2011 IEEE International Conference on Fuzzy Systems
June 27-30, 2011, Taipei, Taiwan

Fuzzy Knowledge Approach to Automatic Disease Diagnosis

Carmen De Maio, Vincenzo Loia
Dipartimento di Informatica
Università degli Studi di Salerno
Fisciano (SA), Italy
{cdemaio, loia}@unisa.it

Giuseppe Fenza, Mariacristina Gallo
CORISA
Università degli Studi di Salerno
Fisciano (SA), Italy
{gfenza, mcgallo}@corisa.it

Roberto Linciano, Aldo Morrone
Azienda Ospedaliera San Camillo-Forlanini
Roma, Italy
rlinciano@yahoo.it,
amorrone@scamilloforlanini.rm.it

Abstract— **Applying best available evidences to clinical decision making requires medical research sharing and (re)using. Recently, computer assisted medical decision making is taking advantage of Semantic Web technologies. In particular, the power of ontologies allows to share medical research and to provide suitable support to the physician's practices. This paper describes a system, named ODINO (Ontological DIsease kNOwledge), aimed at supporting medical decision making through semantic based modeling of medical knowledge base. The system defines an ontology model able to represent relations between medical disease and its symptomatology in a qualitative manner by using fuzzy labels. Medical knowledge is defined according with physician experts members of INMP[1] (National Institute for Health Migration and Poverty). The main aim of ODINO is to provide an effective user interface by using ontologies and controlled vocabularies and by allowing faceted search of diseases. In particular, this work mashes the capabilities of Description Logic reasoners and information retrieval techniques in order to answer to physician's requests. Some experimental results are given in the field of dermatological diseases.**

Keywords- Semantic Web, Ontology, Clinical DSS, Diagnosis

I. INTRODUCTION

Traditional approaches to the medical diagnosis practice have many drawbacks, like as: the huge growth of biomedical information has made difficult the retraining for the individual doctors; the poor dissemination of effective research results; and so on.

New trend is the Evidence-Based Medicine (EBM) which aims to apply the best available evidences gained from the scientific methods to clinical decision making. The challenge of EBM is to define a systematic approach to integrate research results with clinical expertise and patient preferences, and exploit them during the medical diagnosis. So, it is necessary to provide a suitable model to structure medical results and to perform the knowledge spreading and sharing.

There are several ongoing efforts aimed at developing formal models of medical knowledge and reasoning to design decision support systems. These efforts have focused on representing content of clinical guidelines and their logical structure. Semantic Web technologies and ontologies are enabling elements to achieve these aims. In fact, today ontologies are assuming increasingly important role in the area of knowledge based decision support systems by introducing capabilities in terms of logic based reasoning.

This work presents ODINO, a multilingual web based application, that addresses the aim to use semantic web technologies in order to support medical practices through an effective user interface. In particular, ontologies are used to model available medical diseases features (e.g., skin diseases), symptomatologies, treatment protocols and so on. The relations between symptomatologies (i.e., symptoms and signs) and available diseases are represented by using fuzzy labels resulting from the analysis of medical expertise included in [1]. Standard formalisms, like OWL and SKOS, are used to specify domain knowledge and controlled vocabularies, i.e., diseases, symptomatology, active ingredients and clinical tests according to standard specifications, like as ICD-9-CM[2].

Ontologies, controlled vocabularies and information retrieval techniques are exploited to provide typical capabilities of Semantic Web portals [2] and medical decision support. Some of the main features of the system are:

- disease catalogue browsing including images, symptoms and signs, treatments, etc. to support rapid training;
- preliminary medical diagnosis, indeed medical knowledge querying by specifying symptoms, signs and complications, to find eligible diseases;
- faceted search of diseases by enabling multi-criteria selections (i.e., symptomatologies, complications, active ingredients, etc.) to support differential diagnosis.

ODINO results are explained to physician by highlighting symptoms/signs/complications that match the retrieved diagnosis and by suggesting other important features of founded diseases.

The paper is organized as follows. Section II introduce some related works in the applicative domain. Section III describe the knowledge layer of ODINO. The medical decision support methodology is described in Section IV. Then, Section V details features of system and provides the results of the case

[1] http://www.inmp.it/

[2] International Statistical Classification of Diseases, 9th Revision, Clinical Modification

study. Finally, there are conclusions of the paper.

II. Related Works

A medical diagnosis problem consists of the identification of a disease and corresponding treatments that treat it.

In the past years, a great deal of artificial intelligence research has been directed towards the development of expert systems for problem solving in medical diagnosis domain. As examples of developed medical expert systems, we mention: Mycin [3], DXplain [4], Puff [5], Cadiag2 [6], Gideon [7] and Casnet [8]. In particular, interesting approach is used in Cadiag2 that exploits fuzzy set theory to model medical concepts, and fuzzy logic to emulate diagnostic processes. As argued in [9], [10], fuzzy sets offer linguistic label that well approximate medical texts. In addition, fuzzy logic provides reasoning methods capable of making approximate inference [11],[12]. So, fuzzy set theory provides appropriate basis for the development of a computer-based diagnosis system [13]. Nevertheless, there are some problems related with these expert systems consisting in their limited [14],[15]: flexibility, adaptability, extendibility and cooperation capability. Strictly related to our approach, some works, such as [16] and [17], highlight benefits of mashing semantic and fuzzy logic in healthcare. Analogously, ODINO exploits: Semantic Web formalism to represent fuzzy relations between diseases and symptomatologies; and soft computing techniques to arrange diseases that share these relations.

Conversely, solutions like MEDBOLI [18] and ODDIN [19] use Semantic Web Technologies to develop a software that allows users to make diagnosis. In fact, since in medical discipline homogeneity of terminology is mostly problematic, the semantic technologies can be exploited to make known machine-readable latent relationships [20]. Ontologies allow users to understand meaning of each element, and improve reasoning [21], [22]. Regarding efforts which apply semantic techniques, initiatives such as OpenGalen [23] should be mentioned, a not-for-profit organization which provides downloadable open source medical terminology. Other initiatives include, for example, OBO Foundries [24], a collaborative experiment among developers of science-based ontologies. Within the scope of this research there are many resources in use such as Biological Ontologies [25], Ontology-based Support for Human Disease Study and Medical Ontologies to support human disease research and control [26], ‘‘relations in biomedical ontologies’’ [27], and SNOMED CT, a concept-oriented controlled vocabulary [28].

However, today, semantic annotation of Web content with metadata is not very common on health websites, but it is sometimes mandated by, e.g., government standards. The annotation processes are often tedious and require capability in the use of large vocabularies, such as Medical Subject Headings (MeSH) [3]. In this perspective, one of the major contribution of ODINO is that diseases are semantically annotated by using typical standards and controlled vocabularies of medical domain, as well as ICD9. These standards are exploited in order to provide feature of faceted based navigation of diseases. Furthermore, thanks to the benefits of semantic technologies usage (flexibility and configurability of the system), faceted navigation supports multi-criteria selection to discover right diseases. In this sense, ODINO was inspired by previous semantic portals such as SWED [2], the MultimediaN E-culture demonstrator [28] and HealthFinland [30]. Another contribution of ODINO is the definition of sharable *Medical Disease Ontology* model capable to represent relation degree between disease and symptoms coming from statistical analysis included in the book [1].

III. Knowledge Layer

ODINO is a knowledge based system that provides capabilities in terms of clinical decision support system and medical semantic web portal. In particular, ODINO makes an intensive usage of ontologies, controlled vocabularies and other standards typical in the medical domain. This section provides details about medical knowledge modeling of ODINO.

A. Technologies and Standards

Following sections provide some references to used technologies and standards, and introduce their roles in ODINO. Semantic technologies (i.e., OWL and SKOS) have been used to represent medical diseases. On the other hand, knowledge developed in ODINO have been aligned with international standard classification of medical diseases available in ICD9-CM.

1) OWL

Ontologies play an important role in ODINO. Generally, ontologies enable knowledge sharing, exchanging and reusing. Standardization of ontology languages has been an important issue in W3C for years. OWL[4] (Web Ontology Language) is their result and is the standard for ontology definition in the last decade. OWL can be used to explicitly represent the meaning of terms and their relationships. Furthermore, OWL has more facilities for expressing meaning and semantics than others W3C’s standards and thus fosters availability of machine interpretable content on the Web. In particular, OWL has three sublanguages with different expressive power: OWL Lite, OWL DL, and OWL Full, respectively. OWL is based on XML and RDF and all data in an OWL file can be represented as a set of RDF triples.

In this work, we follow W3C's recommendation to build ontologies and construct knowledge bases according to the OWL-DL restrictions. In particular, a *Medical Disease Ontology* has been defined as described in Section III.C.

2) SKOS

SKOS [5] (Simple Knowledge Organization System) is a common data model for knowledge organization systems such as thesauri, classification schemes, subject heading systems and taxonomies. Using SKOS, a knowledge organization system can be expressed as machine-readable data. It can be exchanged between computer applications and published in a machine-readable format in the Web.

[3] http://www.nlm.nih.gov/mesh/

[4] http://www.w3.org/TR/owl-features/

[5] http://www.w3.org/TR/skos-reference/#L895

SKOS is used in our system together with OWL in order to express and share knowledge about medical domain. In particular, ODINO exploits SKOS to write vocabularies and taxonomies, like as: symptomatologies, diseases, drugs, etc. (as described in Section III.B). On the other hand, OWL is used to represent the formal model of disease ontology by defining axioms and constraints.

3) ICD-9-CM

The International Statistical Classification of Diseases and Related Health Problems (ICD) is a way to code and organize diseases and a wide variety of signs, symptoms, abnormal findings, complaints, and external causes of injury or disease, published by the World Health Organization [6] (WHO). It assigns a unique up to six characters long code to every health condition.

In this work, the ninth version of ICD (ICD-9) has been used. In particular, ODINO uses ICD9 - Clinical Modification (ICD-9-CM) that provides additional morbidity details than ICD9. ICD9-CM has been applied to align controlled vocabularies deployed during the knowledge base design. This alignment provides shareability of the knowledge base deployed for ODINO.

B. Controlled Vocabularies & Taxonomies

Some controlled vocabularies and taxonomies (i.e., diseases, symptomatology, clinical tests, active ingredients) have been deployed in ODINO by using SKOS technology. Generally, following SKOS properties have been used to represent vocabularies:

- *preferredLabel* (in both Italian and English languages), to define the favorite name of concept;
- *alternateLabels* (in both Italian and English languages), to associate alternative terms to the same concept in order to specify synonym, and so on;
- *hiddenLabel*, to associate the right ICD-9-CM code to the defined concepts (i.e., disease, symptom, etc.) taking into account results from the alignment process;
- *exactMatch,* to eventually specify relation between concepts with equivalent meaning;
- *broader* and *narrower* to specify hierarchical relations among concepts. These properties are enabling elements to create taxonomies and to support faceted navigation as described below in Section V.C;
- *prefSymbol* and *altSymbol* properties, to join one or more images to concept.

In particular, *exactMatch* allows to maintain ICD9-CM coding and to update the knowledge base with alignment to others classification systems (such as ICD-10 or ICD-11).

Controlled vocabularies and taxonomies developed in ODINO are:

- *Symptomatology* – that contains definitions of concept concerning symptoms and signs, like as: paresthesia, anhidrosis, plaque of psoriasis, etc.. Fig. 1 shows an example of SKOS concept modeling a symptomatology and their hierarchical relations;
- *Disaeses* – that contains medical diseases, their characteristics and hierarchical relations, like as: leprosy, malaria, AIDS, etc.. In particular, broader and narrower properties, in SKOS, allow to define a hierarchy of diseases;
- *Clinical Tests* – that models clinical and laboratory tests useful in diagnostic process, like as: haemochrome, urinalysis, specific blood tests for syphilis, etc.
- *Active ingredients* – that defines the drugs suitable for a specific disease treatment, like as: Dapsone, Rifampicin, Antacids, etc.. Since, ICD9 doesn't contain any drug classification, this vocabulary is not aligned with it.

C. Medical Disease Ontology

This section describes main aspects of knowledge models developed in ODINO. An ontology named *Medical Disease Ontology* has been defined. Through a set of properties and concepts, this ontology models disease characteristics and relations. In particular, *Medical Disease Ontology* includes the definition of a correlation degree between disease and its symptomatology by using fuzzy labels. These properties lead system's diagnosis processes and aid disease information consultation.

One of the main class defined is the concept of Disease which identifies condition of medical disease that causes pain, dysfunction, distress, social problems, and/or death, to afflicted person. Fig. 2 shows a sketch of the ontology by illustrating the relations between *Disease*, *Symptom/Sign* and *Complication* classes. In particular:

- *Symptom/Sign* is the class identifying symptom or sign that influences a disease. Many instances of *Symptom/Sign* may be associated to the same *Disease*. So, the relation *"has symptom"* hasn't any cardinality restriction.
- *Complication* is a medical disease or symptom that represents an unfavorable evolution of disease, health condition or medical treatment. Analogously with *"has symptom",* the relation *"has complication"* hasn't any

```
<owl:Thing rdf:about="#Sin157">
  <rdf:type rdf:resource="&skos;Concept"/>
  <skos:prefLabel xml:lang="en">Hand swelling</skos:prefLabel>
  <skos:prefLabel xml:lang="it">Edema delle mani<skos:prefLabel>
  <skos:altLabel xml:lang="en">Hand edema</skos:altLabel>
   <skos:hiddenLabel>729.81</skos:hiddenLabel>
   <skos:prefSymbol>edema1.jpg</skos:prefSymbol>
   <skos:altSymbol>edema2.jpg</skos:altSymbol>
   <skos:inScheme rdf:resource="#Symptomatology "/>
</owl:Thing>
```

Figure 1. An *example* of symptom specification in SKOS.

[6] http://www.who.int/en/

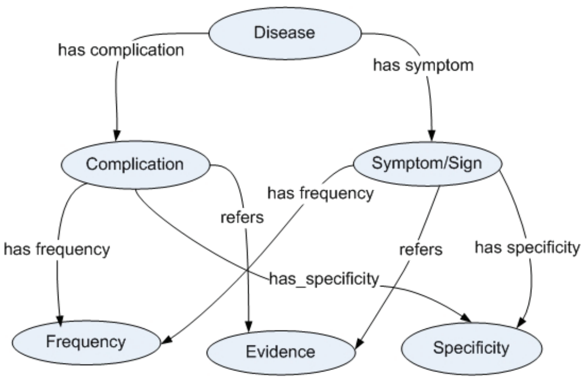

Figure 2. A sketch of *Medical Disease Ontology: Disease, Complication* and *Symptom/Sign.*

cardinality restriction.

Complication and *Symptom/Sign* are themselves related with other classes:

- *Evidence,* that represents the bridging element with the *Symptomatology* SKOS vocabulary;
- *Specificity* and *Frequency,* that allow to indicate how much *Complication* or *Symptom/Sign* implies the presence of that disease.

Furthermore, as shown in Fig. 3, *Disease* is also related with *Clinical test* class. *Clinical test* identifies each possible clinical, laboratory and diagnostic test. Analogously with *Symptom/Sign* and *Complication*, the model foresees the specification of *Precision* degree between a *Clinical test* and a *Disease* (see Fig. 3). In particular, the *Precision* defines how much a *Clinical test* result determines the presence of specific disease.

Frequency, Specificity and *Precision* are enumeration classes defined by using *oneOf* construct available in OWL. More precisely, we have defined, for each of them, some individuals that identify fuzzy labels. Values assumed by these relations are obtained from the analysis of medical practice and experience included in [1]. Specifically:

- *Frequency* - is the rate with a symptom/sign or complication is present in a given disease. Identified levels are: *very frequent, very frequent/frequent, frequent, frequent/occasional, occasional*;
- *Specificity* - is the measure with which manifestation implies a specific disease. Identified levels are: *high, high/medium, medium, medium/low, low;*
- *Precision* - is the level with which *Clinical test* result allows to include or exclude a diagnosis. Identified levels are: *high, high/medium, medium, medium/low, low*.

Furthermore, the model foresees other concepts and relations useful to define additional information of diseases, and recommendation), etiology, epidemiology, transmission of disease, prophylaxis, and so on.

Code listed in Fig. 4 shows an individual definition of Disease as an example. In particular, "Sin189" is of type

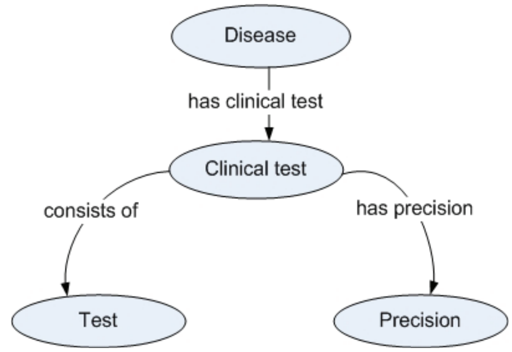

Figure 3. A sketch of some classes of *Medical Disease Ontology*: *Disease* and *Clinical test*.

Evidence and is associated with "*Medium/Low*" specificity and "*Frequent*" frequency degree to disease "*Mal002*". Furthermore, we have associated a complication (*infertility*) and a *NMR (Nuclear Magnetic Resonance)* test.

As detailed in the following section, *Medical Disease Ontology* is the baseline to build our medical diagnosis methodology.

IV. Medical Diagnosys Methodology

A. *Mathematical model to support preliminary diagnosis*

One of the main features supported by ODINO is preliminary medical diagnosis. During medical practices, physician submits a query by specifying one or more evidences that may be symptoms, signs and/or complications. So, the system evaluates correlation degree among incoming query and available diseases by means of an approach based on Information Retrieval techniques and Fuzzy Formal Concept Analysis (FFCA) [31]. Furthermore, system is capable to use relations among concepts in order to augment query results by using a Description Logic reasoner.

As described above, *Specificity* and *Frequency* are the main features to include or exclude diagnosis results. Specifically, a mathematical model of *Disease* and its correlation with *Symptom/Sign* and *Complication* may be obtained in order to perform FFCA.

```
<owl:Thing rdf:about="&Symptomatology ;Sin189">
  <rdf:type rdf:resource="#Evidence"/>
</owl:Thing>
                              ...
<owl:Thing rdf:about="#Sem_fever">
   <rdf:type rdf:resource="#Symptom/Sign"/>
   <refers rdf:resource="&Symptomatology ;Sin189"/>
   <frequency rdf:resource="#Frequent"/>
   <specificity rdf:resource="#Medium-Low"/>
</owl:Thing>
                              ...
<owl:Thing rdf:about="&Diseases;Mal002">
   <rdf:type rdf:resource="#Disease"/>
   <has_semeiotics rdf:resource="#Sem_fever"/>
   <has_complication rdf:resource="#Com_infertility"/>
   <has_clinical_test rdf:resource="#Acc_RMN"/>
</owl:Thing>
```

Figure 4. *Medical Disease Ontology*: individual disease example.

The FFCA, informally, exploits a matrix that represents the fuzzy relation between objects of the input data (i.e., *Diseases*) and some attributes (i.e., *Symptoms/Signs* and *Complications*). The relation is calculated by assessing both *Frequency* and *Specificity* degree between *Disease*, *Symptom/Sign* and *Complication*. In particular, the value of relation, $\mu_{i,j}$, between $Disease_i$ and $Symptoms/Signs_j$ or $Complications_j$ is calculated by performing a linear combination (a weighted sum) of *Specificity* and *Frequency* with parameters λ_1 e λ_2. These parameters consist of constant values chosen based on medical experience, depicted in [1], which shows that, during preliminary diagnosis, *Frequency* is less important than *Specificity*. Formally:

$$\mu_{i,j} = (Specificity * \lambda_1) + (Frequency * \lambda_2), \lambda_1 > \lambda_2 \quad (1)$$

Taking into account the matrix created according to these criteria, FFCA arrange data in a corresponding lattice [31]. Specifically, diseases that share symptoms/signs or complications are arranged together. As argued, the lattice is used to retrieve concepts closer to the incoming query in order to retrieve eligible set of diseases.

B. Identification of candidate disease

In this process the main aim is to get a synthetic value which represents how each resource is relevant with respect to the user query. By revising some Information Retrieval measures (i.e., Precision and Recall), we determine what is the concept (or concepts) most related to the incoming symptoms/signs or complications in the query. In particular, the standard measures of Precision, Recall and F-measure have been adapted to our objective.

Formally, let us consider the following notions:

- $A = \{a_1, a_2, \dots, a_n\}$ the whole set of attributes (i.e., symptoms, signs, complications) belonging to object;
- $O = \{o_1, o_2, \dots, o_t\}$ the whole set of objects (i.e., diseases);
- $C_i = (A_i, O_i)$ *i-th* formal concept of a lattice (where A_i and O_i represent, respectively, terms and resources for the concept C_i);
- $Q_i = \{q_1, q_2, \dots, q_m\}$ a query (e.g., set of symptoms/signs, etc.).

Thus, given a formal concept C_i, we calculate the values of precision P_i and recall R_i as follows:

$$P_i = \frac{|Q_i \cap A_i|}{|A_i|} \qquad R_i = \frac{|Q_i \cap A_i|}{|Q_i|} \quad (2)$$

where *Pi* is computed as ratio between the number of relevant attributes in C_i and all the attributes belonging to C_i. And, R_i represents the number of relevant attributes in C_i divided by the total number of query terms.

Consequently, F-measure value F_i, relative to concept C_i, is computed as follows:

$$F_i = 2 \times \frac{P_i \times R_i}{P_i + R_i} \quad (3)$$

Final result of this computation is the list of the F-measure values $F = \{F_1, \dots, F_s\}$ for all formal concepts in the lattice, given a query.

In order to evaluate diagnosis results, for each object (disease), $o_i \in O$ the system evaluate score as follows:

$$score(o_i) = \sum_{j=1}^{s} \left(\mu_j(o_i) \times F_j\right) \quad (4)$$

where $\mu_j(o_i)$ represents the membership of object o_i (in according to definition of *Fuzzy Formal Context* in FFCA theory [31]). The score represents correlation degree between query and available diseases. Then, a ranked list of diseases may be retrieved in order to answer to a given query.

V. CASE STUDY

Taking into account our experimental results, next subsections detail main features of ODINO. Actually, ODINO has been tested with dermatological disease knowledge. Defined vocabularies include: 700 symptoms/signs; 150 diseases; 125 clinical tests; 260 active ingredients.

A. Disease Catalogue Browsing.

This feature accomplishes the aim to provide rapid training on specific diseases. In fact, ODINO knowledge base includes many multimedia information, such as: definition and historical data, etiological agents, disease transmission, large set of pictures revealing specific symptomatology, complications, clinical tests and treatment protocols, and so on.

B. Preliminary medical diagnosis

This functionality allows to assist physician during preliminary diagnosis practice. Physician starts the process by selecting some clinical manifestations (i.e., symptoms, signs and/or complications). As shown in Fig. 5, ODINO supports selection through two alternative interactions, namely: by browsing ICD-9 tree and specifying ICD-9 categories; by using text based search.

The matching algorithm (described in Section IV.B) allows the system to retrieve eligible diseases. So, taking into account clinical manifestations selected by the user (e.g., *Gangrene* and *Physical decay*), as highlighted in Fig. 6, the system retrieves a ranked list of eligible diagnosis. In particular, the system shows correlation degree useful to rank the results (see Fig. 6(a)) and it allows to access to motivations of each result (in Fig. 6(b)).

Result motivations consist of hints to the physicians about distinctive symptomatology of retrieved disease. These hints may be useful to suggest eventually escaped factors. For example, as shown in Fig. 7, system highlights: submitted symptoms in the request (red outlined box) and *pathognomonic signs* (green outlined box) for the retrieved disease (i.e., in the picture *"Venereal ulcer"*). This is a useful information because the presence of a *pathognomonic sign* means, beyond any doubt, that particular disease is present.

In order to validate preliminary diagnosis performed by ODINO, several queries and relevance sets have been defined with physician experts of dermatological diseases.

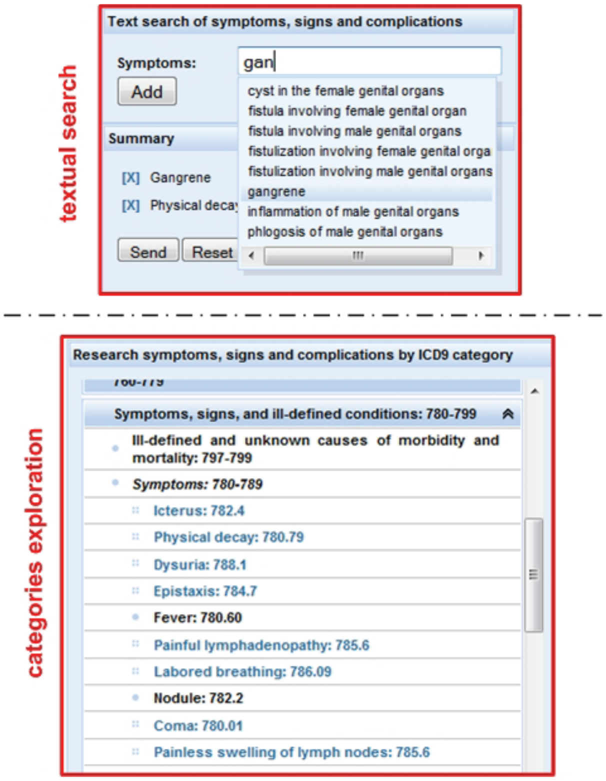

Figure 5. Selection of clinical manifestation by mean textual search (with autocomplete) or categories exploration.

In particular, through the analysis of retrieval performance, we have measured: *Average Uninterpolated Precision* (AUP) [32] and *Precision*. The AUP is defined as the ratio between the sum of the precision value at each point of hierarchical structure (or node of lattice) where a relevant item appears, and the total number of relevant items. The performance for Precision and AUP are shown in Fig. 8, considering a different number N of the clinical manifestations (e.g., *Fever*, *Headache*, *Anemia*, etc.). Let us note that, for N larger than 6, the values of Precision and AUP provide good performance results (between 0.7 and 1). Considering these results we can say that knowledge base and methodology are enough selective for available diseases.

More generally, by analyzing system usage and physician expectancy, experimental results reveal that 86% of diseases retrieved by the system are coherent with results expected by the physicians; just the remaining 14% of results emerge to be too vague. Specifically, by considering wrong results, we have identified that the problem is strictly related to ambiguous clinical manifestations. More precisely, two weakness have been revealed: cases in which huge number of clinical manifestations characterize the same disease (i.e., *AIDS*); and cases in which the same clinical manifestation characterizes many disease with similar *Frequency* and *Specificity* degree (i.e., *Fever*).

C. Faceted search

ODINO also provides a faceted search of diseases through taxonomy constraints (i.e., Symptomatology, etc.). Faceted search [33] is an exploration technique for structured datasets based on the facet theory.

As described in [34], in faceted search the information space is partitioned using orthogonal conceptual dimensions of the data. Each dimension is called facet and represent important disease partitioning feature. The facet has multiple restriction values and the user selects a restriction value to filter relevant items in the information space. The facet based approach can be directly mapped to navigation in semi-structured data, as well as in *Medical Diseases Ontology* model deployed in ODINO. Specifically, available facets are: symptoms, complications, clinical tests and active ingredients. All of these facets are associated with controlled vocabulary defined according to the SKOS formalism.

Obviously, items that ODINO filters step-by-step are diseases. Fig. 9 shows an example of faceted search in ODINO. The left side of Fig. 9 shows constraint “step-by-step” selection to filter diseases (constraints are highlighted with box outlined in green color). At each step of the facet-based search, the system returns an ordered list of eligible diagnosis allowing also further filtering.

With respect to text based search, a faceted search is more useful because allows navigation of an unknown dataset through system suggested restriction values at each step. Additionally, facets provide an intuitive user interface eliminating the need to write exact queries; and prevent empty query results by including restriction values that certainly lead to true results.

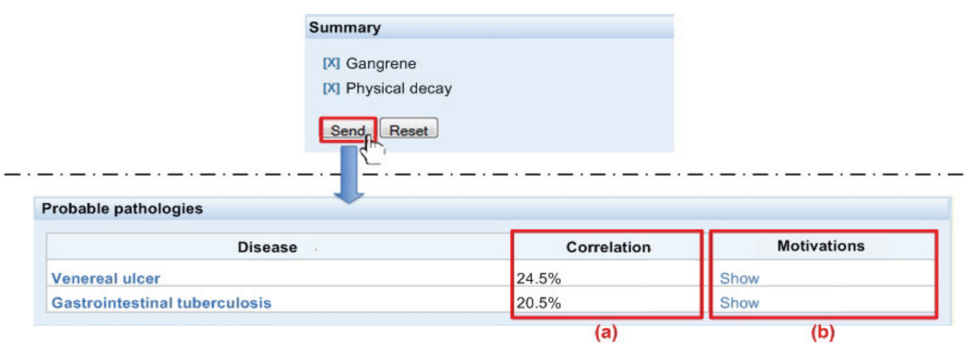

Figure 6. Preliminary diagnosis retrieved results: (a) correlation degree, (b) links to motivations of results.

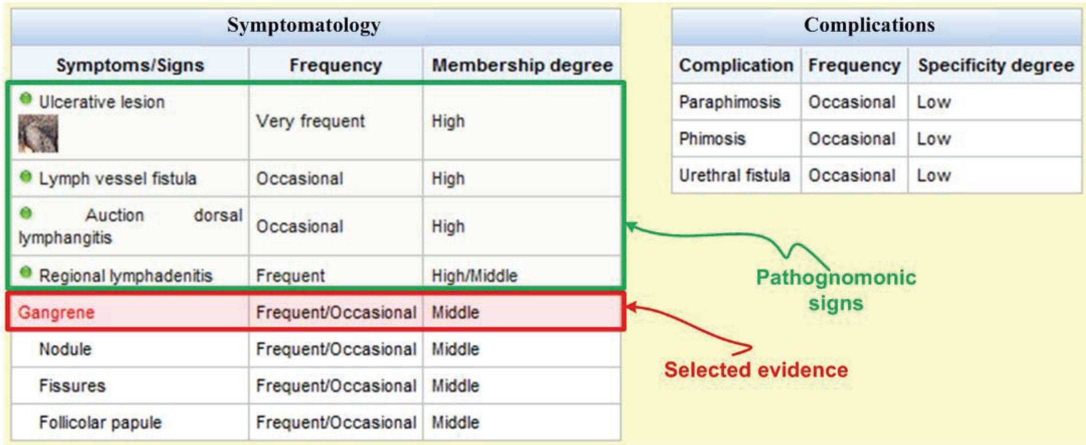

Figure 7. Example of result motivation interface.

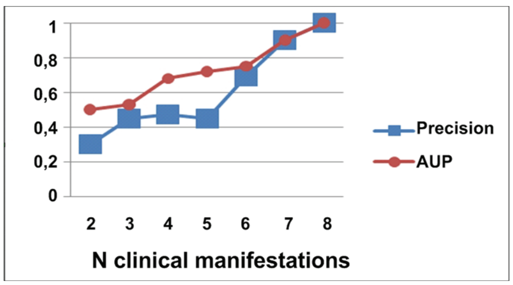

Figure 8. Performance evaluation on Precision and AUP.

VI. Conclusion & Future Works

This paper presents a project named ODINO, an ontology based web application that provides features to daily support medical practices. In particular, ODINO exploits Semantic Web formalisms in order to model medical knowledge. On the other hand, Fuzzy Formal Concept Analysis and ontology inference are applied to support preliminary medical diagnosis.

The availability of rich set of multimedia information allows to provide rapid training feature. Moreover, thanks to the semantic modeling of knowledge base, a faceted search of diseases has been developed.

Benefits deriving from the applied approach are especially related to the reusability of the knowledge layer, scalability and flexibility of the main features, like as faceted search.

The main goal of ODINO is to introduce an effective user interface to provide second opinion especially for the developing country, such as: Etiopia, Kenya, etc. where the availability of economic and human resources are limited. Future works are related to introduce support to improve the exploitation of multimedia information, available in the knowledge base, during faceted search and preliminary diagnosis.

Acknowledgment

The results of the present work have been achieved thanks to the intensive contribution of physician members of INMP that are also involved as authors of [1].

References

[1] A. Morrone. GLOBAL DERMATOLOGY ricerca clinica e logica matematica in Medicina delle Migrazioni, Manuale pratico.

[2] D. Reynolds, P. Shabajee, S. Cayzer, Semantic Information Portals, in: Proceedings of WWW 2004, Alternate track papers & posters, ACM Press, New York, New York, 2004.

[3] E. H. Shortliffe. Computer-Based Medical Consultations: MYCIN. Elsevier, New York, 1976.

[4] Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: an evolving diagnostic decision-support system. JAMA 1987;258:67-74.

[5] J. S. Aikins, J. C. Kunz, E. H. Shortliffe, R. J. Fallat. Puff: An expert system for interpretation of pulmonary function data. Comput. Biomed. Res., 3(16):199–208, 1983.

[6] K. Adlassing. Cardiag 2 expert system. IEEE Transactions on Systems, Man and Cybernetics, SMC-16(2), 1986.

[7] Edberg SC. Global infectious diseases and epidemiology network (GIDEON): a world wide Web-based program for diagnosis and informatics in infectious diseases. Clin Infect Dis. 2005; 40(1): 123-126.

[8] S. M. Kulikowski, C. A. Weiss. Representation of expert knowledge for consultation: the CASNET and EXPERT projects. Artificial Intelligence in medicine. Szolovits, P. (Ed.), Boulder: Westview Press, pages 21–56, 1982.

[9] H. Bossel, S. Klaczko, and N. Muller, A fuzzy-algorithmic approach to the definition of complex or imprecise concepts, in Systems Theory in the Social Sciences, , Eds. Stuttgart: Birkhauser Verlag, 1976, pp. 202-282.

[10] L. A. Zadeh, Linguistic variables, approximate reasoning and dispositions, Med. Iniform, vol. 8, pp. 173-186, 1983.

[11] L. A. Zadeh, Lotfi A., Outline of a new approach to the analysis of complex systems and decision processes, IEEE Trans. Svst., Man, Cybern., vol. 3, pp. 28-44, 1973.

[12] R. E. Bellman and L. A. Zadeh, Local and fuzzy logics, memo. ERL-M584, Electronics Research Laboratory, College of Engineering, University of California, Berkeley 94720, May 11, 1976.

[13] K. -P. Adlassnig, A survey on medical diagnosis and fuzzy subsets, in Approximate Reasoning in Decision Analysis. M. M. Gupta and E. Sanchez Eds. New York: North-Holland, 1982, pp. 203-217.

[14] B. Iantovics, C. Chira, D. Dumitrescu. Principles of the Intelligent Agents. Casa Cartii de Stiinta Press, Cluj-Napoca, 2007.

[15] J. Kuhl, E.J. Graham. Esagent: Expert system control of simulated agent-based mobile robots. Intelligent Systems Research Laboratory, Technical Report TR-ISRL-04-02 University of Louisville, Louisville, 2004.

[16] Acampora, Giovanni; Lee, Chang-Shing; Wang, Mei-Hui; , FML-Based Ontological Agent for Healthcare Application with Diabetes. *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09.* vol.3,

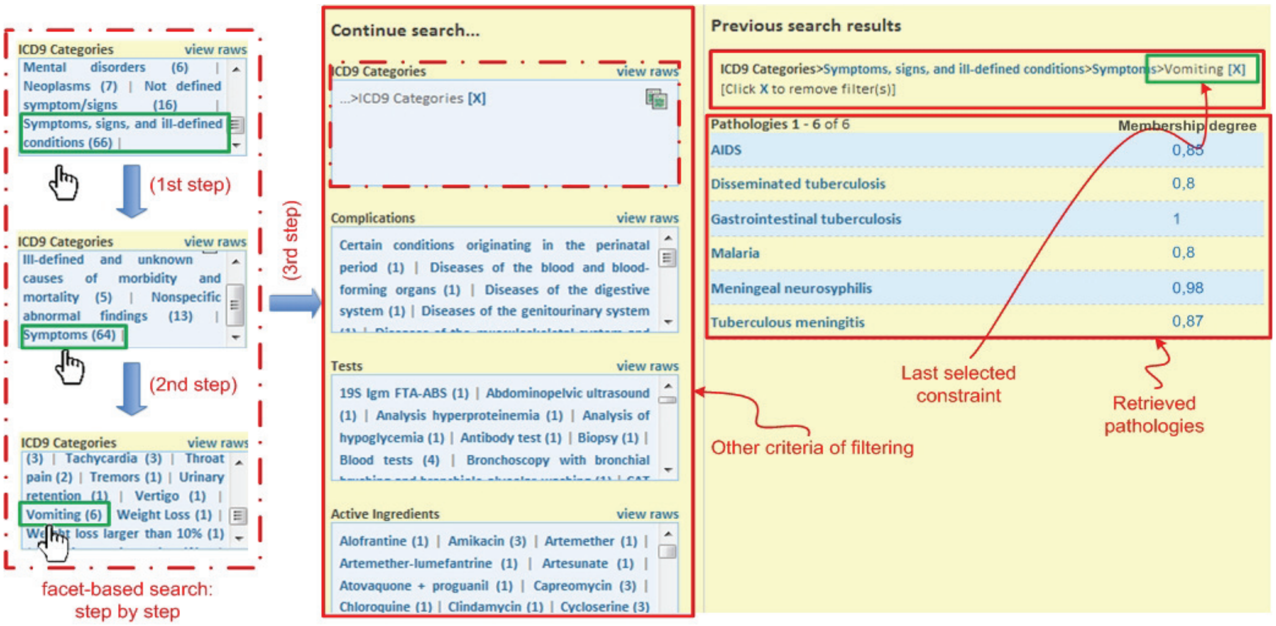

Figure 9. An example of faceted search of diseases.

no., pp.413-416, 15-18 Sept. 2009.

[17] Chang-Shing Lee; Mei-Hui Wang; Acampora, G.; Loia, V.; Chin-Yuan Hsu; , "Ontology-based intelligent fuzzy agent for diabetes application," *Intelligent Agents, 2009. IA '09. IEEE Symposium on* , 2009.

[18] Rodríguez A., Mencke M., Alor Hernandez G., Posada Gómez R. and J. M. Gomez., Medboli: Medical Diagnosis Based on Ontologies and Logical Inference. The Third International Conference on Digital Society, ICDS 2009 (accepted, to appear), 2008

[19] García-Crespo Á., Rodríguez González A., Mencke M., J. M. Gomez., Colomo Palacios R.: ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. Expert Syst. Appl. 37(3): 2621-2628 (2010)

[20] Fuentes-Lorenzo, D., Morato, J., & Gómez, J. M. (2009). Knowledge management in biomedical libraries: A Semantic Web approach. Information Systems Frontiers, 11(4), 471–480.

[21] García Sanchez, F., Fernandez Breis, J., Valencia García, R., Gómez, J. M., & Martínez- Bejar, R. (2008). Combining Semantic Web technologies with multi-agent systems for integrated access to biological resources. Journal of Biomedical Informatics, 41(5), 848–859.

[22] Gomez, J. M., Colomo-Palacios, R., Mayoral, M. R., & Garcia-Crespo, A. (2008). Microarray information and data integration using SAMIDI. In J. R. Rabuñal, J. Dorado, & A. Pazos (Eds.), Encyclopedia of artificial intelligence. Hershey, PA: IGI Global.

[23] Rector, A. L., Rogers, J. E., Zanstra, P. E., & Van Der Haring, E. (2003). OpenGALEN: Open source medical terminology and tools. In Proceedings of the American medical informatics association symposium 2003 (pp. 982–985).

[24] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology, 25(11), 1251–1255.

[25] Lambrix, P., Tan, H., Jakoniene, V., & Strömbäck, L. (2007). Biological ontologies. In C. J. O. Baker & K. H. S. Cheung (Eds.), Semantic Web revolutionizing knowledge discovery in the life sciences (pp. 85–89). Berlin: Springer.

[26] Hadzic, M., & Chang, E. (2005). Medical Ontologies to support human disease research and control. International Journal of Web and Grid Services, 1(2), 139–150.

[27] Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., et al. (2005). Relations in biomedical ontologies. Genome Biology, 6(5), 1425–1433.

[28] Schulz S, Hanser S, Hahn U, Rogers, J. The semantics of procedures and diseases in SNOMED CT, Methods Inf Med. 2006; 45(4): 354-358.

[29] A. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. Kersen, M. Niet, B. Omelayenko, J. Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, B. Wielinga, MultimediaN E-Culture demonstrator, in: Proceedings of ISWC 2006, Athens, GA, Springer–Verlag, Berlin, 2006.

[30] Suominen, O., Hyvönen, E., Viljanen, K. and Hukka, E.: HealthFinland-A National Semantic Publishing Network and Portal for Health Information. Web Semantics: Science, Services and Agents on the World Wide Web, 7(4), pp. 287--297 (2009).

[31] De Maio, C., Fenza, G., Loia, V. and Senatore, S. (2010), Knowledge structuring to support facet-based ontology visualization. International Journal of Intelligent Systems, 25: 1249–1264. doi: 10.1002/int.20451.

[32] N.Nanas, V.Uren and A. de Roeck, "Building and Applying a Concept Hierarchy Representation of a User Profile", In Proceedings of the 26th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, 2003.

[33] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *CHI* 2003.

[34] Eyal Oren, Renaud Delbru, and Stefan Decker. Extending Faceted Navigation for RDF Data. ISWC 2006, volume 4273 of Lecture Notes in Computer Science, chapter 40, pages 559–572.