Evaluation of Concepts Asset in Topic Maps Ontology

PETRA HALUZOVÁ
Department of Informatics and Telecommunications
Faculty of Transportation Sciences
Czech Technical University in Prague
Konviktská 20, 110 00, Prague
CZECH REPUBLIC
haluzpet@fd.cvut.cz  http://www.fd.cvut.cz/

Abstract: In connection with structuring and sharing of knowledge in a form of ontologies the question of their assets evaluation occurs. This paper presents the methodology of asset evaluation of individual instances in ontology based on standard of Topic Maps. The evaluation comes out of weights assignment to instances and associations in ontology and total weight calculation for individual instance and its surroundings. Searching results can be ranked according to this total weight expressing information asset of individual instance.

Key-Words: ontology, information asset evaluation, weights of instances, weights of associations, Topic Maps

1 Introduction
With rapid Semantic Web development was created a lot of ontologies whose main goal is to facilitate knowledge sharing in a given domain of interest [1, 2]. The ontology is defined as an explicit specification of a conceptualization [3]. Ontologies comprise two necessary parts in particular, definitions of concepts and their inter-relationships.

In connection with knowledge evaluation, specifically their gains evaluation and currently discussed ontologies raises the question what the ontology actually yields, e.g. as opposed to the standard searching approach. There is a demand to try to evaluate the asset amount of individual concepts in ontology.

The paper is organized as follows: Section 2 prefers the work related to the theme of ontology or their parts evaluation. Section 3 focuses on Topic Maps standard, Section 4 describes weights assignment to instances and associations and proposes the entire procedure of evaluation. The last section contains conclusions and several notes of further work possible.

2 Related Research
Zouaq and Nkambou [4] mentioned that there is no single right way of evaluating ontologies. It is more practical to focus on the evaluation of different levels of the ontology separately rather than trying to evaluate the ontology as a whole. The levels have been defined variously depending on authors, e.g.: lexical, vocabulary or data layer; hierarchy; semantic relations; context, application level; syntax; structure, architecture [5].

The approach summary of several authors related to ontology evaluation based on its structure with examples of various metrics and weights implementation follows. Yao, Orme and Etzkorn [6] define a set of cohesion metrics. Ontology cohesion refers to the relatedness of OWL classes which are conceptually related by the properties. The high cohesion value points out that the entities are strongly related. Proposed metrics are Number of Root Classes (without any semantic super classes), Number of Leaf Classes (having no semantic subclasses) and Average Depth of Inheritance Tree of all Leaf Nodes.

Huang and Diao [7] put forward six properties for ontology structure characterization. The approach is based on the internal ontology structure, statistics and graph theory. The proposed properties are: Concept Quantity (corresponds to the size of ontology), Property’s Expectation (provides an overview of the abundance of relations among concepts), Property’s Standard Deviation (reflects the situation of the property’s distribution in ontology), Tree Balance (evaluates the analogous structure of concept subclasses), Concept Connectivity (in case two concepts share an instance, a connectivity edge will be added between them and then the number of these edges will be counted), Key Concept Quantity (expresses the number of concepts which have more adjacent concepts than the other).

Rocha, Schwabe and Aragao [8] introduced Weight Mapping – a technique for calculating the numerical weight value for each relation, based on the analysis of the knowledge base structure. It helps to obtain better searching results, because all concepts instances that are related to a given keyword are also taken into account. Three different measures were proposed in order to calculate the weights:
3 Topic Maps

Topic Maps (TM) are ISO standard (ISO/IEC 13250:2003 Information technology – SGML applications – Topic maps) for the knowledge structures description and their information sources connection. It seems to be more intuitive technology rather than RDF or OWL. In comparison with languages of Semantic Web ontologies it is possible to define even the associations with arity higher then 2. The TM language is not intended for machine inference, but only for cataloguing and sources searching and therefore it is not connected with formal-logical tool [11].

The main idea of TM rests in two layers division: the lower layer of original information sources and the upper knowledge layer. The knowledge layer contains topics and their inter-relationships. The connection between the information and knowledge layer is realized in occurrences. The basic elements of TM are:

- **Topic** – represents certain subject which we want to capture in TM. Topics can be grouped into types ergo topic type consists of other topics. The topic has three main characteristics: name, occurrence and association role. It is allowed that one topic has more names.
- **Association** – asserts relations between topics. Associations are mainly binary, in general they can have the arity higher than 2. They are not orientated, thus the orientation of the relationship is expressed through roles.
- **Occurrence** – captures the relationship between the topic and related information source. Occurrences can be internal (saved directly in TM) or external (in a form of links).

It is possible to create a TM ontology e.g. in OKS (Ontopia Knowledge Suite) by Norwegian firm Ontopia. It is a client-server application consisting of several modules: Ontopoly for creating and editing, Omnigator for browsing of created ontology in a form of text and Vizigator for graphic representation.

4 Methodology of Asset Evaluation of Instances

4.1 Ontology Matrix Structure

The ontology in TM consists of three main elements: Topics, Associations and Occurrences. Basically, it is possible to consider the associations and occurrences as attributes, which can gain particular values and they are added to topics. The attribute-association value is another topic in TM. The attribute-occurrence value can be, e.g. www URL, article, reference to specific part of text, concise or detailed description directly in topic, picture, etc. Topics can also comprise other attributes describing their properties, characteristics or parameters.
Each of these three main elements can have its own type and these types become other topics in the same TM. It results in the fact that everything in TM becomes a topic. Nevertheless, it is important to distinguish between topic type and instance. The instance is a final element which can not contain other instances. On the contrary, the topic type can include an instance.

It is necessary to add that for algorithmic processing is more convenient to use matrix form of ontology. E.g. the authors in [12] viewed ontology as a matrix. They used an adjacency and accessibility matrix for analyzing of changes in ontology during its evolution.

IDs are automatically assigned to all topics and associations in XTM (Topic Maps format). There may appear different topics under the same name in TM. But there are adhered different IDs to them. These IDs facilitate the transformation from ontology to matrix or lists. Then it is easier to assign weights and to process the calculation for each individual instance. Fig. 1 shows an example of ID#98 instance connection with surrounding instances. As we can see in this scheme, it arises that we need the matrix of all instances and their attributes (Fig. 2) and the matrix of all interconnected instances including associations types (Fig. 3). This form of ontology facilitates to calculate all the instances again incorporating their weight changes.

$$\sum_{ij} w_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}$$

Figure 2: Matrix of instances and their attributes

![Matrix of instances and their attributes](image)

Figure 3: Matrix of instances and their associations

4.2 Instances and Associations Weights

4.2.1 Partial Weight of Instance
Partial weight of each instance $c = \{1, 2, ..., c \}$ emerges from the description in ontology, i.e. how many and what attributes the instance contains (Fig. 2). We individually assess each attribute according to its importance. For example attribute-occurrence “picture” has greater importance than attribute-occurrence “concise description”. In opposite, attribute-occurrence “external web link” has higher weight value than both of the above mentioned attributes, because we assume that the webpage offers more information than the picture or concise description. There are many attribute types depending on the size and elaboration of ontology so we can simplify weight assignment by attributing them with 1. It means that we perceive all attributes at the same importance level and the partial weight of instance grows evenly depending only on the attributes amount, but not depending on its type.

4.2.2 Association Weight
The weight of each association $a \in (0, 1); a \in R$ is determined by the category to which the association belongs to. It seems applicable to put associations into three categories (A, B, C in Fig. 3):

- **Hierarchical** – express the hierarchical order, i.e. the relationship between more general and less general instances.
- **Defining** – refer to definitions, origin, explanations. They are the strongest out of all these categories.
- **Contextual** – express contexts, examples, notes. They are the weakest out of all these categories.

The categorization of associations is difficult to generalize, because it depends on the specific ontology and on the user’s point of view. For example in ontology about Opera which contains composers, their birthplaces, pieces, etc. we consider the association ‘was learner-teacher’ as hierarchical, ‘based on’, ‘composed by’ as defining and finally ‘has voice type’ as contextual.

We can simplify category weight assignment by attributing them the same value, i.e. we are interested in the existence of association between two instances, not in association type.

It would be possible to assign the weight of each association type as in case of instances attributes and then normalize these weights to interval $(0, 1)$. However, associations are not primarily that important for information asset evaluation as attributes are, therefore it is sufficient to divide them into groups and assign weights according to these groups.

4.2.3 Total Instance Weight
From the above mentioned weights we calculate the evaluation $w_i$ of each individual instance $i$ in ontology as:

$$w_i = c_i + k \cdot \sum_{j} a_k \cdot c_j$$

where $c_i$ is a partial weight of instance $i$, $a_k$ are weights of associations between instance $i$ and surrounding instances $j$, $c_j$ are partial weights of surrounding instances $j$. The coefficient $k \in (0, 1); k \in R$ expresses the weight we attribute to all surrounding instances (independently on partial weights of these instances).
Unless we want partial weights of surrounding instances $c_j$ to influence the total instance weight $w_i$ to a great extent, we choose e.g. $k = 0.5$. In case the partial weights of surrounding instances are of the same importance as a partial weight of central instance (instance $i$), we ascribe $k = 1$.

Let us consider the example in Fig. 1. Because of the simplification we use the same weight values for all association categories, i.e. $a_j = 1$ and $k = 0.5$ (we perceive partial weights of surrounding instances as having the half level of importance in comparison with the central instance). The total instance weight #98 according to (1) is:

$$w_{98} = c_{98} + k \cdot [(a_{98,34} \cdot c_{34}) + (a_{98,196} \cdot c_{196}) + (a_{98,259} \cdot c_{259})] = 5 + 0.5 \cdot (7 + 3 + 3) = 11.5.$$  

After calculating the total instance weight for each individual instance in ontology we normalize all results using the maximal calculated value to interval $(0, 1)$.

### 4.3 Methodology of Asset Evaluation of Instances

We summarize the mentioned procedure in steps as follows:

1. Assign the weights to all attributes which can occur within an instance (simplification: consider weight of all attributes 1).
2. Divide associations into three categories introduced in Section 4.2.2. and assign the weights to these categories (simplification: consider weight of all associations 1).
3. Calculate the total instance weight for each individual instance in ontology according to (1).
4. Normalize all calculated total instance weights in accordance with the maximal value.

### 5 Conclusion and Further Work

This paper introduced the procedure of obtaining the information asset quantification for individual instances in ontology. We assigned the weights to individual concepts (instances) and associations. Based on these weights and on the relation of topics via associations we calculated the total evaluation value of each individual instance and its surroundings. If more instances correspond with searching criterion, we can rank the searching results according to this calculated value. So far the analyses were carried out in a testing ontology created out of several transport standards. Also the sample ontology Opera from the OKS tool was used.

Further work will be focused on the automation of the whole calculating process. The extension of relevant surroundings of each instance will be taken into account. The partial weights of instances in more distant surroundings could be appropriately decreased by means of further coefficient.

### References:


