Monte Carlo Study of Taxonomy Evaluation

Alexander Ulanov, Georgy Shevlyakov, Nikolay Lyubomishchenko, Pankaj Mehra, Vladimir Polutin

HP Laboratories
HPL-2010-147

Keyword(s):

Taxonomy evaluation, taxonomy benchmarks

Abstract:

The problems of taxonomy evaluation criteria comparison and corresponding benchmark creation are considered. The classes of Primitive Ideal Taxonomies (PITs), their WordNet and disrupted versions are proposed as the sets of benchmark taxonomies for the comparison of taxonomy evaluation methods. For WordNet PITs and their perturbations, the performances of the structure-based PageRank, FloorRank, and the corpusbased Information Content criteria are studied in Monte Carlo experiment. It is shown that the proposed approach can be used for the ranking of taxonomy evaluation criteria.

External Posting Date: October 6, 2010 [Fulltext] Approved for External Publication
Internal Posting Date: October 6, 2010 [Fulltext]
To be published and presented at WeBS 2010, Bilbao, Spain.

Monte Carlo Study of Taxonomy Evaluation

Alexander Ulanov*, Georgy Shevlyakov†, Nikolay Lyubomishchenko*, Pankaj Mehra‡ and Vladimir Polutin*

*HP Labs Russia, Saint-Petersburg, Russia
Email: {alexander.ulanov,nikolay.lyubomishchenko,vladimir.polutin}@hp.com
†Department of Applied Mathematics, Saint-Petersburg State Polytechnic University, Saint-Petersburg, Russia
Email: {Georgy.Shevlyakov,nlyubomishchenko}@gmail.com
‡Inlogy Inc., Los-Altos, CA, USA
Email: {pankaj.mehra}@sbcglobal.com

***Abstract*—The problems of taxonomy evaluation criteria comparison and corresponding benchmark creation are considered. The classes of Primitive Ideal Taxonomies (PITs), their WordNet and disrupted versions are proposed as the sets of benchmark taxonomies for the comparison of taxonomy evaluation methods. For WordNet PITs and their perturbations, the performances of the structure-based PageRank, FloorRank, and the corpus-based Information Content criteria are studied in Monte Carlo experiment. It is shown that the proposed approach can be used for the ranking of taxonomy evaluation criteria.**

***Index Terms*—taxonomy evaluation; taxonomy benchmarks.**

I. Introduction

Ontologies are increasingly used in various fields such as knowledge management, information extraction, and the semantic web. However, it is useful to know the quality of a particular ontology before deployment, especially in the case when there are numbers of similar ones. Ontology evaluation is the problem of assessing a given ontology from the point of view of a particular criterion of application, typically in order to determine which of several ontologies would better suit a particular purpose. An ontology contains both taxonomic and factual information. Taxonomic information includes information about concepts and their association usually organized into a hierarchical structure. This paper addresses the evaluation of such taxonomies.

Taxonomy evaluation is based on measures and methods to examine a set of criteria [1]. The approaches range spreads from simple golden-standard and structure to the more complicated corpus and task based [2], [3]. Complex methodologies include different approaches as far as expert assessment. Golden-standard is more an approach to evaluation of taxonomy extraction methods because we need to choose a good taxonomy. Structure based approach tries to relate structural parameters of a taxonomy graph with some semantical criteria. Corpus based approach studies the cohesion to a domain represented in the vocabulary or document collection. Task-based approach tries to model a taxonomy in the settings close to the real deployment. However, in this case it is not very clear what is evaluated - the taxonomy itself or how the embedded method of retrieval or classification works.

Each approach has its own advantages and drawbacks and there is a need for common evaluation of them. In this work, we will study the approaches to taxonomy evaluation, discuss their challenges and make an attempt to create a common benchmark for their evaluation.

The remainder of the paper is as follows. In Section II, we present a survey of the state of the art in ontology evaluation according to the proposed classification. In Section III, the problem of taxonomy evaluation and the corresponding challenges are described. In Section IV, the proposed approach to taxonomy evaluation is given. In Section V, the proposed set of benchmarks and the general description of experiments are presented. In Section VI, the performance evaluation results are given. In Section VII, conclusions are drawn, recommendations proposed and further work highlighted.

II. State of the Art

Since the number of ontologies in the web increases there are plenty of taxonomy and ontology evaluation methods. Except simple syntax checking, existing ontology evaluation methods are based on the following approaches [2], [3]: (i) golden-standard, (ii) structure-based, (iii) corpus-based, (iv) task-based, and (v) complex and expert based.

The golden-standard is the most straight-forward approach. In [4], it is described how to measure the similarity between ontologies using overlap of relations. In [5], a taxonomy from Wikipedia categories structure is derived showing that it can be used for such tasks as word similarity measurement, sense disambiguation etc.

Structure-based approaches assume computing various structure properties of the ontology such as relationship and attribute richness, number of nodes, connectivity, cohesion etc [6]. Studies [3] and [6] show that ”best” ontologies intended for browsing are the most populated with a lot of links. There are studies dedicated to building hierarchies from the network of related notions. The fit between the hierarchy constructed from Wikipedia articles and its category structure is measured using structural approaches, e.g., PageRank, betweenness centrality etc. [7], [8].

The corpus-based approach supposes that there is some corpus against which one can evaluate such taxonomy properties as correctness, completeness etc. [9]. In [10], it is proposed to map the set of terms from a corpus to the ontology. The authors raise an important issue of structural fit, i.e., a measure how the corpus clusters represent an ontology structure. A corpus for evaluation can be derived from Google query results [11].

The task-based approach supposes that ontology is intended for some task and its performance for this task is evaluated. The following two issues are usually addressed: how good is the given ontology for aiding information retrieval and how easy the needful information can be retrieved from the ontology. Study [2] addresses search task, [13] and [14] discuss classification potential of taxonomy. Ease of navigation inside taxonomy are addressed in [6] and [12].

There are a lot of complex ontology evaluation measures that try to incorporate many aspects of ontology quality: AKTiveRank [15], Ontometric [16], OntoClean [17], OntoQA [18]. AKTiveRank includes several measures of ontology evaluation, e.g., class match, density, betweenness etc. Ontometric is based on the notion of multilevel tree of characteristics with scores, which includes design qualities, cost, tools, language characteristics. OntoClean evaluates ontologies on the basis of correctness of the classes and their properties. OntoQA framework allows evaluating ontology design and usage.

III. The Goals of Study

The great variety of taxonomy evaluation criteria together with the accompanying lack of the objective comparisons of their performances causes a definite uncertainty with the choice of an appropriate criterion in the problem of taxonomy evaluation. Thus, even an approximate ranking of taxonomy evaluation criteria with the specification of their potential areas of applicability is a very important task from both theoretical and practical points of view. This problem is naturally connected with the problem of forming the set of taxonomy benchmarks using which the comparisons of taxonomy evaluation criteria could be made.

At present, the main method of comparisons of a newly proposed taxonomy evaluation criterion is, in its essence, *ad hoc* application oriented: the performance evaluation is made on the basis of the data corpus and the related taxonomies used in a particular application. In this case, the obtained results can hardly be extended onto the other applications. Thus, the problem of creation of the set of taxonomy benchmarks suitable for the full range of taxonomy evaluation criteria is very important.

The main goals of our study are as follows: (i) to propose a low-complexity method aimed at ranking taxonomy evaluation criteria, (ii) to elaborate a set of taxonomy benchmarks, (iii) to study the performance of the proposed taxonomy evaluation ranking criterion over the set of taxonomy benchmarks.

IV. Taxonomy Evaluation Methods

A. Preliminaries

In Section II, it is admitted that there exist a lot of taxonomy evaluation methods, and it is generally unclear which method to use since there is no definite order among them. In this study, our goal is not to propose yet another taxonomy evaluation method to the existing list of methods, but to find an approach to their general ranking.

Any conceivable ranking of a given set of taxonomy evaluation methods is a function of a chosen collection of taxonomy benchmarks, hence the problems of taxonomy ranking and of the creation of taxonomy benchmarks are interdependent. The golden-standard approach for taxonomy ranking seems to be quite a natural choice [4], [5], since it is based on the expert work in a particular area of application. Unfortunately, expert work is at the same time very valuable and very expensive, and thus the creation of a golden-standard set of taxonomies is not always viable. However, we can try to find simpler general analogs of golden-standard taxonomies suitable for wide spectra of applications.

B. Primitive Ideal Taxonomies (PITs)

In this paper, we propose to use the ”primitive ideal taxonomies” (PITs) which are nothing but trivial trees either of a dichotomy type, or of a trichotomy type, or of a more complicated structure, as simple aforementioned analogs of golden-standard taxonomies.

The essential advantage of PITs is that all their nodes have obvious (ideal) ranks (not necessarily different), and this fact can be used for constructing a taxonomy ranking measure. Since the application of practically all known taxonomy evaluation criteria involve computation of taxonomy node ranks, we can compare the ideal PIT ranks and the actual PIT ranks obtained by the application of a chosen taxonomy evaluation criterion.

The natural measure of this comparison can be given by the correlation between the ranks, namely, by the Spearman rank correlation coefficient r_S. We may expect that for reasonable taxonomy evaluation criteria the r_S correlations will be close to unit.

C. Monte Carlo Approach to Semantic Studies

Next, we enlist several other convenient features of the PIT model: its structure can be easily adapted to real-life taxonomies by choosing the proper values of the taxonomy depth, of the minimum, maximum and average number of children for a node, of the distance of a node from the taxonomy root, etc. All those taxonomy parameters can be estimated from the real-life taxonomies and used to model randomized PITs. Thus, the Monte Carlo method widely used in statistics can be naturally extended onto the study of taxonomy evaluation criteria performance.

D. Taxonomy Evaluation Criteria for Comparative Study

In our study, we restrict ourselves to structure and corpus based criteria

- The structure based PageRank criterion being a link analysis algorithm, named after Larry Page [7], used by the Google Internet search engine that assigns a

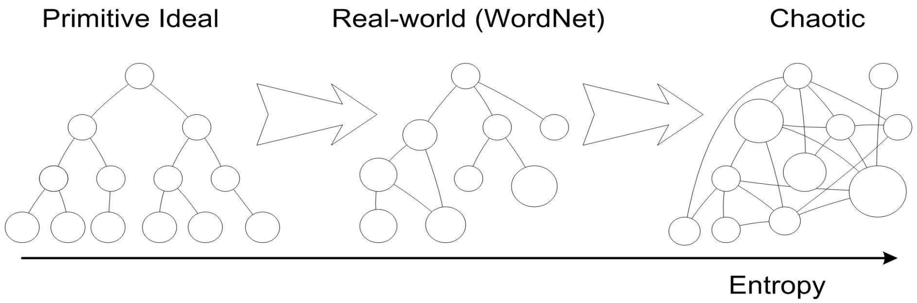

Fig. 1. Taxonomy scale

numerical weighting to each element of a hyperlinked set of documents

$$PR(v_i) = (1-d)/N + d \sum_{v_j \in M(v_i)} PR(v_j) \Big/ L(v_j),$$

where $v_1, \ldots, v_N$ are the vertexes (taxonomy nodes) under consideration; $M(v_i)$ is the set of nodes that link to v_i; $L(v_j)$ is the number of outbound links from the node v_j, and N is the total number of nodes; d is a damping factor. Various studies have tested different damping factors, but it is generally assumed that the damping factor should be set around 0.85 [7].

- The structure based FloorRank - ranking by ”floors”: the ”floor” is a set of nodes with the same distance from the taxonomy root; the numbers of the floors at which a node is located are assigned to it. It is computed by the breadth-first search (BFS) [19]. The FloorRank is an average floor number for a given node

$$FR(v_i) = \sum_{floors(v_i)} floor(v_i) \Big/ floors\ count(v_i).$$

- The corpus based Information Content (IC) criterion [20] defined as

$$IC(v_i) = -\log p_i,$$

where p_i is the probability of node v_i. Generally, the values of IC are taken from special IC-corpuses. In the case of PITs, initially all the leaves are equiprobable: the leaf probability $= 1/(leaf\, count)$ - unit divided by the total number of leaves; the node probability $= \sum(node\, child\, probability)$ - the sum of all the node children probabilities for a given node.

V. Taxonomy Evaluation Benchmarks

A. Preliminaries

In this section, in order to form a set of taxonomy benchmarks, we develop the proposed concept of Primitive Ideal Taxonomies. Initially, PITs are quite regular taxonomies with a simple and clear structure, generally, being very far from real-life taxonomies. Therefore, our main goal is to transform their structure making them closer to real targets of taxonomy. This can be achieved by many ways, say, we may transform them either in a deterministic or in a random way changing taxonomy parameters, allowing undesirable connections, inserting extra nodes with cycles, increasing the perturbations scale, etc.

Next, we may imagine the following taxonomy scale: on its left-hand side, we locate regular PITs of minimum entropy, on its right-hand side, we put fully chaotic and thus senseless taxonomies of maximum entropy, and we may reserve the middle zone for real-life taxonomies as in Fig. 1 with circle sizes proportional to node Information Contents.

Applying different taxonomy evaluation criteria to such a range of taxonomies, we may expect different sensitivity of those criteria, firstly, to qualitatively different taxonomy perturbations, and, secondly, to quantitatively different perturbation scales of the same type. These general considerations are basic for our further studies. In what follows, we use only a few possibilities within the proposed scheme.

B. Using WordNet as Golden-Standard

To make PITs closer to real-life taxonomies, we set their parameters similar to the average parameters of taxonomies generated by the WordNet [21].

The direct way of generating ”real” PITs is to use the WordNet sub-taxonomies parameters in the corresponding algorithms. As a result of the statistical analysis of WordNet sub-taxonomies, the following taxonomy parameter estimates were obtained: the median taxonomy depth is equal to 7; the number of node children is distributed according to the heavy-tailed Pareto law with the numbers from 2 to 5. We used these results to generate random PITs, so now we call them as WordNet PITs.

The next step in creation of a set of taxonomy benchmarks is to perturb the WordNet PITs changing their ”ideal” structures. Here, we use the two types of changes, ”in-floor” and ”inter-floor” perturbations.

In our experiment, we set the maximal taxonomy depth (the number of floors) equal to 7: this choice provides the triangle-shape structure of taxonomies with increasing numbers of leafs as in PITs. However, real WordNet sub-taxonomies may not necessarily have this shape; the analysis of the other types of taxonomy shapes deserves a separate study.

Next, we specify the aforementioned types of perturbations with the fraction of perturbed nodes set to 50%:

• the ”in-floor” (horizontal) perturbations: inserting the additional connections (edges) between the nodes lying at the same distance from a taxonomy root (at the same floor) with the random choice of the perturbed floors and nodes;
• the ”inter-floor” (vertical) perturbations: inserting the additional connections (edges) between the nodes lying at the same taxonomy ”branch” (a taxonomy ”branch” is a sequence of nodes from the root to a leaf) with the random choice of the perturbed branches and nodes.

VI. Performance Evaluation

A. Description of Monte Carlo Experiment

In our research, we use five different types of test taxonomies of depth 7: the initial WordNet PIT and the following four perturbed versions:

- Perturbation 1 – the horizontal perturbation of 50$\%$ of nodes up to the 50$\%$ of floors (down to 4th floor counting from the root) ($HP/50/4$);
- Perturbation 2 – the horizontal perturbation of 50$\%$ of nodes on all floors ($HP/50/7$);
- Perturbation 3 – the vertical perturbation of 50$\%$ of nodes up to the 50$\%$ of floors ($VP/50/4$);
- Perturbation 4 – the vertical perturbation of 50$\%$ of nodes on all floors ($VP/50/7$).

All those taxonomy benchmarks are obtained by Monte Carlo method with the parameters estimated from the WordNet statistics: in our study, we use the data of 100 random trials for each taxonomy type.

B. Comparative Study of Taxonomy Evaluation Criteria

For the comparative study of taxonomy evaluation criteria, we take the structure-based FloorRank and PageRank, and the corpus-based Information-Content method.

First, we compute the rank correlations between the ideal ranks of the initial WordNet PIT and the ranks of all its perturbations obtained by the aforementioned criteria. These Monte Carlo results are exhibited in Tables I–III. Second, we compute the pair-wise correlations between the ranks obtained by the chosen criteria for all the nine types of WordNet PITs. Those results are displayed in Tables IV–VI.

To analyze the obtained data, the following sample statistics of the Spearman rank correlation are computed: the mean Ave, the lower quartile LQ, the median Med, the upper quartile UQ, and the standard deviation S.

In general, the data represented in Tables I–VI show that the obtained correlations are grouping rather tightly – their standard errors can be roughly estimated by dividing their sample standard deviations by $\sqrt{100}$ (for 100 trials), so that the standard errors are approximately of order 10^{-3}.

C. Analysis of Experimental Results

The obtained rank correlations exhibited in Tables I-III can be considered from the two complementary points of view:

- For a given criterion, the obtained rank correlation yields the measure of disorder of the perturbed WordNet PITs — the smaller correlations, the greater disorder (entropy).

TABLE I
FloorRank Correlation

	Ave	LQ	Med	UQ	S
Perturbation 1 (HP/50/4)	0.96	0.95	0.97	0.98	0.02
Perturbation 2 (HP/50/7)	0.79	0.77	0.79	0.81	0.03
Perturbation 3 (VP/50/4)	0.99	0.99	0.99	0.99	0.00
Perturbation 4 (VP/50/7)	0.98	0.98	0.98	0.98	0.00

TABLE II
PageRank Correlation

	Ave	LQ	Med	UQ	S
Perturbation 1 (HP/50/4)	0.82	0.79	0.82	0.84	0.03
Perturbation 2 (HP/50/7)	0.77	0.76	0.77	0.79	0.02
Perturbation 3 (VP/50/4)	0.99	0.98	0.99	0.99	0.01
Perturbation 4 (VP/50/7)	0.92	0.92	0.92	0.93	0.01

TABLE III
Information Content Correlation

	Ave	LQ	Med	UQ	S
Perturbation 1 (HP/50/4)	0.99	0.99	0.99	0.99	0.00
Perturbation 2 (HP/50/7)	0.96	0.96	0.96	0.97	0.01
Perturbation 3 (VP/50/4)	0.99	0.99	0.99	0.99	0.00
Perturbation 4 (VP/50/7)	0.99	0.99	0.99	0.99	0.00

TABLE IV
FloorRank versus PageRank Correlation

	Ave	LQ	Med	UQ	S
Initial WordNet PIT	0.98	0.98	0.98	0.98	0.00
Perturbation 1 (HP/50/4)	0.82	0.80	0.82	0.84	0.04
Perturbation 2 (HP/50/7)	0.82	0.81	0.82	0.84	0.03
Perturbation 3 (VP/50/4)	0.98	0.97	0.98	0.98	0.01
Perturbation 4 (VP/50/7)	0.93	0.92	0.93	0.93	0.01

TABLE V
FloorRank versus Information Content Correlation

	Ave	LQ	Med	UQ	S
Initial WordNet PIT	0.69	0.67	0.69	0.71	0.02
Perturbation 1 (HP/50/4)	0.88	0.86	0.88	0.91	0.04
Perturbation 2 (HP/50/7)	0.77	0.74	0.78	0.80	0.04
Perturbation 3 (VP/50/4)	0.69	0.68	0.70	0.71	0.02
Perturbation 4 (VP/50/7)	0.85	0.84	0.85	0.86	0.02

TABLE VI
PageRank versus Information Content Correlation

	Ave	LQ	Med	UQ	S
Initial WordNet PIT	0.68	0.67	0.68	0.71	0.03
Perturbation 1 (HP/50/4)	0.63	0.61	0.64	0.65	0.03
Perturbation 2 (HP/50/7)	0.45	0.43	0.44	0.46	0.02
Perturbation 3 (VP/50/4)	0.69	0.67	0.69	0.70	0.02
Perturbation 4 (VP/50/7)	0.76	0.75	0.76	0.77	0.02

- For a given perturbed WordNet PIT, the values of the rank correlation reflect the individual sensitivity of criteria to the particular type of perturbations.

The significantly high values of observed correlations show that, in general, the proposed perturbations are rather slight producing the taxonomies being far from chaotic ones.

From Tables I–III, it follows that PageRank is the most sensitive measure to perturbations as the rank of a node highly depends on the number of incoming and outgoing links. PageRank reacts more on horizontal perturbations than on vertical ones. The former are more destructive in the sense of taxonomic structure since they represent the relations between the notions of the same level of abstraction but of different topics. FloorRank also is highly reactive to taxonomy perturbations behaving very similar to PageRank.

Information Content (IC) ranking is practically invariant to perturbations. This can be explained by the nature of IC: the changes in a taxonomy structure redistribute IC between the nodes so that the dispersion of its values stays approximately the same with the growth of node number.

Since all the perturbations we consider are structural, it is not surprising that a semantic measure such as IC performs worse than structural measures (PageRank and FloorRank). The surprising outcome reported here is that in fact IC fails to catch such semantic errors as are induced by drastic structural perturbations.

The data exhibited in Tables IV–VI confirm the aforementioned results: over the wide scale of WordNet PITs and their perturbations, the PageRank and FloorRank criteria are very close to each other in performance at the same time being far from the Information Content method. In general, the pairwise correlations between taxonomy evaluation criteria may give a good basis for their well-grounded multivariate ranking with various further applications.

VII. Conclusions

In this paper, the problems of the comparison of taxonomy evaluation criteria and of the creation of benchmark taxonomies are studied.

The class of Primitive Ideal Taxonomies, their WordNet and perturbed versions are proposed as taxonomy benchmarks for the comparison of taxonomy evaluation methods.

For WordNet PITs and their perturbations, the comparative performances of the structure-based PageRank and FloorRank criteria as well as the corpus-based Information Content criterion are studied in Monte Carlo experiment.

It is shown that the proposed rank correlation approach can be used for the ranking of taxonomy evaluation criteria specifying the individual criterion behavior.

Further work is associated with developing new classes of task-based benchmark taxonomies and a set of task-based benchmarks; extracting real taxonomies for benchmarks; developing approaches to multivariate ranking of taxonomy evaluation criteria.

References

[1] J. Yu, J.A. Thom, and A. Tam. Requirements-oriented methodology for evaluating ontologies. *Information Systems Databases: Their Creation, Management and Utilization,* ISSN: 0306-4379, Vol. 34, No. 8, December 2009, pp. 686-711.

[2] D. Strasunskas, and S.L. Tomassen. Empirical Insights on a Value of Ontology Quality in Ontology-Driven Web Search. In: *R. Meersman and Z. Tari (Eds.)* OTM 2008, Part II, LNCS 5332, Springer-Verlag, pp. 1319-1337.

[3] M. Fernandez, C. Overbeeke, M. Sabou, and E. Motta. What makes a good ontology? A case-study in fine-grained knowledge reuse, *The 4th Asian Semantic Web Conference,* Shanghai, China, 2009.

[4] A. Maedche, S. Staab. Measuring Similarity between Ontologies. In: *Proc. of the European Conference on Knowledge Acquisition and Management - EKAW-2002*, Madrid, Spain, October 1-4, 2002. LNCS, Springer, 2002.

[5] S. P. Ponzetto, M. Strube. Deriving a large scale taxonomy from Wikipedia. In: *Proceedings of the 22nd National Conference on Artificial Intelligence,* 2007.

[6] J. Yu, J.A. Thom, and A. Tam. Ontology Evaluation Using Wikipedia Categories for Browsing, In: *CIKM'07 Proceedings of the 16th ACM Conference on Information and Knowledge Management*, Lisbon, Portugal, November 6-8, 2007, pp. 223-232.

[7] S. Brin S., L. Page. *The Anatomy of a Large-Scale Hypertextual Web Search Engine,* 1998.

[8] L. Muchnik, R. Itzhack, S. Solomon, Y. Louzoun. Self-emergence of knowledge trees: Extraction of the Wikipedia Hierarchies. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, Vol. 76, 016106, 2007.

[9] A. Gomez-Perez. A framework to Verify Knowledge Sharing Technology. *Expert Systems with Application,* Vol. 11, No. 4., 1996, pp. 519-529.

[10] C. Brewster, H. Alani, S. Dasmahapatra, Y. Wilks. Data Driven Ontology Evaluation. In: *Proceedings of International Conference on Language Resources and Evaluation*, 2004.

[11] M. Jones, H. Alani. Content-based Ontology Ranking. In: *9th International Protege Conference*, July 2006.

[12] P. Pirolli. *Information Foraging Theory*. New York: Oxford Univrsity Press, 2007.

[13] T. Weale. Utilizing Wikipedia Categories for Document Classification. *Technical Report*. Ohio State University, Computer Science and Engineering Department, OSU-CISRC-4/08-TR14, 2008.

[14] Y. Netzer, D. Gabay, M. Adler, Y. Goldberg and M. Elhadad, *Ontology Evaluation Through Text Classification,* ENQOIR 2009, Suzhou, China.

[15] H. Alani, C. Brewster. Metrics for Ranking Ontologies. In: *Proceedings of the Workshop on Evaluating Ontologies for the Web EON 2006,* May 2006.

[16] A. Lozano-Tello, A. Gomez-Perez. Ontometric: A method to choose appropriate ontology. *Journal of Database Management* Vol. 15, No. 2, 2004, pp. 1-18.

[17] N. Guarino, C. Welty. An Overview of OntoClean. In: *Handbook on Ontologies,* Springer, 2004, pp. 151-172.

[18] S. Tartir, I. Arpinar, M. Moore, A. Sheth, and B. Aleman-Meza. 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, pp. 45-53.

[19] D. E. Knuth. *The Art Of Computer Programming.* Vol 1. 3rd ed., Boston: Addison-Wesley, 1997.

[20] P. Resnik. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: *Proc. 14th Intern. Joint Conf. on Artificial Intelligence,* Vol 1, Montreal, August 1995, pp. 448-453.

[21] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. *Int. J. Lexicograph.,* Vol. 3, 4, pp. 235-244.