

Cohesion Metrics for Ontology Design and Application

Haining Yao, Anthony Mark Orme and Letha Etzkorn

University of Alabama in Huntsville, Huntsville, Alabama 35899, United States of America

Abstract: Recently, domain specific ontology development has been driven by research on the Semantic Web. Ontologies have been suggested for use in many application areas targeted by the Semantic Web, such as dynamic web service composition and general web service matching. Fundamental characteristics of these ontologies must be determined in order to effectively make use of them: for example, Sirin, Hendler and Parsia have suggested that determining fundamental characteristics of ontologies is important for dynamic web service composition. Our research examines cohesion metrics for ontologies. The cohesion metrics examine the fundamental quality of cohesion as it relates to ontologies.

Key words: Metrics, Cohesion Metrics, Ontology, Ontology Based System, Semantic Web

INTRODUCTION

With the tremendous development of the Internet and combined new technologies, ontologies have become a key technology to provide shared knowledge models to semantic-driven applications on the Internet. Defined as “explicit specification of a conceptualization”[1], ontologies have been used for artificial intelligence, natural language processing, information query systems, agent software systems. With today’s internet-based world, ontologies provide intelligent and adaptive solutions to today’s distributed and dynamic information processing. There have been many knowledge representation languages developed to model domain ontologies. Recently, Web Ontology Language (OWL) is “already being used as an open standard for deploying large scale ontologies on the Web [2]. As an ontology language, OWL supports knowledge representation, domain vocabulary sharing, advanced search, software agents and knowledge management [3, 4].

With the promising benefits provided by ontologies, domain ontology development and management have become more and more important in all kinds of knowledge-driven applications with commercial success. It has become important to determine fundamental characteristics of ontologies [5]. Since software metrics have been used in object-oriented/object-based software systems to assess software quality, we examine ontology metrics that would be a desirable feature in ontology assessment. Similar to software metrics, ontology metrics are expected to give some insight for ontology developers to help them design ontologies, improve ontology quality, anticipate and reduce future maintenance requirements, as well as help ontology users to choose the ontologies that best meet their needs. For ontology-based systems, measuring ontologies in the early phases of the software development life-cycle allows better

management for the later phases, more effective quality assessment and maintenance estimation.

This study, proposes a set of ontology cohesion metrics to measure the modular relatedness of OWL ontologies. These metrics are Number of Root Classes (NoR), Number of Leaf Classes (NoL) and Average Depth of Inheritance Tree of all Leaf Nodes (ADIT-LN). The metrics are collected by using a standard XML DOM parser that parses the XML-based OWL ontology syntactically, but computes cohesion metrics conceptually based on predefined OWL primitives which explicitly defined tree-based semantic hierarchies in OWL ontologies. These metrics are theoretically validated using standard metrics validation frameworks, then are empirically validated by comparing them statistically to assessments performed by a human team of evaluators.

Background

Ontology: An ontology is defined as “a formal, explicit specification of a shared conceptualization [1]. The ontology structure O , proposed by Maedche [4], can be described by a 5-tuple $O = \{C, R, H, rel, \mathring{A}\}$. C is for concepts; R is for relations; H is for concept hierarchy; rel is a function relating the concepts non taxonomically; and \mathring{A} is a set of ontology axioms expressed in appropriate logical language.

As an explicitly defined and machine-processable abstract model, ontologies were developed for the purpose of knowledge sharing to provide shared common understanding about domain knowledge. In order to serve as metadata schemas to help people and machines to communicate concisely and consistently, different ontology representation languages have been used to store domain knowledge so that knowledge can be reused, shared and interchanged. Recently, as an effective and expressive ontology representation language, Web Ontology Language (OWL) is regarded as “an important step for making data on the Web more

machine processable and reusable across applications” [2]. OWL is developed and recommended by the W3C Web Ontology Working Group as a knowledge modeling language to define and instantiate Web ontologies, which describe classes and properties and their instances and from which the logical consequences can be semantically derived, though the facts are not literally presented [4, 6]. OWL is based on RDF/RDF-S (Resource Description Framework and RDF Schema) and evolved from DAML+OIL (DARPA Agent Markup Language + Ontology Interference Layer). As an ontology language with rich representative primitives, OWL provides the following promising benefits: knowledge representation, domain vocabulary sharing, advanced search, software agents and knowledge management [3,4].

Cohesion Metrics: Cohesion traditionally refers to the degree to which the elements in a module belong together. More particularly, in object-oriented software, cohesion refers to the degree of the relatedness or consistency in functionality of the members in a class; strong cohesion is recognized as a desirable property of object-oriented classes because it measures separation of responsibilities, independence of components and control of complexity [7, 8]. Several software cohesion metrics have been proposed, [9-11]. One of the most widely known object-oriented cohesion metrics was proposed by Chidamber and Kemerer [9, 10]: Lack of Cohesion of Methods (LCOM). The value of LCOM is the number of pairs of methods in a class having no common instance variables, $|P|$, minus the number of pairs of methods in the class having common instance variables, $|Q|$. If $|P| < |Q|$, the value of LCOM is set to zero. Value of LCOM is an inverse to cohesion: a higher value of LCOM demonstrates lower cohesion.

Criteria for Analyzing Metrics: It is desirable to have a formal set of metrics evaluation criteria to assess the usefulness and correctness of software metrics. Kitchenham *et al.* [12] proposed a general framework for software measurement validation. They identify concepts that are necessary for measurement:

- * Entities: objects will be researched
- * Attribute: properties will be measured about the entities
- * Units: how to measure an attribute
- * Scale type: nominal, ordinal, interval, or ratio

Described in their framework, a direct measure should have the following four aspects to be theoretically valid:

- * Attribute validity: the entity being evaluated has the attribute.

- * Unit validity: the unit used is appropriate for the attribute.
- * Instrumental validity: the underlying model is valid and the measurement instrument is calibrated.
- * Protocol validity: the protocol used for the measurement is consistent, unambiguous and prevents problems.

Briand *et al.* [11] proposed a set of criteria for different software metrics including more specific criteria for cohesion metrics:

- * Nonnegativity and normalization: the value is never negative and the values can be compared between different modules.
- * Null value: the value is zero if there is no intramodule relationship within a module.
- * Monotonicity: the value never decreases when adding intramodule relationships into a module.
- * Cohesive module: the value of merged modules is never greater than the maximum cohesion of the original modules.

Ontology Cohesion Metrics

Research Description: We consider ontology cohesion metrics as part of a measure for ontology modularity: ontology cohesion refers to the degree of the relatedness of OWL classes, which are semantically/conceptually related by the properties. An ontology has a high cohesion value if its entities are strongly related. The idea behind this is the concepts grouped in an ontology should be conceptually related for a particular domain or a sub-domain in order to achieve common goals. We did related research from software cohesion metrics and noticed that actually the most cited Chidamber and Kemerer software cohesion metrics were theoretically based on concepts similar to those of objects of ontologies [13, 14]: because cohesion metrics are intended to measure modularity, the metrics similar to the software cohesion metrics can be defined to measure relatedness of elements in ontologies.

In this paper, our goal is to define and validate a set of ontology cohesion metrics to measure cohesiveness of OWL ontologies. The idea is that these ontology cohesion metrics, similar to software cohesion metrics, can be used to measure separation of responsibilities and independence of components of ontologies.

To achieve this goal, we studied ontologies and ontology engineering in general in order to propose a set of cohesion metrics based on general characteristics of ontologies, which we then theoretically validated using Kitchenham *et al.* [12] software measurement validation framework and Briand *et al.* [11] more specific validation criteria for cohesion metrics. An automated data collection tool was then developed and implemented to collect samples of our

ontology metrics from a set of OWL ontologies. The results were statistically compared with the assessments of a human evaluation team in order to demonstrate the correlation between the two data sets.

Common Formal Notation: We use the following formal notation to mathematically define the ontologies and the cohesion metrics:

- * Let C_1, C_2, \dots, C_m be the set of m classes explicitly defined in an ontology.
- * Let P_1, P_2, \dots, P_n be the set of n properties which work as relationship between the set of classes.
- * Let $F_{c_1}, F_{c_2}, \dots, F_{c_m}$, be the fanout of each class C_i in the set.
- * Let O_i be an OWL ontology of interest.
- * Let \rightarrow be subtype relationship from C_i to C_j such that $C_i \rightarrow C_j$ if class C_j is a subclass of class C_i .

In ontologies, concepts are typically organized into a taxonomy tree where each node represents a concept and each concept is a specialization of its parent [5]. OWL defines hierarchical constructs to present a tree-based structure for subtype relationships between entities in OWL ontologies. The following terminology used in this paper describes this tree-based relationship.

- * Let A be a finite set and T be a relation, then T is a tree if there is a vertex V_0 in A and there exists a unique path in T from V_0 to every other vertex in A [16].
- * Let A be a finite set and T be a relation on A . A vertex V_n has a Fanout of degree m if there exist m relationships to other vertices in A [16].
- * Let A be finite set and T be a relation on A . A vertex V_q is called a leaf of the tree if it has Fanout of degree 0.

Cohesion Metrics Definitions

Cohesion metric #1: Number of Root Class:

Definition: Number of Root Classes (NoR) is the number of root classes explicitly defined in the ontology O_i . A root class in an ontology means the class has no semantic super class explicitly defined in the ontology O_i . For example, an OWL ontology semantic hierarchy is defined as follows in Fig. 1.

Mathematically, NoR can be formulated as follows:
 $NoR(O_i) = \sum C_j$ for all $1 \leq j \leq n$ (number of root classes in O_i)

In the Fig. 1, the ontology O_i has 3 root classes, C_1, C_5, C_6 , thus $NoR(O_i) = 3$.

Cohesion metric #2: Number of Leaf Classes:

Definition: Number of Leaf Classes (NoL) is the number of leaf classes explicitly defined in the ontology O_i . A leaf class in an ontology means the class has no semantic subclass explicitly defined in the ontology O_i .

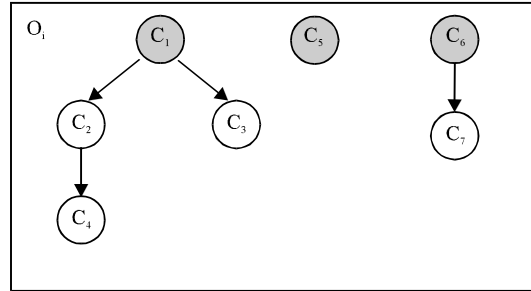


Fig. 1: Number of Root Classes (NoR)

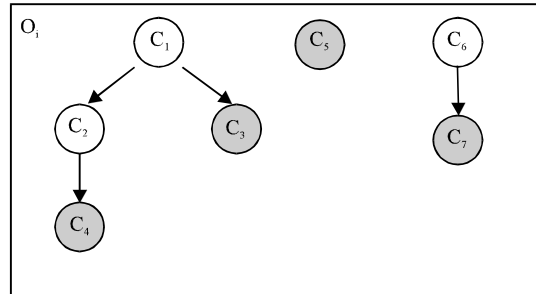


Fig. 2: Number of Leaf Classes (NoL)

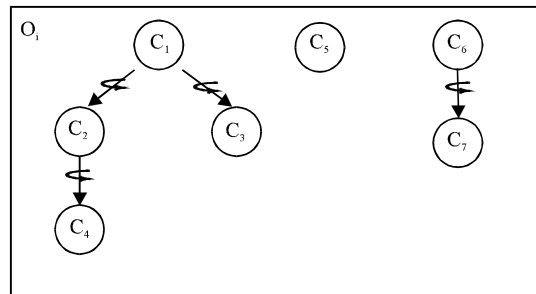


Fig. 3: Average Depth of Inheritance Tree of Leaf Nodes (ADIT-LN)

For example, an OWL ontology semantic hierarchy is defined in Fig. 2.

Mathematically, NoL can be formulated as follows:

$$NoL(O_i) = \sum L_j \text{ for all } 1 \leq j \leq n \text{ (number of leaf classes in } O_i)$$

In the Fig. 2, the ontology O_i has 4 leaf classes, C_4, C_3, C_5, C_7 , thus $NoL(O_i) = 4$.

Cohesion metric #3: Average Depth of Inheritance Tree of Leaf Nodes:

Definition: Average Depth of Inheritance Tree of all Leaf Nodes, ADIT-LN is the sum of depths of all paths divided by the total number of paths. A depth is the total number of nodes. A depth is the total number of nodes starting from the root node to the leaf node in a path. The total number of paths in an ontology is all distinct paths from each root node to each leaf node if there exists an inheritance path from the root node to the leaf node. And root node is the first level in each path. For example, ADIT-LN of an OWL ontology is described in Fig. 3.

Mathematically, ADIT-LN is formulated as follows:
 $ADIT-LN(O_i) = \sum D_j / n$ for all D_j (D_j is total number of nodes on j^{th} path); $1 \leq j \leq n$ (number of paths in O_i);
In the Fig. 3, the ontology O_i has $ADIT-LN(O_i) = 2$.

Analysis of Ontology Cohesion Metrics

Theoretical Validation

Analysis of NoR (Number of Root Classes):
According to Kitchenham *et al.* [12] software measurement validation framework, cohesion metrics NoR is a direct measure, in this case, to count the root nodes (classes) in ontologies. For this measurement:

- * The entity is the ontology O_i being analyzed.
- * The attribute measured is the number of root classes.
- * The unit is the class.
- * The data scale is interval.

The NoR meets to Kitchenham, Pfleeger and Fenton's properties as follows:

- * Attribute validity: the entity (the ontology being analyzed) has the attribute (number of root classes), a measure of the total number of root classes explicitly defined in the ontology.
- * Unit Validity: the attribute is measured by counting the number of root classes.
- * The instrument is valid as long as the metrics collecting tool parses and counts the number of root classes defined in the ontology O_i correctly.
- * Protocol validity: calculations performed according to the formal notation given in this paper will be free from counting errors by counting the number of root classes, which is consistent and unambiguous.

According to Briand, Morasca and Basili's cohesion measurement validation:

- * Nonnegativity and Normalization: the value of NoR is never negative and the values can be compared between different ontologies.
- * Null Value: not applicable to the NoR metric. If there is no intramodule relationship within an ontology, the value of NoR is still the number of root nodes (in this situation, the root nodes are also leaf nodes).
- * Monotonicity: may not applicable to the NoR metric. The value of NoR may decrease if a root node becomes a non root node after adding an intramodule relationship. For example, in Fig. 4, the added relationship (represented by a dashed arrow line) makes the previous root node C_5 a subclass of another node C_3 . Otherwise, the criterion is applicable to the NoR metric.

- * Cohesive Module: the value of NoR of merged modules is never greater than the maximum NoR of the original modules. The number of root nodes of merged modules will never be greater than the maximum number of root nodes in the original modules, because there is no reason that a non-root node should become a root node after merging modules.

Analysis of NoL (Number of Leaf Classes):
According to Kitchenham, Pfleeger and Fenton's software measurement validation framework, cohesion metrics NoL is a direct measure, in this case, to count the leaf nodes (classes) in ontologies. For this measurement:

- * The entity is the ontology O_i being analyzed.
- * The attribute measured is the number of leaf classes.
- * The unit is the class.
- * The data scale is interval.

The NoR meets to Kitchenham *et al.* [12] properties as follows:

- * Attribute validity: the entity (the ontology being analyzed) has the attribute (number of leaf classes), a measure of the total number of leaf classes explicitly defined in the ontology.
- * Unit Validity: the attribute is measured by counting the number of leaf classes.
- * The instrument is valid as long as the metrics collecting tool parses and counts the number of leaf classes defined in the ontology O_i correctly.
- * Protocol validity: calculations performed according to the formal notation given in this paper will be free from counting errors by counting the number of leaf classes, which is consistent and unambiguous.

According to Briand *et al.* [11] cohesion measurement validation:

- * Nonnegativity and normalization: The value of NoL is never negative and the values can be compared between different ontologies.
- * Null value: not applicable to the NoL metric. If there is no intramodule relationship within an ontology, the value of NoL is still the number of leaf nodes (in this situation, the leaf nodes are also root nodes).
- * Monotonicity: may not applicable to the NoL metric. The value of NoL may decrease if a leaf node becomes a non leaf node after adding an intramodule relationship. For example, in Fig. 5 the added relationship (represented by a dashed arrow line) makes the previous leaf node C_3 a superclass

of another node C_4 . Otherwise, the criterion is applicable to the NoL metric.

- * Cohesive module: the value of NoL of merged modules is never greater than the maximum NoL of the original modules. The number of leaf nodes of merged modules will never be greater than the maximum number of leaf nodes in the original modules, because there is no reason that a non-leaf node should become a leaf node after merging modules.

Analysis of ADIT-LN (Average Depth of Inheritance Tree of Leaf Nodes): According to Kitchenham, Pfleeger and Fenton's software measurement validation framework, cohesion metrics ADIT-LN is a direct measure to count the depth of the inheritance tree for all leaf nodes (classes) in ontologies. For this measurement:

- * The entity is the ontology O_i being analyzed.
- * The attribute measured is the average depth of inheritance tree of all leaf classes.
- * The unit is the depth of inheritance.
- * The data scale is interval.

The ADIT-LN meets to Kitchenham *et al.* [12] properties as follows:

- * Attribute validity: the entity (the ontology being analyzed) has the attribute (number of leaf classes), a measure of the average depth of inheritance for all leaf nodes in the ontology interested.
- * Unit Validity: the attribute is measured by counting the depth of all leaf classes.
- * The instrument is valid as long as the metrics collecting tool parses and counts the average depth of all leaf classes defined in the ontology O_i correctly.
- * Protocol validity: calculations performed according to the formal notation given in this paper will be free from counting errors by counting the depth of all leaf nodes in the inheritance tree and the number of all leaf classes, which is consistent and unambiguous.

According to Briand *et al.* [11] cohesion measurement validation:

- * Nonnegativity and normalization: the value of ADIT-LN is never negative and the values can be compared between different ontologies.
- * Null Value: not applicable to the ADIT-LN metric. If there is no intramodule relationship in an ontology, the value of ADIT-LN is the 1, because each leaf node has a depth of one.
- * Monotonicity: may not applicable to the ADIT-LN metric. The value of ADIT-LN may decrease. For

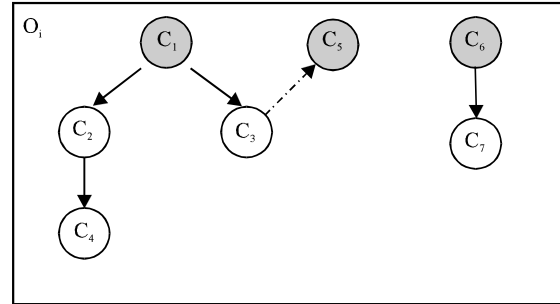


Fig. 4: NoR Decreases after Adding New Intermodule Relationship

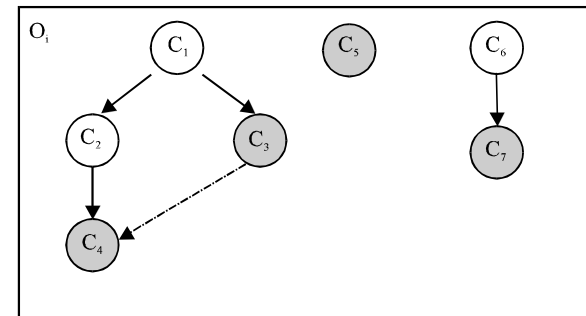


Fig. 5: NoL Decreases after Adding New Intermodule Relationship

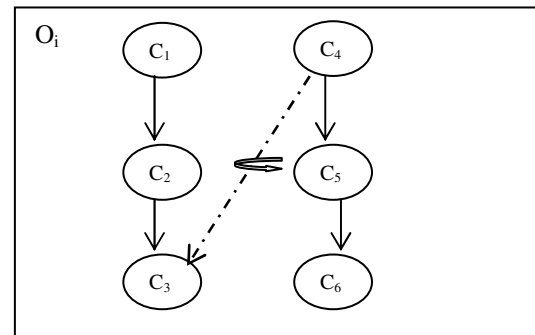


Fig. 6: ADIT-LN Decreases after Adding New Intermodule Relationship

example, in Fig. 6, before adding the relationship (represented by the dashed arrow) the value of ADIT-LN is 3; after adding the relationship the value of ADIT-LN is $8/3$. Otherwise, the criterion is applicable to ADIT-LN metric.

- * Cohesive module: the value ADIT-LN of merged modules is never greater than the maximum ADIT-LN of the original modules. Merging unrelated modules will not increase or decrease the total number of nodes or the total number of paths in the merged modules. The value of overall ADIT-LN is the average of the values of ADIT-LN of each module. So the average of all ADIT-LN of all modules is never greater than the maximum ADIT-LN of the original modules.

Empirical Validation

Description of Empirical Study: To perform our empirical analysis on ontology cohesion metrics, we collected a set of ontologies developed by Creative Commons [17]. The computation was performed by Ontology Metrics Parser (OMP), a XML parser, that parses the XML-based OWL ontology syntactically, but computes cohesion metrics conceptually based on predefined OWL primitives which have explicit semantic meanings.

Then a panel of eighteen evaluators were assembled to assess the set of ontologies to determine cohesiveness of these ontologies. The evaluators have 3 to 5 years average experience in software development, while thirteen of the evaluators had experience with knowledge based systems or knowledge representation. The remaining evaluators were provided ontology training before serving in an evaluation capacity. Each expert was given the set of ontologies and asked to assess the cohesiveness of each ontology. The experts rated the cohesiveness of each ontology on the following scale:

- * 0 = Low
- * 0.25 = Moderately
- * 0.5 = Average
- * 0.75 = High
- * 1.0 = Excellent

First, we determined Interrater reliability. Interrater reliability determines how well the evaluators agree with one another on a particular evaluation. For example, one evaluator rates a particular ontology with 0.25, while another rates the same ontology with 1.0. Therefore, Interrater reliability addresses the consistency of the implementation of a rating system and is expressed as a real number in the range of [0.1]. We computed Interrater reliability from assessment of all eighteen experts and across all thirty three ontologies. We used MiniTab's Two-Way Mixed Effect Model and considered the people effect to be random and the measure effect to be fixed. For our case, the Interrater reliability is 0.9014, which indicates a consistent agreement between the evaluators.

Next, we performed statistical analysis to check the correlation between the averaged evaluation ratings for each of the ontologies and the cohesion metrics. We used Pearson's correlation coefficient with the following hypotheses:

H0: $\rho = 0$ (There is no correlation between the metrics values and the evaluators' values)

H1: $\rho \neq 0$ (There is correlation between the metrics values and the evaluators' values)

The Pearson correlation coefficient measures the strength and direction of linear relationship between a pair of variables. The correlation coefficient assumes between -1 and 1. The larger absolute value of the correlation coefficient means stronger correlation

between the pair of variables. And if the pair of variables are independent, the correlation coefficient between the two variables is 0. The following scales proposed by Cohen [18] and Hopkins [19]:

- * < 0.1 - trivial
- * .10 - .30 - minor
- * .30 - .50 - moderate
- * .50 - .70 - large
- * .70 - .90 - very large
- * .90 - 1.0 - almost perfect

Another quantitative measure for the correlation between a pair of variables is the p-value. P-values are used in hypothesis tests to either reject or fail to reject a null hypothesis. A small p-value indicates that a null hypothesis is false.

Table 1 shows the Pearson correlation coefficients between cohesion metrics values and evaluators' values and p-values for the hypothesis that H0: $\rho = 0$.

Table 1: Pearson Correlation between Cohesion Metrics Value and Evaluators' Value

Metrics	Correlation	
	Coefficient ρ	P-value
NoR	0.386	0.035
NoL	0.687	< 0.0001
ADIT-LN	0.641	< 0.0001

CONCLUSION AND FEATURE RESEARCH

In this study, we introduced cohesion metrics for measuring ontology cohesiveness which can help ontology developers and system developers better understand ontology structures and potentially could help estimate cost and maintenance for the ontology itself and also for the whole lifecycle of ontology-based software system.

We have performed theoretical and empirical analysis of the ontology cohesion metrics #1, 2 and 3. The results from our theoretical studies indicate that the ontology cohesion metrics are theoretically valid. The results from our empirical studies indicate that a good correlation exists between evaluators' opinions of ontology cohesiveness and the cohesiveness measured by the cohesion metrics proposed by this study.

In the future, we will continue work on the ontology cohesion metrics and other ontology metrics. In this research, we mainly considered the subclass relationship of classes in ontologies. We may add more metrics to the ontology cohesion metrics set. Also, future research may include additional uses for the metrics and how the metrics can be used to effect ontology system development.

REFERENCES

1. Gruber, T.R., 1993. A translation approach to portable ontology specification. Knowledge Acquisition, 5:199-220.

2. Berners-Lee, T., 2003. News release: World wide web consortium issues web ontology language candidate recommendations, <http://lists.w3.org/Archives/Public/www-webont-wg/2003Aug/0091.html>, Aug., 2003.
3. Heflin, J. (Eds.), 2004. OWL Web Ontology Language Use Cases and Requirements. W3C Recommendation 10 Feb., <http://www.w3.org/TR/webont-req/#section-goals>.
4. Maedche, A., 2002. *Ontology Learning for the Semantic Web*. Kluwer Academic Publishers.
5. Sirin, E., J. Hendler and B. Parsia, 2003. Semi-automatic composition of web services using semantic description. *Web Services: Modeling, Architecture and Infrastructure Workshop in Conjunction with ICEIS2003*.
6. Smith, M.K., C. Welty and D.L. McGuinness (Eds.), 2004. *OWL Web Ontology Language Guide*. W3C Recommendation 10 February 2004, <http://www.w3.org/TR/owl-guide/>.
7. Bansiya, J., L. Etzkorn, C. Davis and W. Li, 1999. A class cohesion metric for object-oriented designs. *J. Object-Oriented Programming*, 11: 47-52.
8. Chae, H.S., Y.R. Kwon and D.H. Bae, 2000. A cohesion measure for object-oriented classes. *Software Practice and Experience*, 30: 1405-1431.
9. Chidamber, S.R. and C.F. Kemerer, 1991. Towards a metrics suite for object oriented design. *ACT SIGPLAN Notices*, 26: 197-211.
10. Chidamber, S.R. and C.F. Kemerer, 1994. A metrics suite for object oriented design. *IEEE Trans. Softw. Eng.*, 20: 476-493.
11. Briand, L.C., S. Morasca and V.R. Basili, 1996. Property-based software engineering measurement. *IEEE Trans. Softw. Eng.*, 22: 68-86.
12. Kitchenham, B., S. Pfleeger and N. Fenton, 1995. Towards a framework for software measurement validation. *IEEE Trans. Softw. Eng.*, 21: 929-944.
13. Bunge, M., 1977. *Treatise on Basic Philosophy: Ontology I: The Furniture of the World*. Boston, Riedel.
14. Bunge, M., 1979. *Treatise on Basic Philosophy: Ontology II: The World of Systems*. Boston, Riedel.
15. Doan, A., J. Madhavan, R. Dhamankar, P. Domingos and A. Halevy, 2003. Learning to match ontologies on the semantic web. *VLDB J.*, 12: 303-319.
16. Kolman, B. and R.C. Busby, 1987. *Discrete Mathematical Structures for Computer Science*. 2nd Edn. Prentice-Hall, Inc., Englewood Cliffs, N.J.
17. Chen, H., 2004. Creative Commons: <http://daml.umbc.edu/ontologies/cobra>. Last accessed 06/07/04.
18. Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edn. Lawrence Erlbaum Publishing Co., Mahwah, NJ.
19. Hopkins, W., *A new view of statistics*, SportScience, Dunedin, New Zealand, <http://www.sportsci.org/resource/stats>.