

A Flexible Biomedical Ontology Selection Tool

GILBERT MAIGA and DDEMBE WILLIAMS*
Faculty of Computing and IT, Makerere University

Abstract.

The wide adoption and reuse of existing biomedical ontologies available in various libraries is limited by the lack of suitable tools with metrics for their evaluation by both naïve users and expert ontologists. Existing evaluation tools perform technical evaluation of the structure of an ontology as presented by its design and knowledge representation. These find little use in evaluating the processes and representation of granularity presented by biomedical ontology. In this paper we present an evaluation tool as part of a flexible framework that enables users to select a suitable biomedical ontology for use in building applications that integrate clinical and biological data. Requirements for such a tool were elicited in a descriptive survey using questionnaires, and a prototype developed. The tool also enables ontology modelers to iteratively elicit new requirements for improving upon existing biomedical ontologies, leading to new ones that are able to integrate data across structure, processes and granularity. By facilitating biomedical ontology evaluation, the tool contributes towards reuse of existing biomedical ontologies. This helps to avoid the need for costly time consuming tasks of developing entirely new biomedical ontologies. The utility of this tool was demonstrated in experiments that evaluated the infectious disease ontology. The results were validated using a questionnaire based human assessment.

Categories and Subject descriptors: H. [Information Systems]: H.0.General

General terms: Evaluation Tool; Biomedical Ontology; Ontology Integration; Ontology Selection Tool

Additional key words and phrases: Biological data; Clinical data.

IJCIR Reference Format:

Gilbert Maiga and Ddembe Williams. A Flexible Biomedical Ontology Selection Tool. International Journal of Computing and ICT Research, Special Issue Vol. 3, No. 1, pp. 53-66. <http://www.ijcir.org/Special-Issuevolume3-numbe1/article6.pdf>.

1. INTRODUCTION

The wide adoption and reuse of existing biomedical ontologies available in various libraries is limited by the lack of suitable tools and lack of knowledge about properties users require to select a suitable ontology for a task [Alani and Brewster, 2006; Kalfoglou and Schorlmer, 2006]. Existing evaluation tools are known to differ by scope, evaluation type, purpose, inputs, process, metrics used, outputs and the level of user involvement. These perform technical evaluations of the structure of an ontology as presented by its design and knowledge representation [Alani& Brewster, 2006; Fernandez *et al.*, 2006; Hartmann *et al.*, 2004; Lozano-Tello & Gomez-Perez, 2004; Tartir *et al.*, 2005]. These tools therefore find little use in evaluating the processes and representation of granularity presented by biomedical ontology.

In this paper we present an evaluation tool as part of a flexible framework that enables users to select a suitable biomedical ontology for use in building applications that integrate clinical and biological data.

* Author's Address: Gilbert Maiga and Ddembe Williams. Faculty of Computing and IT Makerere University gmaiga@cit.mak.ac.ug

"Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than IJCIR must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee."

© International Journal of Computing and ICT Research 2009.

International Journal of Computing and ICT Research, ISSN 1818-1139 (Print), ISSN 1996-1065 (Online), Special Issue Vol.3, No.1 pp. 53-66, October 2009.

International Journal of Computing and ICT Research, Special Issue Vol. 3, No. 1, October 2009.

Requirements for such a tool were elicited in a descriptive survey using questionnaires, and a prototype developed. The tool also enables ontology modelers to iteratively elicit new requirements for improving upon existing biomedical ontologies, leading to new ones that are able to integrate data across structure, processes and granularity.

The utility of the tool was demonstrated in experiments that evaluated the infectious disease ontology. The results were validated using a questionnaire based human assessment. The tool and human assessment results had high positive correlation (>0.5) when matched by scope, granularity and biomedical integration metric scores for the different use case scenarios. This is an indication of the success of the tool used, and the validity of our approach. By facilitating biomedical ontology evaluation, the tool contributes towards reuse of existing biomedical ontologies. This helps to avoid the need for costly time consuming task of developing entirely new biomedical ontologies.

The rest of this paper is organized as follows. Section 2 discusses current approaches to ontology evaluation. Section 3 explores existing ontology evaluation tools and their metrics, while section 4 explains how the tool was derived. Section 5 explains the prototype tool and its utility in evaluating the infectious disease ontology. Conclusions are given in section 6.

2. CURRENT APPROACHES TO ONTOLOGY EVALUATION

There is no unifying definition of ontology evaluation [Gangemi *et al.*, 2005]. Evaluation determines the quality and adequacy of an ontology for use in a specific context for a specific goal [Fernandez *et al.*, 2006]. It is a technical judgment of the contents of an ontology with respect to requirements specifications, competency questions, or a meta ontology as a frame of reference [Gangemi *et al.*, 2005; Gomez-Perez, 2004; Guarino and Welty, 2002]. Ontology evaluation is important in order for them to become widely adopted and reused by industry and the wider web community [Alani and Brewster, 2006; Kalfoglou and Schorlmer, 2006]. There is also lack of knowledge about properties users require to select a suitable ontology for a task [Alani and Brewster, 2006]. Existing approaches to ontology evaluation use various contexts and conduct evaluation at different levels of complexity. A taxonomy of evaluation approaches based on type and purpose that adopts levels [aspects] of vocabulary, taxonomy, semantic relations, application, syntax, structure and design is provided by Brank *et al.* [2005] and given in table 1.

- 1) Gold standard approaches that compare an ontology to a manually built golden standard ontology or other representation of the problem domain for which an appropriate ontology is needed [Gomez_Perez, 1994uarino, 1998]. Lexical or conceptual levels of evaluation may be used. Lexical comparison uses similarity between the lexicons of two ontologies. At the conceptual level, taxonomic structure and relations are used for comparison [Sabou, *et al.*, 2006].
- 2) Task based approaches that use the ontology in an application and evaluate the quality of results [Porzel and Malaka 2004]. This approach has problems in that: (i) it is difficult to assess the quality of the supported task; (ii) it is difficult to create a neutral experimental environment where there are no other factors influencing the application performance. [Sabou, *et al.*, 2006]
- 3) Data or corpus driven approaches. These evaluate the congruence of an ontology with a given corpus to determine how appropriate it is for the representation of knowledge of the domain represented by the texts [Brewster *et al.* 2004].
- 4) Assessment by humans to show how well the ontology meets a set of predefined criteria, standards, requirements as in OntoMetric [Lozano-Tello and Gomez-Perez 2004] and the peer-review based approach [Supekar 2005].

Table 1. A level based Taxonomy of Ontology Evaluation approaches [Brank *et al.*, 2005]

Evaluation level	Evaluation approach			
	Golden standard	Application based	Data or corpus driven	Human assessment
Lexical, vocabulary	X = applied	X	X	X
Hierarchy	X	X	X	X
Semantic relations	X	X	X	X
Content application	X	X	Not applied	X
Syntactic	X	Not applied	X	X
Structure, architecture, design	Not applied	Not applied	Not applied	X

Table 1 points to existing approaches being mainly for conducting technical evaluation of the structure of an ontology domain as represented by its taxonomy, using different levels. It indicates lack of a unifying approach for evaluating ontologies, creating an obstacle for their reuse [Alani and Brewster, 2006]. There is also lack of knowledge about properties users require to select a suitable ontology for a task [ibid]. The approaches in table 1 evaluate for the ontology structure (taxonomy) and not for biomedical processes and representation of granularity. With these approaches, evaluating biomedical ontologies remains challenging as they seek to integrate clinical and biological data across structure, processes and granularity so as to achieve interoperability between sources [Kumar *et al.*, 2006; Yugyung *et al.*, 2006].

3. ONTOLOGY EVALUATION TOOLS AND METRICS

The need for tools to evaluate ontologies before their reuse is justified by the expense, time and effort required to build new ones [Fernandez *et al.*, 2006]. Existing evaluation tools include AKTiveRank [Alani& Brewster, 2006], OntoQA [Tartir *et al.*, 2005], OntoMetric [Lozano-Tello & Gomez-Perez, 2004], ODEVal and OntoManager [Hartmann *et al.*, 2004]. They differ by scope, evaluation type, purpose, inputs, process, metrics used, outputs and the level of user involvement (table 2) as explained here.

3.1 OntoQA

The OntoQA tool measures the quality of an ontology using schema and instance metrics. The schema metrics of OntoQA address the design of the ontology schema while instance metrics measure the size and distribution of instance data [Tartir *et al.*, 2005]. Schema metrics are; relationships diversity (RD) and schema deepness (SD). RD is the ratio of the number of non-inheritance relationships, divided by the total number of relationships defined in the schema. It indicates the diversity of relationships in an ontology. SD is the average number of subclasses per class. It describes the distribution of classes across different levels of the ontology inheritance tree.

The instance metrics are of two types, class and relationship metrics. Class metrics are: class utilization (CU), cohesion (Coh), class instance distribution (CID), class connectivity (CC) and class importance (CI). CU is the ratio of the number of populated classes divided by the total number of classes defined in the ontology schema. It is used to indicate which of two populated ontologies has a richer KB. Coh of an ontology is the number of connected components of the graph representing the KB. It represents the number of connected components in the KB. CID is the standard deviation in the number of instances per class. It provides an indication on how instances are spread across the classes of the schema. CC indicates the centrality of a class. Formally CC is the total number of relationships instances the class as with instances of other classes. Class Importance is the number of instances that belong to the inheritance sub tree compared to the total number of class instances in the KB. It indicates where developers should focus on getting data if the intention is to get a consistent coverage of all classes in the schema.

OntoQA's relationship instance metrics are relational richness (RR) and importance (RI). RR indicates how the relationships defined for each class in the schema are being used at the instances level. It is number of relationships that are being used by instances that belong to an ontology compared to the number of relationships that are defined at the schema level. Relationship Importance measures the percentage of instances of a relationship with respect to the total number of relationship instances in the KB. This metric is important in that it helps in identifying which schema relationships were in focus when the instances were extracted and inform the user of the suitability for their intended use.

OntoQA enables knowledge engineers and researchers to find and analyze useful ontologies on the semantic web. It takes as input a crawled populated ontology or a set of user supplied search terms and ranks them according to some schema or instance level metrics related to various aspects of an ontology. A user is able to tune ontology ranking features to suit their needs. The tool has been validated by comparing its result to other approaches and to expert users [Tartir *et al.*, 2005].

3.2 AKTiveRank

AKTiveRank [Alani& Brewster,2006] is another tool proposed for ranking ontologies using metrics of the Class Match Measure (CMM), the Density measure (DM), Semantic similarity (SS) and Betweenness (BM) measures. The CMM assesses the coverage of an ontology for a given search term. The DM estimates information-content and level of knowledge detail of classes. The SS calculates how close the classes that match the search terms are in an ontology. The BM determines classes that are central to an

ontology. AKTiveRank uses as input search terms provided by a knowledge engineer and ontologies from a search engine (e.g. Swoogle). It ranks the ontologies using a number of classical metrics and compares the results with a questionnaire based human study. The results show that AKTiveRank has great utility although there is potential for improvement [ibid]. Its shortcoming is that it is that search terms can only be matched with ontology classes, and not with properties or comments [Alani & Brewster, 2006].

3.3 OntoMetric

OntoMetric [Lozano-Tello & Gomez-Perez, 2004] is a multi criteria decision making method that helps knowledge engineers to determine the suitability of a particular ontology for a project. A user selects an ontology using dimensions that specify the: 1) ontology content; 2) implementation language; 3) development methodology; 4) software used to build ontology; 5) costs of using the ontology in the system [Hartmann *et al.*, 2004]. OntoMetric is based on the analytic hierarchy process [Saaty, 1977], a selection process with four steps: 1) decide on selection criteria; 2) rate the relative importance of these criteria using pair-wise comparisons; 3) rate each potential choice relative to each other choice on the basis of each selection criterion, via pair wise comparisons of the choices; 4) combine ratings in steps 2 and 3 to get an overall rating for each potential choice. OntoMetric therefore offers the flexibility to select the hierarchy for the decision criteria to be used in evaluations. It however offers no specific support for evaluating integrated systems operating in dynamic environments like biomedicine.

3.4 ODEval

ODEval [Hartmann *et al.*, 2004] is used by ontology designers to evaluate knowledge representation for ontologies implemented in Semantic Web languages before they are used in applications. It detects possible knowledge representation inconsistency, incompleteness and redundancy in concepts for each considered language. ODEval has a high level of user involvement. It uses unpublished ontologies as input and generates ontology descriptions as output. No metrics are specified for ranking ontologies.

3.5. OntoManager

OntoManager is used by administrators, domain experts and business analysts to determine the truthfulness of an ontology with respect to its problem domain. It supports semi-automatic ontology improvement in response to the users' needs analysis. Its architecture has four functions of Monitor, Analyze, Plan and Execute. These monitor user interactions, analyze collected data, plan actions required for the changes discovered, and execute the changes to update the underlying ontology based application. It supports optimization of an ontology according to the users' needs. It is easy for end users to apply and use. However, evaluation quality is low. It is best applied in domains where information on ontology usage is available to identify relevant concepts of an ontology. The limitation on usage information does not allow to evaluate an ontology in general [Hartmann *et al.*, 2004] (Table 2)

Table 2. Comparison of some Ontology Evaluation tools

Criteria	OntoManager	OntoQA	AKTiveRank	OntoMetric
Users	Domain experts, admins., analysts	Knowledge engineers	Knowledge engineers [KE]	Knowledge engineers
Purpose	Fit ontology to requirements	Ontology ranking	Ontology ranking	Fit ontology to domain task
Evaluation context	Structure & user evaluation	Structure & knowledge base	Structure evaluation	Structure, cost, methodology, language, software
Input	Ontology in use	Search terms	Search terms	Ontology in use
Metric or criteria	User needs	Schema and instance metrics	Class density, semantic similarity, centrality measure	Multi-level tree [MTC] of 160 characteristics
Output	Ontology fitting requirements	Ranked ontology	Ranked ontology	Ontology fit to project
Validation type used	Not defined	Expert users	Questionnaire assessment	Not defined

Table 2 reveals that none of the current evaluation tools can be used across scope, in all evaluation contexts, inputs, metrics and outputs. However, as evident from the table, while the tools have been used to conduct structural and user evaluations, they do not define metrics useful for the dynamic processes and granularity presented by biomedical ontology. This lack of concern about dealing with heterogeneity including granularity has also been pointed out by Sabou, *et al.* [2006]. Furthermore, emphasis has been placed on defining structural metric on the basis of concepts while ignoring relations between concepts, a major limitation since relations provide valuable information in the search for the right ontologies, at the correct level of granularity [ibid].

4. AN EVALUATION TOOL FOR BIOMEDICAL ONTOLOGIES

In this paper we present a tool for evaluation of biomedical ontologies. Tool development is guided by:

- i. Existing literature on ontology evaluation frameworks [Maiga and Williams, 2008].
- ii. General requirements for ontology evaluation tools as given in the literature, revealed (table 2).
- iii. Specific requirements for biomedical ontology evaluation tools derived by a descriptive survey.

4.1. Specific Requirements for a biomedical Ontology Evaluation Tool

Literature review and document analysis were used to describe the theoretical requirements for a tool to help users assess and select a biomedical ontology suitable to their task. A questionnaire based descriptive survey was then used to validate the theoretical requirements with the proposed beneficiaries (biologists and medical doctors) of the tool. This objective guided the framing of the following key data collection questions on biomedical data integration model and evaluation tools namely: What is the scope (users and use cases) of a biomedical ontology evaluation tool? What general properties should the tool have? How should biomedical ontologies be compared using the tool? What are its inputs processes and outputs? How should such a tool be built?

The survey tested for agreement by respondents to characteristics of the proposed evaluation tool. Structured interviews and the questionnaire were pretested on twenty health care workers. Questionnaires were used to collect data and clarify requirements for the meta-model and tool. The survey then used questionnaires in which 630 randomly selected biomedical workers (580 medical doctors and 50 biologists) in Uganda were used as the study population. Filled questionnaires were returned by 404 medical doctors and 46 biologists. In the survey, these potential users of a biomedical integration system were asked for their level of agreement with proposed characteristics of the tool. The statistical package for social sciences (SPSS) was used to determine the level (%) of user agreement with the proposed characteristics of the tool. The results of the field study are presented in tables 3.

4.2. The Results

Tables 3 provide user perspectives on the requirements of a biomedical ontology evaluation tool.

Table 3 Requirements for the Evaluation Tool

Tool requirement	Molecular Biologists		Medical doctors	
	No. of responses	Agreement Level [%]	No. of responses	Agreement Level [%]
Ontology Visualization	46	91	404	70
Determine scope	46	85	404	73
Generate task requirements	46	90	404	89
Match requirements to ontology	46	68	404	72
Provide user feedback	46	90	404	91
Ease of use	46	94	404	95

The table reveals a high level of user preference for the following features and processes in a tool to evaluate a biomedical ontology: visualize an ontology, determine its scope, generate requirements for a task

and match them to an integration model, provide user feedback for re-specifying requirements and be easy to use. The requirements inform the principles adopted in the tool design.

4.3. A flexible User centered Evaluation Framework

The utility of using requirements in table 3 was demonstrated by a prototype tool developed as part of a flexible user centered framework for evaluating biomedical ontologies [Maiga and Williams, 2008]. The framework (figure. 1) provides different users with the flexibility to iteratively search through an ontology library using multiple criteria. This is more likely to result into the selection of a suitable ontology for a given task, or re-specification of new requirements for an ontology to fit the task. In the framework user needs motivate the assessment criteria in a formative evaluation, providing systematic feedback to designers and implementers in order to influence the process of developing new biomedical ontologies. Feedbacks in the framework provide a way to re-specify user requirements when extending, pruning and improve existing ontologies. This helps to avoid the huge effort of starting or building entirely new ontologies [Alani and Brewster 2006].

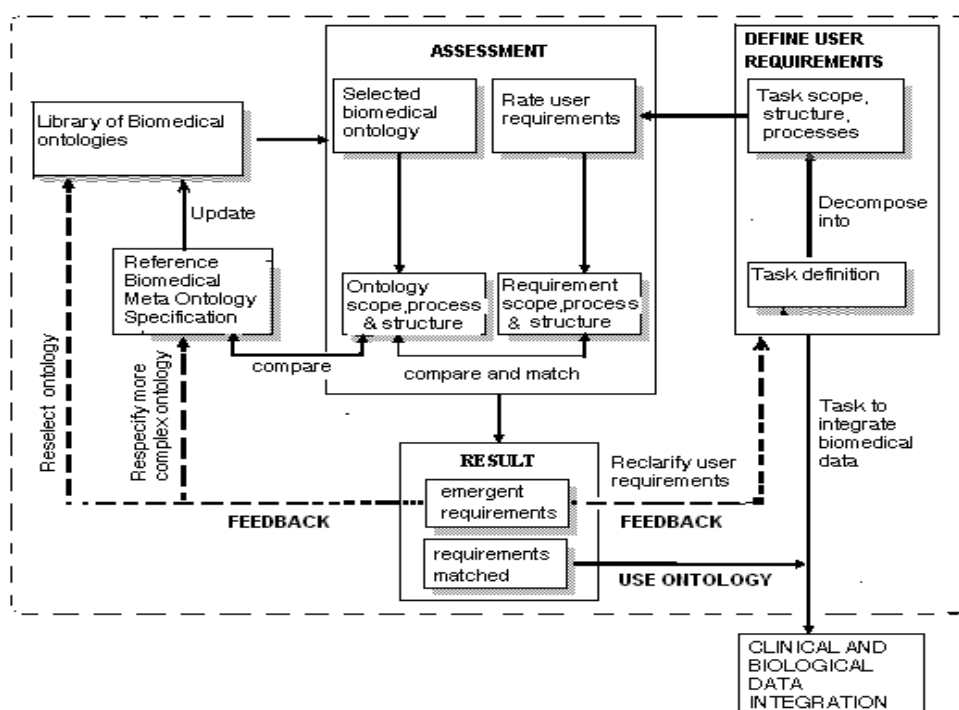


Figure 1. User Centered Framework to Biomedical Ontology Evaluation [Maiga & Williams, 2008].

4.4. Overview of the Evaluation Tool

The prototype tool described in this paper assumes a mixed context for assessing a biomedical ontology, applied in a flexible evaluation framework described in fig.2. It enables a user to select a biomedical ontology that supports development of a suitable application for their integration task. Scope, granular density and biomedical ontology structural integration are used as metrics for selecting an ontology.

4.5. Tool Architecture and Design

The design of the evaluation tool is guided by:

- i. Requirements for a biomedical evaluation tool [table 3]
- ii. Literature on the design of other tools for evaluating ontologies.

Existing ontology evaluation tools differ in their design [Alani& Brewster, 2006; Tartir *et al.*, 2005; Fernandez *et al.*, 2006]. They vary in the type and format of data inputs, processing and outputs generated as presented [table 1]. These differences mean that users have to learn to work with new tools,

data formats and outputs. The prototype tool design follows a process [activity] centered architecture guided by both requirements [table 3] and current knowledge on the design of ontology evaluation tools. The design therefore represents key requirements for such a tool as indicated in tables 1 and 3.

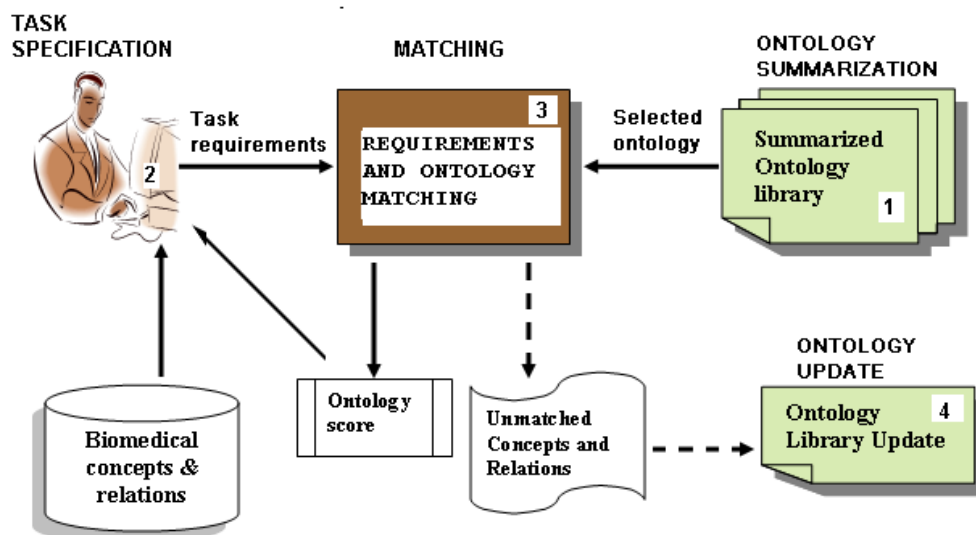


Fig. 2. Architecture of the Evaluation tool.

Tool design incrementally builds on existing tools in order to enable effective evaluation, promote user familiarity and system up-take by encouraging compatibility with previous data formats [Boardman, 2004]. An incremental approach that builds on existing tools design is advocated and used here because:

- 1) It builds on existing knowledge and technology, promotes user familiarity and compatibility with previous data formats, leading to a tool that is easy to use, a requirement for the evaluation tools.
- 2) It helps to achieve design goals that are implementable within the limited time and man-power resources in which research is conducted [ibid].

An incremental approach therefore enabled this study to build on previous design expertise manifested in already available tools. The tool is designed around the following core processes or activities of: 1) ontology summarization; 2) task determination; 3) matching and update, shown in the design.

Ontology summarization abstracts the top level classes and relations in an ontology. Concepts and relations representing the thematic categories of an ontology are extracted to enable users understand and select an ontology [Noy, 2004; Zhang *et al.*, 2007]. Summary concepts are used to build an ontology database to facilitate quick comprehension prior to its assessment and selection by non expert users.

The task specification enables a user to select or specify a biomedical data integration activity from a database. A task is a statement of the purpose for integrating biomedical data e.g. "a need to relate gene profile data to a disorder in the population." A task is decomposed into a set of requirements or search terms using objects and relations. The tool computes metrics for rating a biomedical ontology using a matching algorithm. User requirements are compared to an ontology selected from a library. The tool calculates and returns metrics and a description of any unmatched terms

4.6. The Tool's Matching Algorithm

The tool computes metrics for rating a biomedical ontology using a matching algorithm. User requirements are compared to an ontology selected from a library. The comparison process calculates and returns a metric and a description of any unmatched queries. On the basis of the returned result, a user can make a decision on whether to use the evaluated ontology, redefine the task or select another ontology. The UML activity diagram in Figure 3 is used to present the logic of an algorithm.

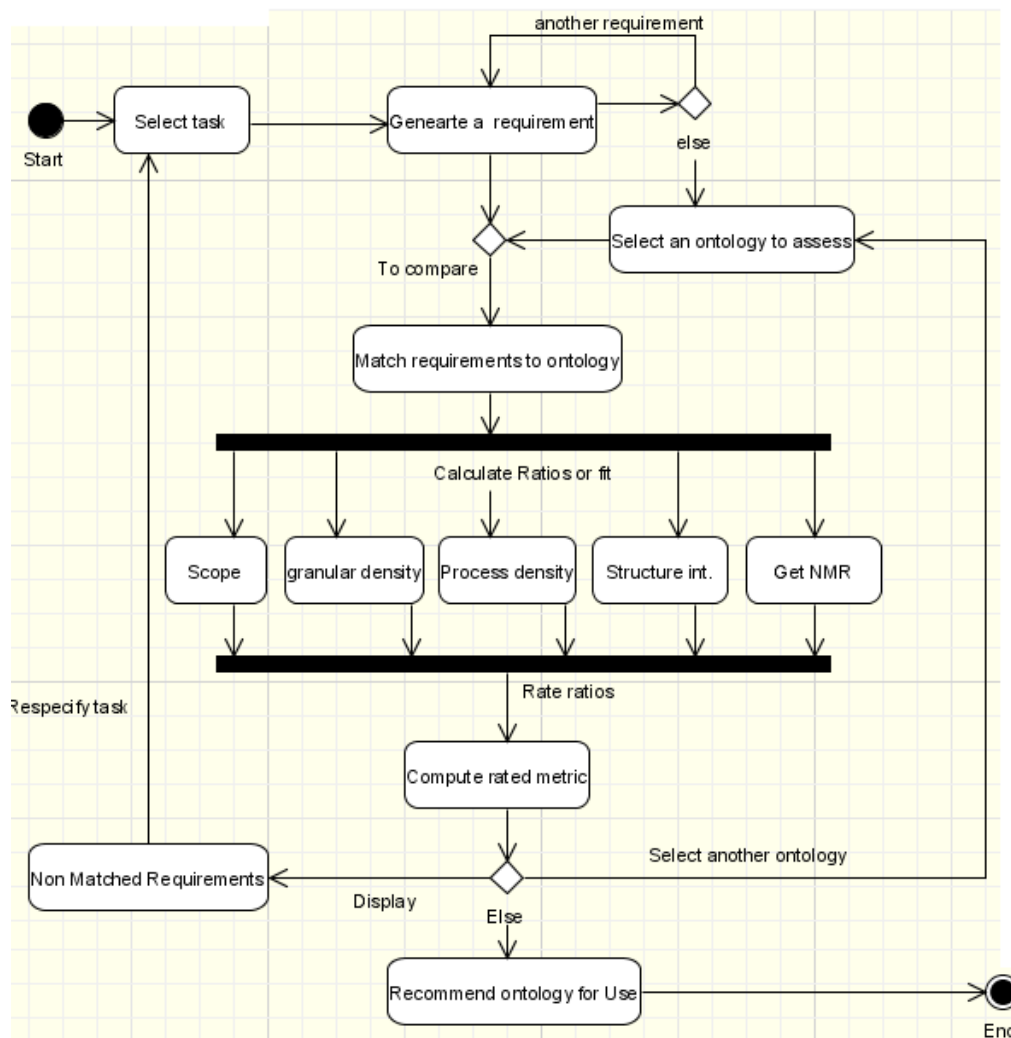


Figure 3. A UML Activity Diagram of the tool's Matching Algorithm

4.7. The Evaluation Metrics

In this paper, we describe three types of measures used to rate a biomedical ontology. These are scope, granular density and biomedical ontology integration.

4.7.1. Granular Density

Granularity, articulated as collectives linked by relations assumes a perspective that aggregates individuals into collections [Rector and Rodgers, 2005]. A related perspective articulates granularity based on categories of non-scale dependency (NSD), relationships between levels, set theory and mereology [Keet, 2006]. NSD granularity provides levels ordered through primitive relations like structural *part_of*, the spatial *contained_in* and the subsumption *is_a*. Logically, set theory also supports the *is_a* relation while mereology supports the *part_of* relation. Assuming multiple perspectives of granularity [as in this paper] that combine NSD granularity, relations between perspectives, set theory and mereology makes it possible to seamlessly shift from one granular perspective to another [ibid].

In a biomedical ontology, granularity is articulated using relations between classes like *part_of* or *contained_in*. The *part_of* relation is both an intra and trans ontological as it can be used to model relationships within and between biological and clinical classes respectively. The ratio of relations [relational richness] has been used as a structure based metric for comparing ontologies [Tatir *et al.*, 2005]. The perspective of granularity as collectives linked by relations is adopted here and used to define a metric for assessing the ability of the biomedical meta-ontology to represent granularity. Intra and trans-ontological relations in the meta ontology are used to define the metric. Trans ontological relations like

part_of, *has_part* and *contained-in* model collectives. The density of trans ontological relations [granular density] in a biomedical ontology is an indicator of its ability to model collectives. A metric for measuring the level to which an ontology models granularity can therefore be expressed by the ratio of trans ontological relations that model collectives to all trans ontological relations. This ratio or the granular density [G_D] is given by:

Let : R_{io} represent an intra ontological relation in the ontology; R_{ic} represent an intra ontological relation modeling collectives; R_{to} represent a trans ontological relation in the ontology, and R_{co} represent a trans ontological relation modeling collectives, then the Granular density G_D is given by the equation

$$G_D = \frac{[\sum R_{ic} + \sum R_{co}]}{[\sum R_{io} + \sum R_{to}]} \dots\dots\dots (1)$$

The metric provides a way to select between various ontologies for the task of integrating biomedical data across levels of granularity that are common in biomedicine. Good representation of granularity in a model enables tracking of entities and attributes across levels, leading to building of better data integration systems [Kumar *et al.*, 2006].

4.7.2. Biomedical Ontology Structure Integration

Clinical and biological classes and their relations are part of biomedical ontology structure. Between these classes, trans domain relations [e.g. *participates_in*] are identified. The term semantic density has been used to describe the level of connectedness of an ontology using relations [D'Aquin *et al.*, 2007]. The more trans ontological relations that are defined, the greater is the degree of connectedness in the ontology. The density of trans ontological relations between entities in an ontology is here used as an indicator of the level of integration [connectedness] and overlap between clinical and biological ontologies. A measure of the level of integration between clinical and biological data in an ontology can therefore be expressed as the proportion of trans ontological relations to all relations as defined in equation 2.

Let : R_{io} represent an intra ontological relation in the ontology; R_{to} represent a trans ontological relation in the ontology, then biomedical ontology structure integration (BOS_{int}) is given by

$$BOS_{int} = \frac{\sum R_{to}}{[\sum R_{io} + \sum R_{to}]} \dots\dots\dots (2)$$

This metric provides a user with way to compare biomedical ontologies and select between various alternatives, a suitable one for a biomedical data integration task.

4.7.3 Scope

Scope is a measure of how well an ontology's classes represent requirements for an integration task. For the tool, it is expressed as the number of objects [% search terms] in the requirements that are present or found in the ontology model. These scores are basis for a user to select the ontology, re-specify requirements or select another ontology to rate. The result also displays any unmatched requirements [search terms].

5. THE PROTOTYPE

The tool was implemented using visual studio 2005 integrated development environment, in C# and SqlExpress2005 database management system (DBMS). The database summarized the human phenotypic ontology (HPO) and the infectious disease ontology (IDO), selected from the open biomedical ontologies (OBO) library. The OBO library is a collection of standardized ontologies (60 plus) being reformed or developed according to a set of principles that represents biomedical reality in a way that enables them to interoperate in the task of integrating biomedical data [Smith *et al.*, 2007]. The emphasis on standardization informed selection of biomedical ontologies from the OBO library.

5.1. The Tool's User and Results Interfaces

The tool's user interface (figure 4.) has text boxes with drop down menus for selecting: 1) a biomedical integration task; 2) a biomedical object; 3) a clinical object; 4) a relation for the objects; 5) the ontology to be assessed. The interface also has panes for displaying requirements and the selected ontology. A user selects an integration task and its relevant biomedical objects and relations. The objects and

relations constitute requirements for the task and are ranked on a scale of one to ten before being displayed. A user also selects an ontology, displays and compares it to the requirements. The biomedical objects and relations are used as the search terms against which an ontology selected from the database is assessed. The tool returns metric scores and a description of any unmatched search terms (figure 4).

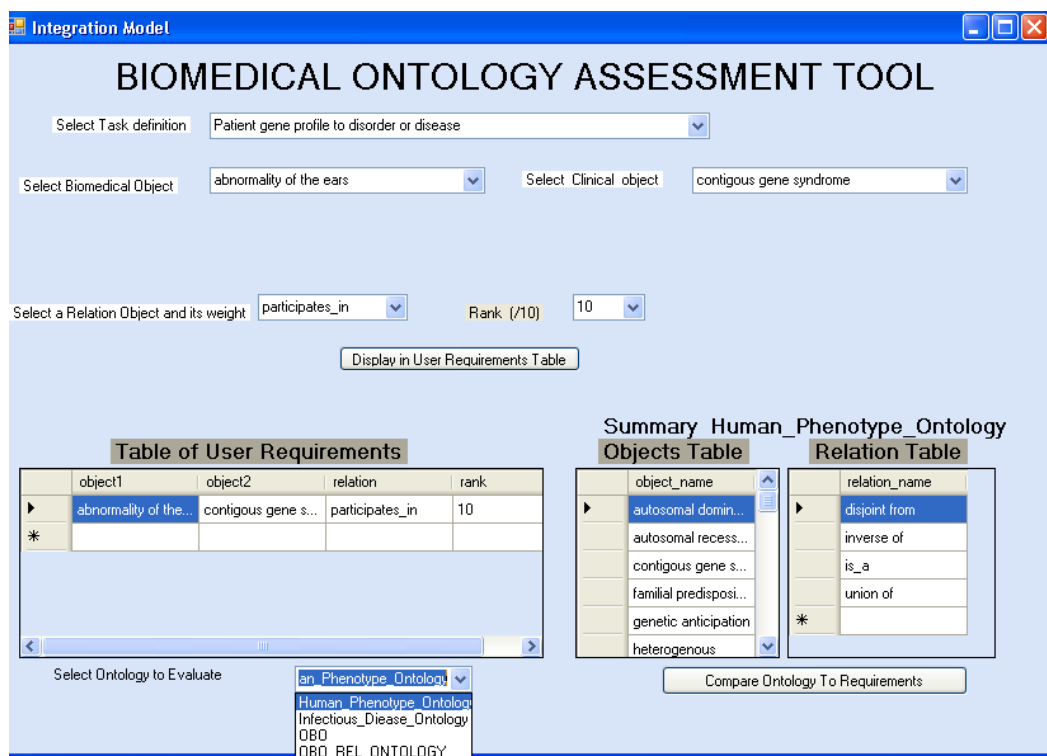


Figure 4. The Tool's User Interface

5.2. Evaluating the infectious disease Ontology

The tool was used to assess the infectious disease ontology [IDO], a set of interoperable ontologies that provide coverage of the infectious disease domain. IDO defines general entities relevant to biomedical and clinical aspects of infectious diseases. IDO is being built as a coordinated effort using best practices of the OBO consortium.

Experiments using the tool compared the ontology for their ability support different use case scenarios of biomedical data integration applications. For each use case, similar steps followed to assess an ontology are: 1) select a task; 2) generate requirements as a set of search terms; 3) rate the relative importance of requirements for the task; 4) display requirements; 5) select and visualize an ontology to be assessed; 6) compare search terms to selected ontology and generate any unmatched requirements. These may be used to re-specify a task, recommend ontology update or for use to build applications for the user's integration task or process.

Use case scenarios differed by the integration task selected, the requirements generated as a set of search terms and by the relative importance of these requirements or search terms for the users task. The output of using the tool to assess the IDO ontology for the different use case scenarios as illustrated in fig. 5 are presented in table 4. The results are displayed as scores of scope, granular density and biomedical integration.

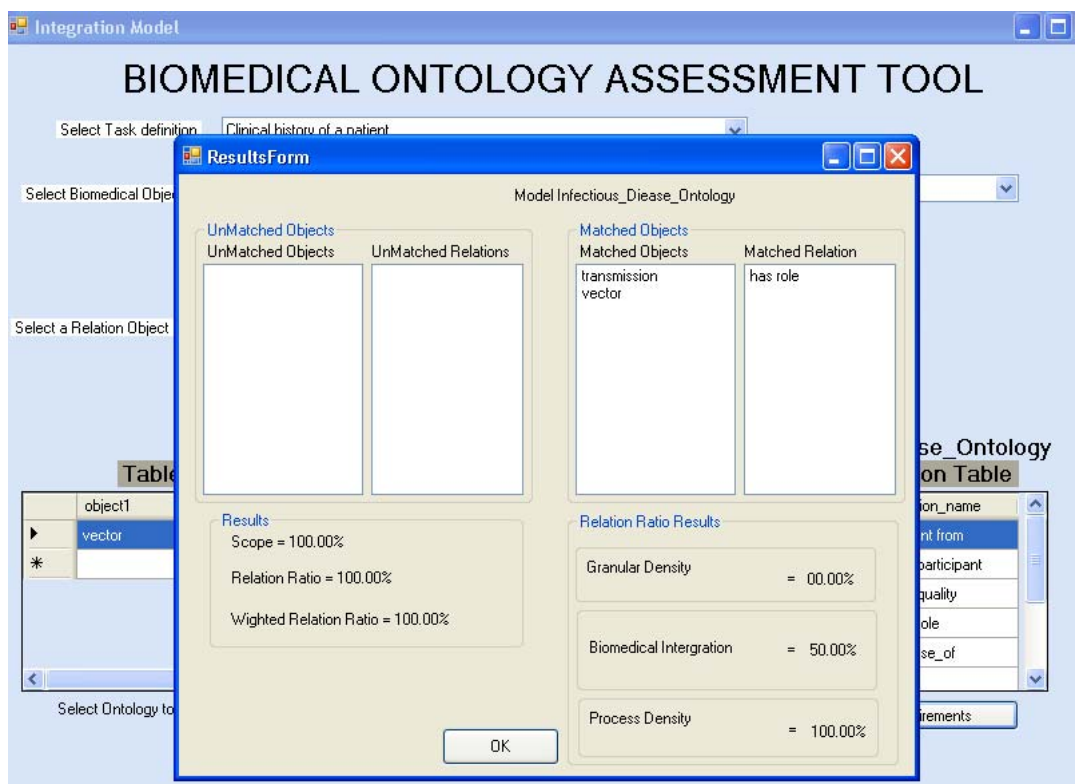


Figure 5 . The Tool's Results Interface

5.3. Tool Validation

Previous attempts to validate ontology evaluation tools have included: 1) comparing the results of assessing an ontology in experiments using the tool to those by other approaches as controls [Tartir *et al.*, 2005]; 2) comparing the results of assessing an ontology using the tool to human assessment by expert users, ontology consumers and domain experts [Tartir *et al.*, 2005; Alani& Brewster, 2006; Cross and Pal, 2006],

The first approach is ideal for mature ontologies that have previously been evaluated by other tools using similar or related metrics, with data available for comparison. For IDO, no such data was available. The results of assessing the IDO with the tool (TA) are therefore compared to those by a questionnaire based human assessment (HA) study using a Pearson correlation [r] for significance. A random selection of 32 biomedical workers (18 medical doctors and 14 biologists) was used for this test. The questionnaire had a screen shot of the IDO model, two tables with six use case scenarios for applying the model and instructions on how to fill in the tables. For each scenario (search terms and relation), a question was asked about the models (IDO) ability to represent scope, biomedical relation and granularity. The answers to these questions were used to calculate the scope, granular density and biomedical integration ratios. The tool results (TA) of assessing IDO and corresponding ones from a questionnaire based human assessment [HA] for the different use case scenarios (expressed as percentages) are presented in table 4. The corresponding Pearson correlation coefficients are given in table 5.

Table 4 Tool Assessment (TA) and Human Assessment (HA) of different use case scenarios

		Scope		Granularity		Biomed	
		TA	HA	TA	HA	TA	HA
1	Vector <i>has _role</i> in transmission mode of disease	100	100	00	0	50	0
2	y-linked <i>has participant</i> abnormality incubation period <i>part of</i> progression rate	50	50	100	50	67	50
3	y-linked <i>has participant</i> abnormality infectious disease course <i>has _quality</i> phenotypic variability	25	15	50	0	67.7	50
4	Host <i>has participant</i> dormancy period	100	75	100	0	50	45
5	Host <i>has _quality</i> familial predisposition	50	55	00	0	50	0
6	Vector <i>is_a</i> transmission mode of disease	100	100	100	100	50	50

5.8. The Pearson Correlation

Table 5. Pearson Correlation [r] Comparison of Tool Scores against Human Assessments

Metric scored	Pearson correlation
Scope	0.945
Granular Density	0.608
Biomedical Ontology Integration	0.576

The r values from the tool and corresponding ones from the questionnaire based human assessment are not matched exactly. They however show a moderate to strong positive correlation between the two sets [$r = 0$ means zero correlation; $r = 1$ means strong positive correlation; $r = -1$ means strong negative correlation]. The scope ration is more highly correlated with biomedical ontology integration being the least.

The tool and human assessment had exact matches of scope, granularity and biomedical integration for the use case scenario of the *is_a* relationship. The tool and human assessment also had exact or near exact matches of scope and granularity for the use case scenarios of the *has_role*, *has_part* and *has_quality* relations, an indication of the success of the tool used, and the validity of our approach.

6. DISCUSSION AND CONCLUSIONS

The evaluation tool presented in this paper provides users with the flexibility to select a suitable ontology for use in building applications that integrate clinical and biological data. Given the large numbers of biomedical ontologies available in libraries, this tool can save time for both naïve users and expert ontologists by enabling them to quickly assess and select an ontology from a large collection (of summarized ontologies) before further evaluation of its taxonomy.

The tool can also enable ontology modelers to quickly gather new requirements for improving on existing biomedical ontologies, leading to new biomedical ontologies that are better able to integrate data across structure, function, processes and granularity. This is likely to contribute towards solving the problem of the persistent lack of a generic all purpose methodology for integrating biomedical data, as earlier recognized by Grenon *et.al.* [2004], leading to their re-use and adoption.

The evaluation process has some limitations. Human assessment of the infectious disease ontology presented some difficulties that are associated with ontology evaluation by human subject. Respondents were only exposed to screen shots of the IDO without sufficient accompanying documentation about it, which may have affected the quality of the response and choices made on the questionnaire. The respondents [medical doctors and biologists] are not ontologists and reported difficulty when interpreting and reasoning about the different use case scenarios of an ontology, even with detailed and clear instructions.

7. ACKNOWLEDGEMENTS

We wish to thank Makerere University's school of Graduate studies and the Faculty of Computing and Information technology for sponsoring this work. Special thanks to Dr Mathias Brochhausen and staff of the Institute for Formal Ontology and Medical Information Science, Saarland University, Germany, for hosting us and making the facilities of the institute for formal Ontology and medical information Science [IFOMIS] available to us when they were required.

8. REFERENCES

- ALANI, H. & BREWSTER, C. 2006. Metrics for Ranking Ontologies. In *Proceedings of the 4th International Conference on Evaluation of Ontologies for the Web*. EON Workshop Edinburgh International Conference Center. Edinburgh UK. May 22nd 2006.
- BOARDMAN, R. 2004. Improving Tool Support for Personal Information Management. PhD Thesis.
- BRANK, J., GROBENLINK, M. AND MLADENIC, D. 2005. A Survey of Ontology Evaluation Techniques. *At the conference on Data Mining and Data Warehouses* SiKDD.
- BREWSTER, C., ALANI, H., DASMAHAPATRA, S., AND WILKA, Y. 2004. Data Driven Ontology Evaluation. *Proceedings of the Language Resources and Evaluation Conference LREC 2004*, Lisbon, Portugal.
- CROSS, V. AND PAL, A. ONTOCAT: 2006. An Ontology Consumer Analysis Tool and Its Use on Product Services Categorization Standards. Available: <http://protege.stanford.edu/conference>
- D'AQUIN, M. SABOU, M. DZBOR, C. BALDASSARRE, L. GRIDINOC, S. ANGELETOU, & E. MOTTA. 2007. WATSON: A Gateway for the next Generation Semantic Web Applications. *ESWC 2007*
- FERNANDEZ, M., CANTADOR, I., & P CASTELLS. 2006. CORE: A Tool for Collaborative Ontology Reuse and Evaluation.
- GANGEMI, A., CATENACCI, C., CIARAMITA, M. & LEHMANN, J. 2005. A theoretical framework for ontology evaluation and validation. *In Proceedings of SWAP 2005*.
- Gomez-Perez, A. 1994. Some ideas and Examples to Evaluate Ontologies. Technical Report KSL-94-65. Knowledge System Laboratory. Stanford University. Also in Proceedings of the 11 the Conference on Artificial Intelligence for Applications, CAIA94.
- GUARINO, N. & WELTY, C. 2002. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*. 45, No. 2.
- GÓMEZ-PÉREZ, A. 2004. Ontology Evaluation. *Handbook on Ontologies*. 251-274.
- GRENON, P., SMITH, B., AND GOLGBERG, L. 2004. Biodynamic Ontology: Applying BFO in the Biomedical Domain. *From D. M. Pisanelli ed., Ontologies in Medicine, Amsterdam: IOS Press*, 20-38
- HARTMANN, J., SPYNS, P., GIBOIN, A., MAYNARD, D., CUEL, R., SUÁREZ-FIGUEROA, M. C. & SURE, Y. 2005. Methods for ontology evaluation. EU-IST Network of Excellence NoE IST-2004-507482 KWEB.
- KALFOGLOU, Y. & SCHORLMER, M. 2006. Issues with evaluating and using publicly available ontologies. In *Proceedings 4th International EON Workshop, Evaluating Ontologies for the Web*, Edinburgh, Scotland.
- KEET, C.M. 2006. A Taxonomy of types of granularity. *IEEE International Conference on Granular Computing GrC*.
- KUMAR, A., LINAYIP, Y., SMITH, B. & GRENON, P. 2006. Bridging the gap between medical and bioinformatics: An ontological case study in colon carcinoma. *Computers in Biology and Medicine* 36, 694-711
- LAMBRIX, P. & TAN, H. 2006. Sambo. A system for aligning and merging biomedical ontologies *Web Semantics: Science, Services and Agents on the World Wide Web*, 4 196 – 206
- LOZANO-TELLO, A., GOMEZ-PEREZ, A. AND E. SOSA 2004. Selection of Ontologies for the Semantic Web.
- MAIGA, G. & WILLIAMS, D. 2008. A Flexible Approach for User Evaluation of Biomedical Ontologies. *International Journal of Computing and ICT Research*. ISSN 1996-1065. Volume 2 Dec. 2008, 62-74; Available: <http://www.ijcir.org/volume2-number2/index.php>
- NOY F. & MUSEN, M.A. 2002. Evaluating Ontology-Mapping Tools: Requirements and Experience. Available: <http://seer.nj.nec.com>. Accessed 6 may 2006.

- PORZEL., R. AND R. MALAKA. 2004. A task-based approach for ontology evaluation. Proc. of the 16th European Conf. on Artificial Intelligence (ECAI 2004), Valencia, Spain, August 2004.
- RECTOR. A., AND ROGERS, J. 2005. Granularity Scale & Collectivity: When size does and doesn't matter.
- SAATY, T. 1977. "A Scaling Method for Priorities in Hierarchical Structures". *Journal of Mathematical Psychology*, Vol.15, (1977) 234-281.
- SABOU, M., LOPEZ, V, MOTTA, E. AND UREN, V.(2006. Ontology Selection: Evaluation on the Real Semantic Web, *Fourth International Evaluation of Ontologies for the Web Workshop (EON2006)*, Edinburgh, UK, May (2006).
- SMITH, B., CEUSTERS, W., KLAGGES, B., KÖHLER, J. KUMAR, A.,LOMAX, J., MUNGALL, C., NEUHAUS, F., RECTOR, AL., ROSSE, C., 2005. Relations in Biomedical Ontologies. *Genome Biology*.
- SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, JL., EILBECK, K., IRELAND, A., MUNGALL, C.J., LEONTIS, N., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S., SCHEUERMANN, R.H., SHAH, N., WHETZEL, P.L., & LEWIS, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, Vol 25 (11).
- SUPEKAR, K. 2005. A peer-review approach for ontology evaluation. *Proc.8th Intl. Protégé Conference*, Madrid, Spain, July 18–21, 2005.
- TARTIR,S., ARPINAR B., MOORE, M., SHETH, A.AND ALEMAN-MEZA, B. (2005).OntoQA: Metric-Based Ontology Quality Analysis. In *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, Houston, TX, USA, IEEE Computer Society (2005) 45 -53
- ZHANG, H., CHENG, G., AND QU YUZHONG.2007. Ontology summarization based on rdf sentence graph. In *lproceedings of the 16th international conference on World Wide Web*. Banff, Alberta, Canada