The use of Medical Ontology in a knowledge-based semantic Fusion system

Roxana - Oana Teodorescu, Vladimir – Ioan Cretu* and Daniel Racoceanu***

**Faculty of Automation and Computers, Department of Computer Science, "Politehnica" University of Timisoara, Faculty of Automation and Computer Science, 2 V.Parvan Str. Timisoara, ROMANIA, e-mail: Vladimir.cretu@cs.upt.ro, roxana.teodorescu@cs.upt.ro*

*** Faculty of Sciences, University of Besancon- France, IPAL-Image Perception, Access and Language, UMI CNRS 2955, NUS, UJF, I2R - A*STAR, Singapore, e-mail: daniel.racoceanu@univ-fcomte.fr*

Abstract – *This article states the role of ontology in a fusion system for integrating the features extracted from medical images. Our aim is to generate semantic trees from ontology by using a semantic network. The theoretical system proposed for placing image features into the ontology is presented in this article. It is based on the UMLS ontology and solves the semantic gap. We also briefly talk about placing this system into nowadays CBIR systems and performing maintenance of it by using the links between OWL and medical ontology*.

Keywords: medical ontology, fusion, UMLS, semantic trees, neurodegenerative diseases

I. INTRODUCTION

The main purpose of the article is to present a fusion proposition technique and show the environment in which it will be applied, along with the elements linked to it. The medical images in this study are brain representations used in the detection of neurodegenerative diseases. The proposed fusion method is based on medical knowledge expressed by decision rules, as well as medical ontology.

The Unified Medical Language System[1] (UMLS) represents, in the system's architecture, the level on which the fusion can be done – i.e. the semantic dimension. In order to transform the features from low level to the UMLS semantic level, we propose to use the Protégé[2] ontology tool.

This paper is structured as follows: ***section II*** presents a short sate of art on the medical systems that are used for Computer aided diagnosis; ***section III*** looks at the problems that are present in the use of ontology, as well as a high-level categorization system for the medical concepts present there; ***section IV*** concludes this article and looks at further possibilities for extending our system and applying it to different areas.

II. EXISTING ONTOLOGY – BASED SYSTEMS

The Medical Algorithm Project (MedAl[3]) is a web-based resource, but also provides computer-executable forms from medical algorithms.

The lack of semantic level qualities gives an advantage to the approaches having a comprehensive organization and description.

[1] *http://www.nlm.nih.gov/research/umls/*

[2] *http://protege.standford.edu/*

[3] *http:// www.medal.org/*

The latter provide the possibility of managing unstructured or semi-structured data.
Another ontology that incorporates concepts from UMLS is the Medical Computational Problems (MCP) ontology, used in the knowledge base from the KnowBaSICS-M system. This ontology describes the medical terms via a controlled vocabulary [1], where the conceptualizations of the domain knowledge for MCP descriptions are constructed as an OWL (Ontology Web Language) model. The medical problem space, the algorithm space and the implementation space represent the dimensions of the MCP ontology. The ontology-based system for semantic management in medicine, KnowBaSICS-M, created by Bratsas et al. [2], which uses this ontology, provides an open environment for the MCP. It is a modular system based on an ontology-based model. This model is created using concepts from UMLS and the Information Retrieval (IR) part is relies on an ontology-based vector space model (VSM).
GALEN is used in a formal representation for International Representation of Diseases (ICD10) [2] and creates a knowledge intensive coding support tool. The result of the representation of ICD10 is an ontology based on description logic. This ontology has been converted to OWL DL. The GALEN system shows a performance of 84% recall and 45% precision [2]. This project started 10 years ago and has its own terminology system (GALEN CRM). The main characteristic of this system is that it provides a model composed by a set of building blocks and constraints from which concepts are made. It also has a description language, GRAIL, which is a logic-like language used in DL and is similar to conceptual graphs. It uses rules at the conceptual level - as opposed to the role definition level - and the Natural Language Processing (NLP) module of the system is a statistical component augmented with a thesaurus.

III. MERGING TWO DIFFERENT IMAGE TYPES INTO A SEMANTIC NETWORK

The current ***Problems*** in the ontology, when generating and maintaining a set of accurate annotated terms, are identified by the difficulty of generating rules and annotations [3][5][6].
The Semantic Gap, in our case, denotes the difference between the features extracted from the images and the medical concepts from the ontology. The features extracted from fMRI[4] and SPECT[5] images - the ones we use for our application - are of visual nature (i.e. color, texture, shape, …) and do not match with the medical terms. The difference between the visual features and the medical terms causes the aforementioned *semantic gap*.
In order to better understand this concept we now explain the UMLS structure of the gap.
The ***Unified Medical Language System*** has the purpose of offering to the Computer Aided Diagnosis (CAD) [4] systems a place where the medical terms are explained, in order to be used. It includes very large, multi-purpose, and multi-lingual vocabulary databases with information about biomedical and health related concepts, various names, and the relationships among them, called the Metathesaurus[6]. Another side of UMLS is illustrated by the Semantic Network[7], which offers a semantic categorization. The concepts are linked together using relations (e.g. anatomical-part-of, spatially-related-to, conceptually-related-to, etc.) in a structure representing a higher level of data processing. We need to bring the extracted features to this level of granularity and give those elements a categorization not only based on a specific relationship, but also on a prioritization of these categories. Finally, the fact that one concept appertains to several

[4] *fMRI = functional MRI(Magnetic Resonance Image)*

[5] *SPECT = Single Proton Emission Computed Tomography*

[6] *http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html*

[7] *http://www.nlm.nih.gov/research/umls/meta3.html*

categories should also be taken into account [6][7].

A possible Solution to the semantic gap is the use of ontology rules to map the visual features, in order to link them with the medical concepts from the Metathesaurus. The transformation from visual features to medical concepts can be done using ontology-related rules.

The concepts mapped from the features, which were part of the semantic network all along, are classified into semantic trees, in order to improve annotation.

The concept of semantic tree has the purpose of conserving the relations from the semantic network. Based on that, we are able to classify the concepts. In this case, the nodes are the medical concepts, and the arcs are the relationships chosen for a certain tree (e.g. the Anatomical tree could contain the "anatomical-part-of" and "spatially-related-to" relationships) (see Fig. 1).

The three classes that we have chosen denote the three dimensions for medical analysis: anatomy, pathology and physiology. Once the semantic trees are generated, they contain all the concepts that are needed from the Metathesaurus, and the learning phase for the trees is complete. The new feature requests activate only some parts of the trees. Each activated part has a value within its own tree, depending on its importance and the image from which it comes – fMRI is better with the anatomical features, whereas the SPECT provides better pathology [15][16].

The fact that the trees are generated using the relationships from the semantic network means that the same concept can be a part of several trees, with the same or different degrees of importance. Also, a concept may or may not be activated from a set of images (see Fig. 2).

In order to obtain an ontology that is annotated and efficient, based on the three medical axes, we have to fuse the common concepts together. In this way, only one instance of a concept, containing the degree of importance given by its tree, will be present in the end.

The fusion process does not loose any of the initial meaning; the relationships between concepts are still there, as are the concepts, but they now each have a value, in the form of an importance degree that can be used for further processing. In our case, we use the fused concepts in order to obtain an early diagnosis system.

The Fusion Process takes place at the semantic level, after the trees have been created. The fusion suited for this level of data granularity is a high level fusion type, which takes into account uncertainty and fuzziness.

For our approach we decided to take into account several fusion operators from the high level fusion methods (table 1).

Table 1: Operators used from the high level fusion

	Probabilistic	Possibility		Evidence
Basic concept	probabilities	possible values		mass function belief
Operation	E - event to be evaluated x_1,x_2 information from sensors $p(E\|x_1, x_2) = \frac{p(x_2\|E)p(x_1\|E)p(E)}{p(x_1)p(x_2)}$ in case of independence between sources	T-norm	$\forall(x,y) \in [0,1]^2$, $i(x,y) \leq min(x,y)$	$\bigoplus_{i=1}^{n} m_i(A) = 1/(1-k)$ $\sum_{B_1 \bigcap B_2 ... \bigcap B_n = A} =$ $m_1(B_1)m_2(B_2)...m_n(B_n)$ k= $\sum_{B_1 \bigcap B_2 ... \bigcap B_n = \emptyset}$ $m_1(B_1)m_2(B_2)...m_n(B_n)$
		T-conorm	$\forall(x,y) \in [0,1]^2$, $i(x,y) \geq max(x,y)$	
		mean	$min(x,y)$ $\leq m(x,y)$ $\leq max(x,y)$, $m \neq min$, $m \neq max$, $m(x,y) = m(y,x)$	

In table 1 [12] there are the specific operations for the high level fusion algorithms, with the basic concepts. The operations presented are context independent with constant behavior, meaning the fact that neither the context in which the operators are applied, nor the behavior of the operators changes. This kind of operations is useful because they provide impartiality for the operators, so that only the influence given by the applied coefficients at each step of the processing count, together with the initial values from the extraction step.

The Coefficients used for giving importance to the coefficients that the medical doctors find more important than the others and also for taking more important rules into account first.

The low-level features extracted from the images are processed in the first step in the system's data flow:

$$C_{i,j} = \alpha_{imgi} \times x_j \quad (1)$$

where $C_{i,j}$ represents the coefficient used as the input for the first step of the system; α_{imgi} represents the degree of trust and x represents the value of the j feature extracted. In this case the input data for the fusion will represent the concepts with their degree of trust according to the network that they are attached to. Also the information that is given to the fusion module must contain, according to the image source providing each concept, a degree of trust. This degree of trust represents the confidence that is given to each feature extracted from each image, according to the image type (e.g.: for the anatomical concepts the MRI is more trusted than the CT). In this way the features extracted from an image type that represent strong points that type of image have more importance than the ones extracted from another type of image.

$$v_{CUIj} = \eta_{A/B} \times C_{i,j} \circledR R_a \quad (2)$$

where v_{CUIj} represents the value obtained at the semantic level for the concept j (CUI is the concept unique identifier from the UMLS); $\eta_{A/B}$ is the importance degree of the feature j extracted from the image i and designated from the coefficient $C_{i,j}$ and obtained from the Level 0 of the system; R_a designates the rule a used for reaching the semantic level. The input data for the fusion level are first prepared for that process by mapping them on the semantic network, but the values and the characteristics of these elements must be transferred to the next level too and even increased. An anatomical and a functional network can be generated from the semantic network using the relationships existent in the Semantic Network and, according to the importance for each element; an importance degree for each concept can be attached before the fusion step, at the object level.

The rules used for mapping the coefficients to the semantic network have different importance, giving different levels of priority, as in the case of differential ontology. The importance of each rule determines the power invested by the medical knowledge in the specified rule. This power degree of each rule is computed as the combination of rules applied and the succession of these rules give the parameter R_a:

$$R_a = \sum \rho_{apl} \times \chi_k \quad (3)$$

where ρ_{apl} represents the coefficients for the rules applied and verified and that can be extended in the Semantic Network using the χ_k relation.

The semantic trees have different priority. The priority is given by the medical knowledge and it determines the values of the coefficients applied for each tree.

The Application of this high level semantic network consists in generating a diagnosis system based on medical knowledge and semantic rules, which include all the image features.

It can also be applied into the Content-Based Image Retrieval (CBIR) systems in the preprocessing phase, in order to analyze the content of the images.

Finally, it can be used for extending the data meaning, by using the UMLS co-occurrence system in order to expand the concept. This process can help in enhancing the extracted meaning from the images, by using concepts that come together with those that have been extracted already, using past experience.

IV. CONCLUSION

The presented semantic categorization represents an important step in the information fusion process, as it uses the highest level of data processing.

The fact that the trees are generated using a set of rules that choose the relationship between the features extracted and the UMLS concepts they belong to, also offers the possibility to update and change them in order to achieve better results, or to combine the features for activating new concepts.

What's more, during the fusion of the trees in the final step, the rules and the operators can be changed in order to affect the weight of each concept.

We intend to apply this solution for neurodegenerative diseases, but the process is the same for all medical areas. It can be extended and used in multiple diagnoses, offering the possibility base a decision on several concepts, between two or more diseases.

The biggest advantage of this system is the fact that it takes into account medical knowledge, in the sense that the images are analyzed using a system from the anatomical, pathological and physiological point of view. The final result is the fusion of knowledge extracted from two different types of images and of medical knowledge.

ACKNOLEGEMENTS

This paper is part of Roxana Teodorescu's research thesis, as final stage on the research of the ontology state of the art and also as a possible solution in order to achieve a diagnosis system.

REFERENCES

[1] Charalampos Brastas, Vassilis Koutkia, Evangelos Kaimakamis et al. " KnowBaSICS-M : An ontology-based system for semantic management of medical problems and computerized algorithmic solutions" - Computer Methods and Programs in Biomedicine 88 (2007) pg.39-51

[2] Gergely Heja, Gyorgy Surjan, Gergely Lukasy et al." GALEN based formal representation of ICD10" – International Journal of Medical Informatics 76 (2007) pg.118-123

[3] Manuela Atencia and Marco Schrolemmer- "Situated Semantic Alignment" – 2006

[4] Christine Golbreight, Olivier Bierlaire, Olivier Dameron and Bernard Gibaud "What Reasoning support for Ontology and Rules?" –Workshop Protégé with rules, July 2005

[5] B.Baghieri Hariri, H. Abolhassani and A.Khodaei "A new structural Similarity Measure for Ontology Alignment", 2005

[6] Fausto Giunchiglia, PavelShivaiko and Mikalai Yatskevich " Discovering Missing Background Knowledge in Ontology Matching" – University of Trento, 2006

[7] Anand Kumar, Matteo Piazza, Barry Smith et al. " Formalizing UMLS Relations Using Semantic Partitions in the Context of Task-Based Clinical Guidelines Model", 2004

[8] OpenClinical : Ontologies : http://www.openclinical.org/ontologies.html

[9] Unified Medical Language System : http://www.nlm.nih.gov/research/umls/

[10] The Unified Medical Language System Metathesaurus http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html

[11] The UMLS Semantic Network : http://semanticnetwork.nlm.nih.gov/

[12] Isabelle Bloch," Information combination operators for data fusion: A comparative review with classification" - IEEE Transactions on Systems, Man, and Cybernetics-Part A: Sytsems and Humans 26 (1996), no. 1, 52 – 67.

[13] Audrey Baneyx, Jean Charlet, and Marie-Christine Jaulent, "Building ontology of pulmonary diseases with natural language processing tools using textual corpora" - International Journal of Medical Informatics 76 (2007), 208–215.

[14] A. Bozzali and M. Cherbini, "Diffusion tensor mri to investigate dementias: a brief review" - Magnetic Resonance Imaging 25 (2007), 969–977

[15] James D. Eastwood, Michael H. Lev, and et al., "Correlation of early dynamic ct perfusion imaging with whole-brain mr -diffusion and perfusion imaging in acute hemispheric stroke," - American Society of Neuroradiology 23 (2003), 1869–1875.

[16] Uwe Pietrzyk, Karl Heroltz, and et al., "Clinical applications of registration and fusion of multimodality brain images from pet, spect, ct and mri" - European Journal of Radiology 21 (1996), 174–182.

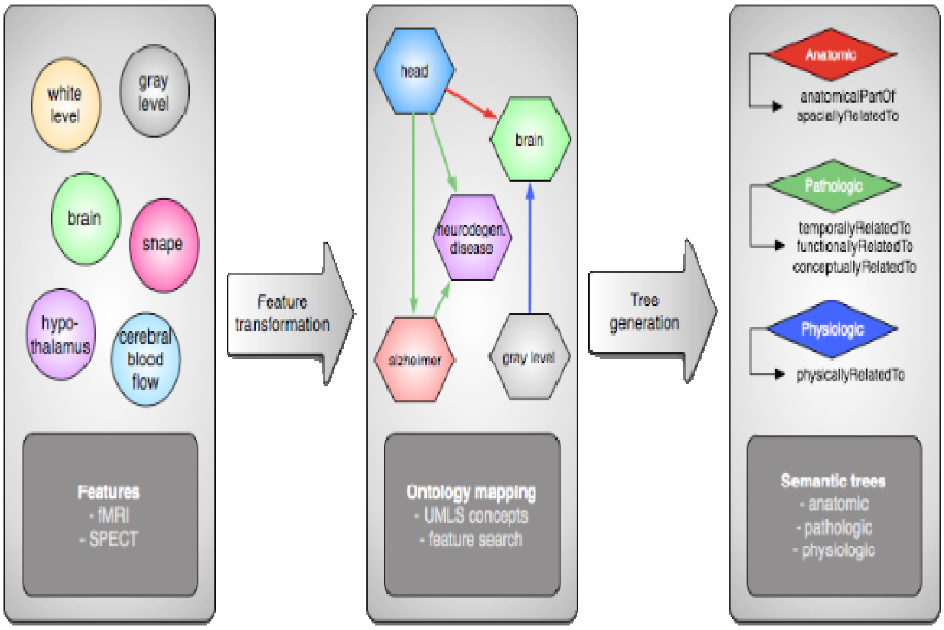

Figure 1 : Ontology tree generation process

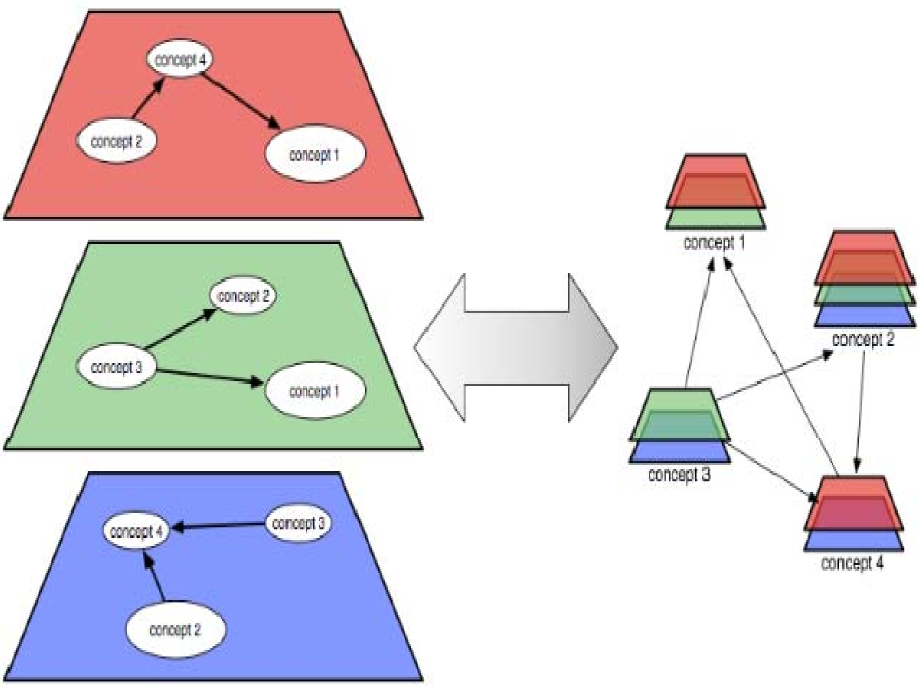

Figure 2: Semantic Fusion Process