarity expressed by L is mapped to the Hamming distance d in the N'-space. The concept c_i can be stored in more then one free location. These locations are at the distance d from the all locations where c_j was already stored.

IV. THE ASSOCIATIVE INFORMATION RETRIEVAL ALGORITHM

Set of concepts stored at the associative level is denoted by C, while the set of concepts at the semantic level is denoted by D. Note that D ⊂ C and Card(C) >> Card(D), where Card denotes a cardinality of a set, i.e. the number of concepts stored at the associative level is much larger compared to the number of concepts stored at the semantic level.

The main task of the associative information retrieval algorithm is to solve the problem when the inference processes at the semantic level encounter an unknown concept of interest, i.e. $c_i \notin$ D. Note that the concept of interest c_i is likely to be stored in one or more locations at the associative level due to Card(C) >> Card(D).

The aim of the proposed associative information retrieval algorithm is to retrieve all the concepts from the associative level that have defined similarity measure (expressed with a value of the linguistic variable L) to the concept of interest. Among retrieved concepts it is expected that some of the concepts also belong to the semantic level. Such "inducted" concepts are used to initialise the inference process at the semantic level. If there are no such concepts, the inference process at the semantic level cannot be started.

The associative information retrieval algorithm is formally given as follows:

Input: A concept of interest c_i, name of its group NGC_k (optionally), a user-defined measure of similarity related to the concept of interest expressed with the value l of the linguistic variable L (optionally).

Step 1:

i) **IF** the concept of interest $c_i \notin$ C, **THEN** return the message: "The concept of interest is not at the associative level". **End.**

ii) **IF** the concept of interest $c_i \in$ C **AND** the name of group NGC_k is not given, **THEN** retrieve all concept-address lists $\left(c_i, NGC_u, \left(a_1^{i,u}, ..., a_{s1}^{i,u}\right)\right)$, $\left(c_i, NGC_v, \left(a_1^{i,v}, ..., a_{s2}^{i,v}\right)\right)$, … , where $c_i \in NGC_u$, $c_i \in NGC_v$, … .

iii) **IF** the concept of interest $c_i \in$ C **AND** the name of its group NGC_k are given, **THEN** retrieve the concept-address list $\left(c_i, NGC_k, \left(a_1^{i,k}, ..., a_s^{i,k}\right)\right)$ from the set of concept-address lists (addresses $a_b^{i,k}, b = 1,...,s$ for concept $c_i \in C$, from group with name NGC_k, are defined as n-bit vectors determining the locations in N'-space where the concept c_i is stored, the superscript i of $a_b^{i,k}$ corresponds to the concept index, while k is group index.

Step 2:

i) **IF** the measure of similarity is not given, **THEN** set the measure of dissimilarity $h = 0$ what means "*extremely* similar".

ii) **IF** the measure of similarity is given by the value l of the linguistic variable L, **THEN** determine the measure of dissimilarity $h \in H$ by using the function f_h: $L \rightarrow H$, where $l \in L$, H is the set of the measures of dissimilarity $H = \{0, 1, …, z\text{-}1\}$, and z is the total number of values of the linguistic variable L. The function f_h is given in TABLE I.

Step 3:

The radii r_1 and r_2 define the n-dimensional hollow sphere where the locations with defined measure of similarity lie. Fig. 1. illustrates 2-dimensional example of N'-space where the potential addresses of locations can lie.

Determine the radii r_1 and r_2 as follows:

i) **IF** $h = 0$ **THEN** set $r_1 = 0$, and calculate $r_2 = \lceil r_p \cdot \beta \rceil$, the potential addresses lie in the n-dimensional sphere with radius r_2 ;

ii) **IF** $h > 0$ **THEN** calculate $r_1 = \left\lceil r_p \cdot \left(\alpha \cdot \frac{h-1}{z-1} + \beta \right) \right\rceil$, and

$$r_2 = \left\lceil r_p \cdot \left(\alpha \cdot \frac{h}{z-1} + \beta \right) \right\rceil;$$ where

- $\lceil \ \rceil$ denotes the ceiling of a value of an expression,
- radius r_p, called the basic radius, defines the x'-centred circle $O(r_p, x')$; $x' \in$ N'-space, which contains on average p proportion of points y' in N'-space that satisfy the relation $O(r_p, x') = \{ y' \mid d(x', y') \le r_p \}$. The radius r_p is obtained from TABLE II based on the number of dimensions n of N'-space and the proportion p of N'-space within r_p bits of a point x'. The proportion p is determined experimentally and its value is between 10^{-3} to 10^{-6},
- parameter β; $1 \le \beta \le \frac{n}{2 \cdot r_p}$ defines the minimum value of r_2 when the concepts are extremely similar ($h = 0$, $r_1 = 0$, $r_2 = \lceil r_p \cdot \beta \rceil$),
- parameter α; $0 < \alpha < \frac{n}{2 \cdot r_p} - \beta$, determines the incremental step of the radii r_1 and r_2 according to the dissimilarity measure h.

Step 4:

i) **IF** a name of group NGC_k is not given, **THEN** select all addresses $a_b^{i,t}$, where $t = u, v, \ldots$ and b = $1, \ldots, s1, 1, \ldots, s2, \ldots$ from retrieved concept-address lists $\left(c_i, NGC_u, \left(a_1^{i,u}, \ldots, a_{s1}^{i,u}\right)\right)$, $\left(c_i, NGC_v, \left(a_1^{i,v}, \ldots, a_{s2}^{i,v}\right)\right)$, … obtained in **Step 1**.

ii) **IF** a name of a group NGC_k is given, **THEN** select every address $a_b^{i,t}$, where $t = k$ and $b = 1, \ldots, s$, from retrieved concept-address list $\left(c_i, NGC_k, \left(a_1^{i,k}, \ldots, a_s^{i,k}\right)\right)$ obtained in **Step 1.**

Step 5: For all selected addresses $a_b^{i,t}$ from **Step 4**. find all nonempty locations y_m, $m = 1, 2, \ldots$ from the N'-space, with addresses $a_e^{j,w}$ that satisfy the following conditions: $d(a_b^{i,t}, a_e^{j,w}) > r_1$ and $d(a_b^{i,t}, a_e^{j,w}) \le r_2$,

- addresses $a_e^{j,w}$ define locations where some concepts $c_j \in C$ are stored. These concepts belong to groups with names NGC_w, $w \in \{1, \ldots, g\}$, g is the number of groups,
- $e \in \mathbf{N}$ is "dummy" index that specifies address storage order of a concept c_j .

Step 6: Return the list of concepts that are stored in locations y_m, $m = 1, 2, \ldots$ obtained in **STEP 5**.

TABLE I
MAPPING OF THE VALUES OF THE LINGUISTIC VARIABLE L TO THE VALUES OF THE MEASURE OF DISSIMILARITY H

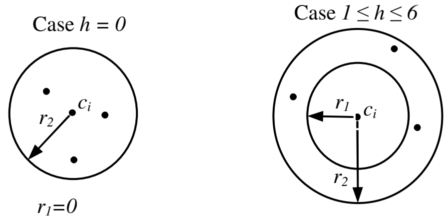
$f_h: L \to H$	
Values of the linguistic variable $l \in L$	Values of the measure of dissimilarity, $h \in H$
extremely	0
very	1
considerably	2
moderately	3
more-or-less	4
minorly	5
minimally	6

TABLE II
RADII r_p OF x-CENTRED CIRCLES THAT ENCLOSE A SPECIFIED PROPORTION p OF N'-SPACE [13]

Proportion p of the N'-space within r_p, bits of point x	Number of dimensions n		
	n=100	n=1,000	n=1,000, 000
	r_p	r_p	r_p
0.000000001	20	405	4,700
0.00000001	23	411	4,720
0.0000001	24	418	4,740
0.000001	26	425	4,760
0.00001	29	433	4,790
0.0001	31	441	4,810
0.0005	34	448	4,840
0.001	35	451	4,850
0.01	38	463	4,880
0.01	44	480	4,940
0.25	47	489	4,970
0.5	50	500	5,000

Case $h = 0$

Case $1 \le h \le 6$



r_1=0

c_i - *concept of interest*

Dots• represent concepts that are *extremely* ($h = 0$) similar to c_i

c_i - *concept of interest*

Dots• represent concepts that are *minimally* ($h = 6$) or *minorly* ($h = 5$) or *more-or-less* ($h = 4$) or *moderately* ($h = 3$), *considerably* ($h = 2$) or *very* ($h = 1$) similar to c_i

Fig. 1. A simplified two-dimensional example of space where the potential addresses of locations can lie.

V. AN EXAMPLE

Let us suppose that N'-space is created as subspace of N-space; N=2^n; n = 100. The N'-space is defined by the following parameters: N' =1,000,000, g = 20, p = 0.0001, α = 0.5 and β = 1. The number of storage locations in every group is equal to N'/g = 50,000. Based on selected values of p and n, the radius r_p= 31 is determined (see TABLE II). Based on associative connections between concepts that are obtained form the psychological experiment [21] concepts were stored by means of concept storing algorithm presented in [20]. TABLE III represents the addresses of locations where the concepts from the psychological experiment are stored at the associative level of the hierarchical heterogeneous knowledge-base. Note, that addresses were determined based on the frequency of the first association to given stimuli word: *needle* [20], [21]. Results depicted in TABLE III are obtained by the program simulator described in [20]. Note that all concepts are stored with assumption they belong to the same group.

TABLE III
RESULTS OF THE CONCEPT STORING ALGORITHM FOR CONCEPTS OBTAINED BY PSYCHOLOGICAL EXPERIMENT [21]

Concepts	Store address (100 bit vector, represented as hex number)
needle	54237B12DF036C08781725B2C
thread	7262F9C212636E4A6C3FA6E26
pin(s)	662773FEB5516488593F3E900
sharp	97014B206BCFC4485A2E2592E
sew(s)	55D159785343586B6135242A5
sewing	0AE25D307F7B8D023A072DF37
steel	3BCF51C2C28AA94E78377D36C
point	12035F96D5EF6AD72C4FE894E
instrument	120B5206C66D5104DDB616BEE
eye	B6834FFE79AA5E9CF536457E9
thimble	F8331B2FA712704B94550C09A

From the TABLE III we can see, for example, that the Hamming distance between the addresses where the concepts *needle* and *thread* are stored is d(54237B12DF036C08781725B2C, 7262F9C212636E4A-6C3FA6E26) = 31. Let us suppose that the concept of interest is c_2 = "*thread*". Let us retrieve all stored concepts for which the measure of the similarity to the concept of interest is "*moderately*". Note that concept "*thread*" is stored with similarity measure related only to the concept "*needle*" and that the similarities to all other concepts are unknown for now.

Associative information retrieval algorithm is applied as follows:

Input: A concept of interest is c_2 = "*thread*", name of its group is NGC_1, measure of similarity is expressed with the value "*moderately*".

Step 1: The concept-address list $\left(thread, NGC_1, \left(a_1^{2,1}\right)\right)$ is retrieved from the set of concept-address lists.

Step 2: For the value "*moderately*" of the linguistic variable L, based on TABLE I, the measure of dissimilarity h = 3 is obtained.

Step 3: Radii r_1 and r_2 are:

$$r_1 = \left\lceil r_p \cdot \left(\alpha \cdot \frac{h-1}{z-1} + \beta \right) \right\rceil, \; r_1 = \left\lceil 31 \cdot \left(0.5 \cdot \frac{3-1}{7-1} + 1 \right) \right\rceil = 37,$$

$$r_2 = \left\lceil r_p \cdot \left(\alpha \cdot \frac{h}{z-1} + \beta \right) \right\rceil, \; r_2 = \left\lceil 31 \cdot \left(0.5 \cdot \frac{3}{7-1} + 1 \right) \right\rceil = 39.$$

Step 4: Select address $a_1^{2,1}$, from $\left(thread, NGC_1, \left(a_1^{2,1}\right)\right)$. The address is 7262F9C212636E4A6C3FA6E26

Step 5: For address $a_1^{2,1}$, from $\left(thread, NGC_1, \left(a_1^{2,1}\right)\right)$, retrieve locations with addresses $a_e^{j,w}$ that satisfy the following conditions: $d(a_1^{2,1}, a_e^{j,w}) > 37$ and $d(a_1^{2,1}, a_e^{j,w}) \leq 39$. Addresses $a_e^{j,w} = a_1^{3,1}$ and $a_e^{j,w} = a_1^{4,1}$ (where concepts c_3="*pin*" and c_4="*sharp*" are stored) satisfy these conditions, i.e. $d(a_1^{2,1}, a_1^{3,1}) > 37$, $d(a_1^{2,1}, a_1^{3,1}) \leq 39$, $d(a_1^{2,1}, a_1^{4,1}) > 37$, $d(a_1^{2,1}, a_1^{4,1}) \leq 39$.

Note that $d(a_1^{2,1}, a_1^{3,1})$ = (7262F9C212636E4A6C3FA-6E26, 662773FEB5516488593F3E900) = 38 and $d(a_1^{2,1}, a_1^{4,1})$ = (7262F9C212636E4A6C3FA6E26, 97014-B206BCFC4485A2E2592E) = 38.

Step 6: Return output list {*pin*, *sharp*}.

It means that concepts *pin* and *sharp* are *moderately* similar to the concept of interest *thread.*

Some already stored concepts (see TABLE III), based on the psychological experiment [21], are used to demonstrate the associative information retrieval algorithm. The results for five stored concepts are obtained by the program simulator and given in TABLE IV.

TABLE IV
SOME RESULTS OF THE ASSOCIATIVE INFORMATION RETRIEVAL ALGORITHM FOR FIVE STORED CONCEPTS FROM THE EXAMPLE

INPUT		OUTPUT
Concept of interest	L	Output list
needle	*extremely*	{*thread, pin, sharp*}
thread	*moderately*	{*pin, sharp*}
sewing	*very*	{*needle*}
steel	*minorly*	{*sew, sewing, point*}
thimble	*more-or-less*	{*needle*}

VI. CONCLUSION

This paper presents the associative information retrieval algorithm for the associative level of the hierarchical heterogeneous knowledge-base model. The model of knowledge-base was inspired by biological and psychological models obtained by analyzing how animal and human brains abstract knowledge from the interaction with the environment. The associative information retrieval algorithm retrieves concepts based on the similarity between concepts expressed by a value of the linguistic variable. This mechanism tries to solve the problem when the inference processes at the semantic level encounter an unknown concept of interest. In this case, the attention is focused back to the associative level and the algorithm retrieves all the concepts that lie in *n*-dimensional hollow sphere corresponding to the value of linguistic variable. Among retrieved concepts it is expected that some of them belong to the semantic level. Such concepts are used to initialise the inference processes at the semantic level.

Future work will involve the implementation of the hierarchical heterogeneous knowledge-base model consisting of multi-levels, starting from an associative level, through to the semantic, rule-based and description-generator level as the top level in the hierarchy. The ultimate research goal is to develop robust and efficient search engine for information retrieval and the World Wide Web queries.

ACKNOWLEDGEMENTS

This work has been supported within the project “Theory, modeling and application of autonomy-oriented computing structures” (no. 036-0361935-1954); Ministry of Science and Technology, Croatia.

REFERENCES

[1] C.C. Hilgetag, M.A. O’Neill and M.P. Young, “Hierarchical organisation of macaque and cat cortical sensory systems explored with a novel network processor”, *Philosophical Transactions of the Royal Society of London Biological Sciences*, vol. 355, p. 71–89, 2000.

[2] D.H. Hubel and T.M. Wiesel, “Receptive fields of single neurons in the cat's striate cortex”, *The Journal of Physiology*, vol. 148, p. 574–591, 1959.

[3] D.H. Hubel and T.M. Wiesel, ”Receptive fields, binocular interaction and functional architecture in the cat's visual cortex”, *The Journal of Physiology*, vol. 160, p. 106–154, 1962.

[4] J. S. Albus, *Brains, Behavior and Robotics*, BYTE Books, McGraw-Hill, New York, 1981.

[5] J. Piaget and B. Inhelder, *The Psychology of the Child*, Basic Books, New York, 1969.

[6] K. Koffka, *Principles of Gestalt Psychology*, Routledge and Kegan Paul, London, 1955.

[7] W. Kohler, *Gestalt Psychology*, Mentor Books, NewYork, 1964.

[8] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, 1995.

[9] Y. Xiang, F. V. Jensen and X. Chen, “Inference in Multiply Sectioned Bayesian Networks: Methods and Performance Comparison”, *IEEE Trans. on System, Man, and Cybernetics*, Vol. 36, No. 3, p. 546-558, 2006.

[10] C. G. Looney and L. R. Liang, “Inference via Fuzzy Belief Petri Nets”, *Proceed. of the 15th Int. Conf. on Tools with Artificial Intelligence (ICTAI '03)*, p. 510-514, 2003.

[11] C. G. Looney, “Fuzzy Petri Nets for Rule-based Decision making”, *IEEE Trans. on the System, Man and Cybernetics*, vol. 18, num. 1, p. 178-183, 1988.

[12] R. R. Yager, M. Federizzi and J. Kacprzyk (eds.), *Advances in the Dempster-Shafer Theory of Evidence*, J. Wiley, 1994.

[13] P. Kanerva, *Sparse Distributed Memory*, The MIT Press, Cambridge, 1988.

[14] S. Ribarić and B. Lastrić, “HETHI- Heterogeneous Hierarchical Knowledge Base”, *Journal of Computing and Information Technology (CIT)*, vol. 2, p 167-179, 1994.

[15] A. Ashraf and S. Franklin, “Sparse Distributed Memory for ‘Conscious’ Software Agents”, *Cognitive System Research*, vol. 4, p 339-354, 2003.

[16] S. Franklin and A. Graesser, “Modeling Cognition with Software Agents”, *Proceedings of the 23rd Annual Conference of the Cognitive Science Society NA*, p. 301-306, 2001.

[17] S. Franklin and D. Jones, “A Triage Information Agent (TIA) based on the IDA Technology”, *AAAI Fall Symposium on Dialogue Systems for Health Communication*, Washington, DC, 2004.

[18] R. Sun, “Desiderata for cognitive architectures”, *Philosophical Psychology*, vol. 17, num. 3, p. 341-373, 2004.

[19] M. Seguran, D. Boulanger and G. Talens, “Semantic Conflict Resolution in Heterogeneous Databases: Interaction Protocols for Domain Ontologies Evolution”, *Proceedings of the International Conference on Artificial Intelligence, IC-AI '04*, vol. 2, p. 1053-1059, 2004.

[20] S. Ribarić and D. Marčetić, “A Storage Algorithm for a Kanerva-like Memory Model”, *32nd International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO*, p. 221-226, 2009.

[21] R. S. Woodworth, *Experimental Psychology*, Henry Holt & Co., New York, 1938.