

Ontology Based Medical Diagnosis Decision Support System

Dr. M.S. Anbarasi

Pondicherry Engineering College, Pondicherry, India

Naveen. P[#], Selvaganapathi. S^{*}, Mohamed Nowsath Ali. I[#]

[#]Information Technology, Pondicherry University

Puducherry, India

^{}Information Technology, Pondicherry University
Puducherry, India*

Abstract—Worldwide health centre scientists, physicians and other patients are accessing, analysing, integrating, and storing massive amounts of digital medical data day by day. To transfer and integrate data from all possible resources, a deeper understanding of all these data sets is required. Since the data users are not the data producers, they face challenges to integrate heterogeneous data. In order to obtain the ability to integrate heterogeneous data there is an urgent need of an evidence-based medicine community of biomedical data integration. Hence we propose ontology based system which can effectively represent the data for future use. Ontology mapping is employed by way of comparing several ontologies. This is also improved using partitioning. Thus we can integrate medical data from different perspectives and take preventive measures based on ontology-derived symptoms in heterogeneous medical records.

Keywords — Serialization, Ontology Slots, Facets, Reasoner, Asserted and Inferred hierarchy, Portioning.

I. INTRODUCTION

The representation of knowledge plays a vital role in the medical field. In our day-to-day life enormous amount of medical data are wasted by just dumping without representing it in some format. Representing these medical records in the form of ontology will be useful in the future to take decisions. Representing the medical records in some common form provides means by which other systems can access and process them and allows to be shared over multiple systems which helps the patient to keep all their records with them for better treatment in the future. Worldwide health centre scientists are accessing, analysing, integrating, and storing massive amounts of digital medical data day by day. To transfer and integrate data from all possible

resources, a deeper understanding of all these data sets is required. Since the data users are not the data producers, they face challenges to integrate heterogeneous data. In order to obtain the ability to integrate heterogeneous data there is an urgent need of an evidence-based medicine community [1] of biomedical data integration [1]. Hence we propose ontology mapping system to integrate medical data from different perspectives and take preventive measures based on ontology-derived symptoms [3] in heterogeneous medical records.

II. RELATED WORK

Ontology is one among the concepts for which researches have been undertaken worldwide in order to extract efficient use from it in the design of medical decision support systems. The following are some of the important related works for our proposed OBMD2S2.

Alan Jovic et al., [1] have elaborated on the work structure of medical ontologies and their construction. For each domain one has to specify the scope of the ontology, acquire knowledge on the domain of concern, select a tool and an ontology language, design the ontology and present it in an appropriate way.

Jingshan Huang et al., [3] discuss in detail about their survey on ontology-based knowledge discovery and sharing systems in bioinformatics. Worldwide health centre scientists, physicians and other patients are accessing, analysing, integrating, and storing massive amounts of

digital medical data day by day. To transfer and integrate data from all possible resources, a deeper understanding of all these data sets is required. Since the data users are not the data producers, they face challenges to integrate heterogeneous data. In order to obtain the ability to integrate heterogeneous data there is an urgent need of an evidence-based medicine community of biomedical data integration.

Marc Ehrig and York Sure [2] have given an account on an integrated approach for ontology mapping. Semantic mapping between ontologies is a core issue to solve for enabling interoperability across the Semantic web. To handle the increasing number of individual ontologies it becomes necessary to develop automatic approaches.

AnHai Doan et al., [5] define a machine learning approach for ontology mapping. They discuss about the problem of finding the semantic mappings between two given ontologies. This problem lies at the heart of numerous information processing applications. Virtually any application that involves multiple ontologies must establish semantic mappings among them, to ensure interoperability. Despite its pervasiveness, today ontology matching is still largely conducted by hand, in a labour-intensive and error-prone process. The manual matching has now become a key bottleneck in building large-scale information management systems. The advent of technologies such as the WWW, XML, and the emerging Semantic Web will further fuel information sharing applications and exacerbate the problem. Hence, the development of tools to assist in the ontology matching process has become crucial for the success of a wide variety of information management applications.

Ashiq Anjum et al., [6] elucidate the requirements for ontologies in medical data integration. According to their research, Evidence-based medicine is critically dependent on three sources of information: a medical knowledge base, the patient's medical record and knowledge of available resources, including, where appropriate, clinical protocols. Patient data is often scattered in a variety of databases and may, in a distributed model, be held across

several disparate repositories. Consequently addressing the needs of an evidence-based medicine community presents issues of biomedical data integration, clinical interpretation and knowledge management

III. ONTOLOGY CONCEPTS

Ontology is a computational model of some portion or domain of the world. The medical ontology describes the semantics of the terms used in the medical domain. Ontology consists of a finite set of concepts, along with these concepts' properties and relationships. In addition, most real-world ontologies have very few or no instances. Medical Ontology is a model of the knowledge from a medical domain. It contains all of the relevant concepts related to the causes, diseases, symptoms and other patient data. The purpose of designing ontology is to allow the system to be capable of knowledge inference and reasoning.

Formally, an ontology O can be defined as a 4-tuple $O = (C, R, I, A)$, where,

C is the set of concepts;

R is the set of binary relations,

I is the set of instances, and

A is the set of axioms.

Thus, according to this definition, each ontology should primarily have four sets of components, which are described below.

1) *Concepts*: Concepts (also called Classes) of an ontology are abstract object (i.e., whose existences are independent of time and location) categories or types in real world. Concepts can generalize (i.e., contain), or can specialize (i.e., subsume) other concepts. For example, the concept 'Person' in a clinical domain can further be specialized by other concepts like 'Patient', 'Physician', 'Nurse' etc.

2) *Relations*: Relations (also known as properties) in ontology are binary predicates which relate between two concepts, or two relations.

3) *Instances*: The Instances are the basic 'ground level' objects for concepts in ontology [7]. For example, 'Diabetes Type 2' can be an instance of

the concept 'Medical-Problem' in a medical ontology.

4) *Axioms*: Axioms in an ontology are formulas (i.e., propositions in mathematics) to specify the interdependencies of concepts or relations on other components (i.e., on other concepts, relations, instances) of that ontology.

Ontology Heterogeneity is an inherent characteristic of ontologies developed by different parties for the same domains. The heterogeneous semantics may occur in two ways. Different terminologies could be used by different ontologies to describe the same conceptual model. In simpler terms, same concept could use different terms, or different concepts could adopt an identical term. Even though the two ontologies use the same name for a concept, the associated properties and the relationships with other concepts are not same. Ontology Matching is the process of determining correspondences between concepts from heterogeneous ontologies which are often designed by distributed parties. Ontology Mapping is also known as Ontology Schema Matching or Ontology Alignment. Such a correspondence may include many relationships. Some of the examples for the relationships are equivalentWith, subclassOf, superClassOf and siblings. Terminological databases can be categorized based on their basic organization unit from a linguistic point of view. There are two types of terminologies. They are Headword with its synonyms and a concept with its different wordings.

IV. ONTOLOGY STRUCTURE

Ontology is made up of classes, properties or slots, relationships between classes and individuals. Individuals are elements of the particular domain. Classes are collections or groups of individuals. Properties or Slots are the relationship between classes or individuals. An example of a medical ontology class is «Disease». It is the super-class of all the other Disease types. All the diseases come under the class «Disease» as sub-class. All the diseases of Circulatory System will come under the class «Circulatory

System» which has a sub-class «ChronicRheumatics» but the super-class of this class «Circulatory System» is «Disease». A class can be more general (upper class) or more specific (subclass), e.g. a specific class of «Disease» is «Circulatory System». Ontology always has a most general class. In our case the «Disease» class acts as the most general class. There is no strict and unambiguous way in which medical knowledge must be represented. The class «Cause» can be placed on the first level of the reason for the each and every disease hierarchy. Bacteria are one of the causes for certain diseases. So the class «Bacteria» is specified in the second level with an individual; «Intracellular Bacteria».

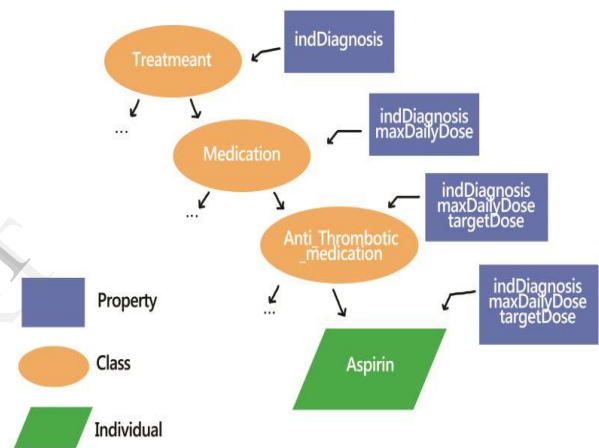


Fig. 1 Ontology Structure

V. ONTOLOGY CREATION

It was mentioned that there exists no single protocol on how to construct a medical ontology or any other type of ontology. An ontology can be constructed manually [2] or (semi)automatically [2]. Manual extraction has been done for the heart failure ontology. In any case, a person who constructs the ontology needs to have some experience in ontology construction and some knowledge of the domain. Usually, domain experts are consulted to explain the meaning of domain-specific concepts. The process of ontology construction can be divided into several steps.

A. Source for ontology creation

Constructing ontologies usually starts with the specification of the desired area of reasoning, especially determining the model boundaries and the level of detail. There is an option to use already existing ontologies or some of their parts in the designing process. Which parts of the existing ontologies are used depends on the domain and application. After the ontology has been finished, it becomes possible to import it into some previously constructed ontology of a higher generalization level as well as to reuse it later in a similar domain. This is the preferred way to achieve cooperation with existing knowledge models.

When one has to construct a higher-level ontology, then one also has to use concepts that are more abstract. In this case, many higher-level classes would have to be only abstract, thus containing no individuals. These classes would create a framework for other, more specific classes to fit in. Discerning relevant from irrelevant concepts should be pursued. This will determine the level of detail that the ontology models.

In addition to scope, it is important to determine the sources of medical information. The most common case in building ontology is to base the ontology vocabulary on related medical guidelines. This means that all the relevant data from the guidelines has to be represented in a systematic way using a hierarchy of concepts and relations. Other sources of medical knowledge include medical articles, other medical ontologies or terminologies and most importantly, experts' knowledge. The manual extraction of facts and terms by human reading from sources of medical knowledge is a reliable method when one has to construct ontologies for decision support tasks.

B. Tools and Languages

After determining the knowledge sources, the next step is to decide which tool and language will be used in order to design the ontology. The choice of the language is usually between Frames and OWL, although other open ontology languages like DAML+OIL can be used. If reasoning and web presentation should be

supported and the open-world is assumed, the OWL is the best choice. If the purpose is only knowledge sharing and terminology/taxonomy, while the closed world assumption is required, then Frames ontology is both sufficient and adequate [8].

The choice between ontology representation tools is another matter. There is always an option of constructing the ontology by directly writing an OWL/RDF file. However, this approach is not practical and requires in-depth understanding of both OWL and RDF syntax and semantics. Graphical tools for the ontology development such as Protégé, SWOOP and many others are freely available. It is the opinion of the authors that Protégé is one of the best choices for a free software ontology development platform. SWOOP is practical when one wants to consult the existing ontologies on the web and compare them or use them as a reference.

C. Ontology design

After a language and a tool have been selected, the process of designing the ontology begins. Essentially, there are two standard approaches to the ontology design. First one is that smaller parts of the ontology are constructed first and then later integrated to form the ontology using higher-level abstract classes. This is the bottom-up approach that is not used often in medical applications, but can be used in, for example, chemical engineering [1]. The other way is to principally design the upper classes (i.e. the skeleton of ontology) and then develop small pieces of the hierarchy, so called top-down approach. This is used for large medical ontologies as well as terminologies [1]. Though, probably the best way of creating ontology is to combine both approaches in an iterative way. It is recommendable to begin the process by creating classes first, then add properties or slots and finally conclude with individuals.

It is noteworthy to mention that there exist some regularity concerning the ontology classes, which the ontology creator should bear in mind. First, the concept from which a class is named should be known and already described in some terminology. This is particularly true for the

smaller scale, lower-level classes. For instance, «Hypertension» is a class that exists in most of the medical terminologies and signifies a disorder of high blood pressure. It can be further divided into two classes or individuals called «Systolic hypertension» and «Diastolic hypertension». It is prudent to give a class the most recognized name for that concept.

«Hypertension» could also be named «High blood pressure», but it should not be named «An elevation of the vein pressure». Second, there should be at least one reference per class to a known medical terminology like UMLS or ICD. If the class has no references in any medical terminology, then there should exist at least a reference to a guideline page, or an article from which this concept was taken. It is possible, though, to have higher-level classes with no references, since they represent more general concepts that sometimes do not exist in the medical terminologies, like «Classification» or «Feature». This should be avoided for lower level classes and especially for individuals.

The number of properties that a class possesses should always be kept as minimal as possible. In larger ontologies, it is usual that two or more classes use the same property. However, the semantics of this property can differ. For instance, the property «Weight» is a general property that describes a physical property of an object. When this property is used for the class «Patient» and for the class «Aldosterone_receptor_blocker» (which is a medication group), the meaning is quite different. Patient's weight is presented in kilograms. It also varies frequently in time. An Aldosterone receptor blocker's weight is the weight of a pill, given in miligrams and usually a constant value. The solution is to reorganize the property «Weight» into two properties, «PatientWeight» and «PillWeight».

D. Ontology and Reasoner

It is important to point out that any ontology is only a knowledge base. If one wants to reason using the ontology, one has to design and implement a decision support system. An example is given in. This figure illustrates an example scenario in the experimental decision

support system in which the ontology has a central position. The event [1] that occurred in a system is served through the DSS interface to the DSS control unit [1]. The control unit initiates [1] the extraction of factual knowledge from the database [1]. Relevant patient data is then transformed to the ontology format [7] and prepared for reasoning [9] as a set of facts. The reasoning process is performed and conclusions reached are loaded back into the ontology [10], which is then analysed by the ontology interpreter [8]. The information acquired by the analysis is served through the DSS interface [1] back to the system user [5].

One of the main services offered by the reasoner is to test whether or not one class is a subclass of another class. We can construct inferred ontology by performing such tests on the classes. Another service offered by the reasoner is, consistency checking. Based on the description of a class the reasoner can check whether or not it is possible for any class to have instance and hence consistent.

Different OWL Reasoners can be plugged into standard tools like Protégé. FaCT++ is the default reasoner that comes along with Protégé. FaCT++ is the new generation of the well-known Fact OWL-DL reasoner. FaCT++ uses the established FaCT algorithms, but with a different internal architecture. Additionally, FaCT++ is implemented using C++ in order to create a more efficient software tool, and to maximise portability.

VI. ONTOLOGY MAPPING

Ontology mapping is seen as a solution provider in today's landscape of ontology research. As the number of ontologies that are made publicly available and accessible on the Web increases steadily, so does the need for applications to use them increases. A single ontology is no longer enough to support the tasks envisaged by a distributed environment like the Semantic Web. Multiple ontologies need to be accessed from several applications. Mapping could provide a common layer from which several ontologies could be accessed and hence could exchange information in semantically

sound manners. Developing such mappings has been the focus of a variety of works originating from diverse communities.

A. Similarity in ontology

The basic assumption is that knowledge is captured in an arbitrary ontology encoding. Based on the consistent semantics the coherences modeled within the ontology become understandable and interpretable. From this it is possible to derive additional knowledge such as, in our case, similarity of entities in different ontologies. An example shall clarify how to get from encoded semantics to similarity: by understanding that labels describe entities in natural language one can derive that entities having the same labels are similar. This is not a rule which always holds true, but it is a strong indicator for similarity. Other constructs as subclass relations or type definition can be interpreted similarly.

The formal definition of similarity for ontologies is as follows:

- O_i : ontology, with ontology index $i \in \mathbb{N}$
- $\text{sim}(x, y)$: similarity function
- e_{ij} : entities of O_i , with $e_{ij} \in \{C_i, R_i, I_i\}$, entity index $j \in \mathbb{N}$
- $\text{sim}(e_{i_1j_1}, e_{i_2j_2})$: similarity function between two entities $e_{i_1j_1}$ and $e_{i_2j_2}$ ($i_1 \neq i_2$); as shown later this function makes use of the ontologies of the entities compared

Due to the wide range of expressions used in this area (merging, alignment, integration etc.), we want to describe our understanding of the term "mapping". We define mapping as [2]: "Given two ontologies A and B, mapping one ontology with another means that for each concept (node) in ontology A, we try to find a corresponding concept (node), which has the same or similar semantics, in ontology B and vice versa." Other but similar definitions are given by [2]. We want to stick to this definition, more specific we will demand the same semantic meaning of two entities.

Formally an ontology mapping function can be defined the following way:

- $\text{map} : O_{i_1} \rightarrow O_{i_2}$

- $\text{map}(e_{i_1j_1}) = e_{i_2j_2}$, if $\text{sim}(e_{i_1j_1}, e_{i_2j_2}) > t$ with t being the threshold entity $e_{i_1j_1}$ is mapped onto $e_{i_2j_2}$; they are semantically identical, each entity $e_{i_1j_1}$ is mapped to at most one entity $e_{i_2j_2}$

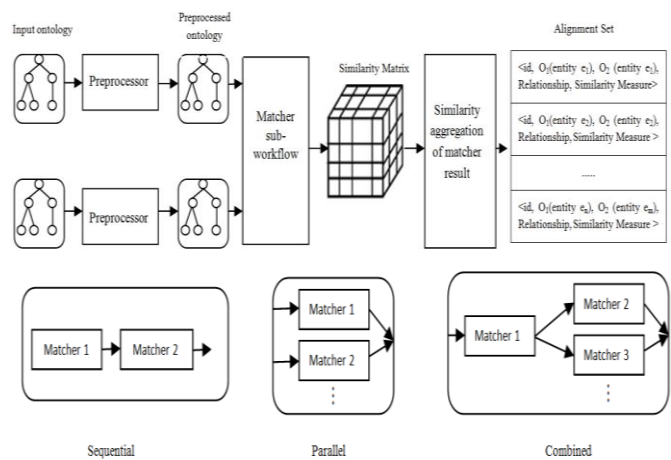


Fig. 2 Ontology mapping system

B. Partitioning in Large Scale ontologies

Large-scale ontologies are a kind of ontologies created to describe complex real world domains. Large class hierarchies are one of the most common kinds of large-scale ontologies. These large ontologies or class hierarchies for the same domain aren't unique. Examples can be found in: (a) Web directory structures, e.g., Google and Yahoo [1]; (b) product description standards, e.g., NAICS1 and UNSPSC2; and (c) medicine or biology, e.g., GALEN3 and FMA4. In order to achieve interoperability among Semantic Web applications using these large ontologies or class hierarchies, ontology matching is necessary. However, the size and the monolithic nature of these large ontologies or class hierarchies cause a new challenge to current ontology matching techniques. Therefore, some novel solutions are required.

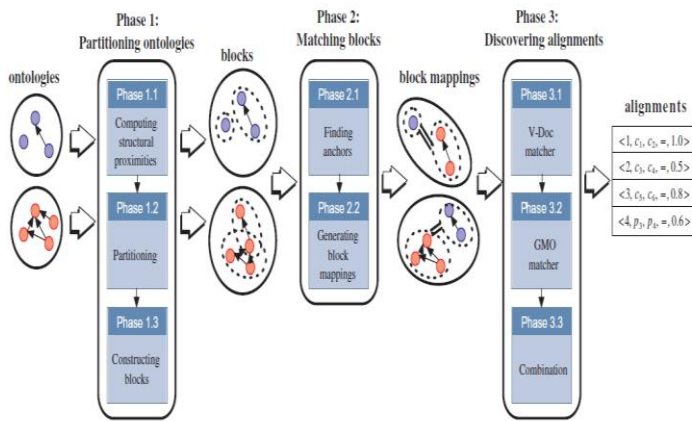


Fig. 3 Partition based Ontology Mapping system

Our partitioning algorithm is an agglomerative hierarchical partitioning algorithm mainly inspired by ROCK []. The main difference between ROCK and ours is that ROCK assumes that all the links between classes are the same; while we import the notion of weighted links, which reflect the information about the closeness between classes. Our algorithm accepts as input the set of n blocks to be clustered, which is denoted by B , and the desired number of blocks k , which is initially determined by application requirement. In each partitioning iteration, it selects the block having the maximum cohesiveness firstly, then choose the block having the maximum coupling with it, and finally merge these two blocks into a new block. The pseudo code of the algorithm is presented here.

```

procedure( $B; k$ )
for each block  $B_i$  in  $B$ , do begin
  initialize the internal sum of links within  $B_i$ ,
  called cohesiveness;
  initialize the sum of links between  $B_i$  and others,
  called coupling;
end
while the number of current blocks  $m > k$  do begin
  choose the best block  $B_i$ , which has the
  maximum cohesiveness;
  choose one block from the rest, which has the
  maximum coupling;
  merge block  $B_i$  and  $B_j$  named  $B_p$ ;
  update  $B_p$ 's cohesiveness and coupling;
  remove  $B_i$  and  $B_j$ ;
for each block other than  $B_p$ , update its coupling;

```

```

 $m := m - 1$ ;
end
end

```

The time complexity of this algorithm is $O(n^2)$. Compared with most other clustering or partitioning algorithms, it is quite efficient. Though k-means method is faster, it is worthy of noting that the means of the blocks are virtual entities, and if we change the means to the real entities (called k medoids method), the time complexity also becomes $O(n^2)$

VII. PERFORMANCE ANALYSIS OF OBMD2S2

The performance parameters used are:

- Precision is the measure of correctness

$$\text{Precision} = |A \cap R| / |A|$$

Where A is the Alignment Set of the algorithm and R is the Reference Alignment Set

- Recall is the measure of completeness

$$\text{Recall} = |A \cap R| / |R|$$

- Execution Time

The mapping problem arises in many scenarios.

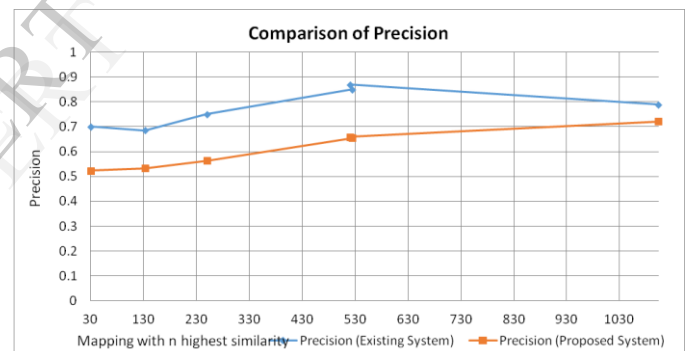


Fig. 4: Comparison of Precision

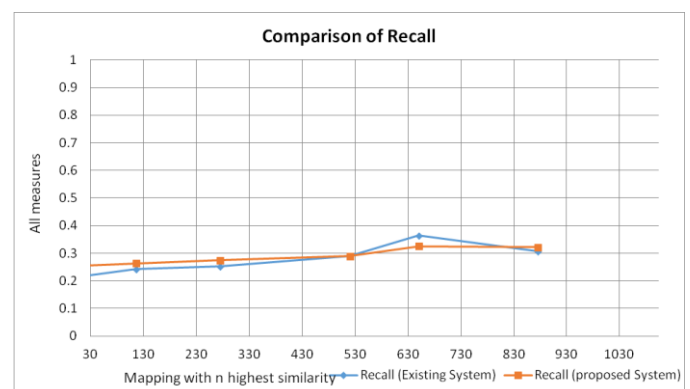


Fig. 5 Comparison of Recall

We have shown a methodology for identifying mappings between two ontologies based on the intelligent combination of manually encoded rules. Evaluation proved our initial hypothesis, i.e. the combination of our presented similarity measures led to considerably better results than the usage of one at a time. One can summarize that precision, recall, and f-measure increase by 20% compared to label-based approaches. Semantics helps bridging the mapping gap.

VIII. CONCLUSION AND FUTURE WORK

Thus the medical ontology has been created and the performance of the decision support system has been increased considerably by way of ontology mapping. The basic advantages of using ontology representation like: standardization of medical terms, knowledge sharing, and support for automatic reasoning have been achieved. The contribution of the work is presentation of the construction process for medical ontologies. The lesson learned from the presented work is that OWL+SWRL is an interesting combination for reasoning in complex medical systems. The problem with large scale ontologies that arise during comparisons is also overcome by partitioning.

Even though the shown approach retrieves good results, the results are not 100% correct. This might be tolerable in some scenarios. Unfortunately, if full-automatic mapping is done, and inference builds on top of it, wrong results can bring down the value of the whole mapping process. Implications of this will have to be understood well when using it. A common

approach to circumvent this problem is to declare the process as semi-automatic rather than doing full-automatic mapping.

REFERENCES

- [1] Alan Jovic, Marin Prcela, Dragan Gamberger, "Ontologies in Medical Knowledge Representation". Rudjer Boskovic Institute, Laboratory of Informational Systems.
- [2] Marc Ehrig, York Sure, "Ontology Mapping – An Integrated Approach". Institut für Angewandte Informatik und Formale Beschreibungsverfahren.
- [3] Yannis Kalfoglou and Marco Schorlemmer, "Ontology mapping: the state of the art". The Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC)
- [4] Matthew Horridge, "A Practical Guide To Building OWL Ontologies Using Protege 4" Edition 1.3. The University of Manchester
- [5] AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy, "Ontology Matching: A Machine Learning Approach". Department of Computer Science University of Illinois, Urbana-Champaign, IL, U.S.A.
- [6] Ashiq Anjum, Peter Bloodsworth, Andrew Branson, Tamás Hauer, Richard McClatchey, Kamran Munir, Dmitry Rogulin, Jetendr Shamdasani, "The Requirements for Ontologies in Medical Data Integration". CCS Research Centre, CEMS Faculty, University of the West of England.
- [7] Kambiz Houshiaryana, Il Kon Kimb, YunSik Kwakb, Hune Chob, "Ontology for Patient Medical Record in Healthcare Organizations". Daegu, Computer Science, Intelligent Information Laboratory, Kyungbook National University South Korea.
- [8] Paul Buitelaar, Philipp Cimiano and Bernardo Magnini, "Ontology Learning from Text". DFKI, Language Technology Lab.
- [9] Barbara Heller, Heinrich Herre, Kristin Lippoldt, "The Theory of Top-Level Ontological Mappings and its Application to Clinical Trial Protocols". Onto-Med Research Group Institute for Medical Informatics, Statistics and Epidemiology (IMISE)
- [10] Wei Hu, Yuanyuan Zhao, and Yuzhong Qu, "Partition-Based Block Matching of Large Class Hierarchies". School of Computer Science and Engineering, Southeast University, Nanjing 210096, P. R. China.
- [11] Ying Wang, Weiru Liu, and David Bell, "A Concept Hierarchy based Ontology Mapping Approach". School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast.
- [12] David CORSAR, Laura MOSS, Derek SLEEMAN Malcolm SIM, "Supporting the Development of Medical Ontologies". Department of Computing Science, University of Aberdeen, Aberdeen, UK.
- [13] owl.cs.manchester.ac.uk
- [14] protege.stanford.edu