WHAT ARE MAIN CONCEPTS IN AN OWL DOMAIN ONTOLOGY?

Christian Kop
Institute of Applied Informatics, Alpen-Adria-Universitaet Klagenfurt
Universitaetsstrasse 65-67, 9020 Klagenfurt, Austria

Keywords: Focus of ontology, Ontology structure, Main concept.

Abstract: Whereas OWL is suitable for machine interpretation, it is hard to read for a human and it is hard to understand what the real focus of the ontology is. This paper will discuss if measures based on the ontology structure can help.

1 INTRODUCTION

Whereas OWL (McGuiness et.al., 2004) is suitable for machine interpretation, it is hard for a human to get an overview of the ontology focus in order to decide if the content of the ontology is correctly expressed by it's name or it's introductive comments. For instance, does the name of an ontology (e.g. *pizza.owl*) also reflect the ontology's content (it's structure)? Such information is useful to decide if the ontology or it's elements are appropriate in a similar domain.

In order to get this kind of information, the ontology structure is searched for concepts (main concepts) which seem to be important (Huang et.al. 2006).

Therefore the paper is structured as follows. In the next section it will be discussed how main concepts can be detected. Afterwards, possible application scenarios are given in section 3. Section 4 finally gives a conclusion.

2 WHAT ARE MAIN CONCEPTS?

With regard to an ontology purpose there are always concepts which have a richer description within an ontology than others to underline their importance for that purpose. Such descriptions might be the involvement in a generalization hierarchy or relationships to other concepts. These concepts are called **main concepts** in this paper.

If these concepts can be detected, then they are helpful to get an overview of the ontology structure. Thus, the expectations of the human can be compared with the ontology focus derived from the structure.

2.1 Measures based on the Structure

Whereas, in (Huang et.al. 2006) only one measure was used, this approach discusses four measures namely: Weighted number of successors (**wNS**), number of object properties (**P**), weighted number of object properties (**wP**) and instances of an OWL class (**I**). They reflect possible different ways how main concepts can be modeled (Guarino, 1995).

Weighted Number of Successors (wNS). The modeling construct often used in ontologies is generalization. In (Bezerra et.al, 2009) the number of direct children is counted. Alternatively, the number of successors (**NS**) in the whole subclass hierarchy of a certain concept is counted here. To avoid that OWL classes at the top of the hierarchy will always be the winners, the number of successors are multiplied with a weighting (figure 1).

This weighting is defined as follows: The root element ("Thing") is weighted with 0. From the root element the path to the concept is calculated as "distance from root" (**dfr**). Additionally, the maximum distance to a leaf from that concept is calculated (**maxdtl**).

For a class which is a leave in the hierarchy, maxdtl is defined as 1. The weighting factor is **dfr / maxdtl**. The weighted number of successors for a concept X then is: **wNS(X) = NS * (dfr / maxdtl)**

Kop C. (2009).
WHAT ARE MAIN CONCEPTS IN AN OWL DOMAIN ONTOLOGY?.
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pages 404-407
DOI: 10.5220/0002294204040407
Copyright © SciTePress

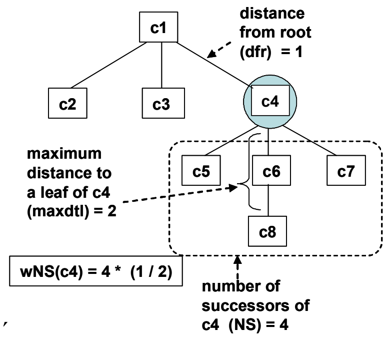

Figure 1: Example for wNS calculation.

Number of Object Properties for a Domain Class (P). An OWL object property points from the domain class to the range class. In other words, the domain is described with the range. Thus domain classes are more likely main concepts. If a certain class is involved as a domain in an object property, then a counter is incremented by 1 for that class.

Weighted Number of Object Properties (wP). A more refined version to count is to add the weighted number of successors of the range class to the domain class. However, if **wNS** for the range is 0 then **wP** is incremented by 1. In this case, it degrades to **P**. With this strategy the importance (weight) of the range is forwarded to the domain.

Instances of an OWL Class (I). Defining many instances of a class can be an additional way to describe a main concept.

2.2 Discussion

The measures in the previous section were tested on 6 ontologies which were found on the web site: *http://krono.act.uji.es/Links/ontologies/*. Particularly the well known pizza-, food- and wine-ontology as well as an ontology of photography, of rheumatic patients and a simple ontology of an university were selected. These ontologies were chosen for the following reasons: According to the filenames it was expected, that they describe typical domain ontologies. Furthermore these domain ontologies are not too specific (e.g. very specific medical, biological or technical domains) but domains, which represent general knowledge also known to the author. Hence it is possible to easily compare the results of the measures with the author's understanding of the ontology.

Pizza Ontology. Using the **wNS**, *Named Pizza* followed by *Pizza*, *Pizza Topping*, *Vegetable Topping*, *Cheese Topping*, *Domain Concept* are the best ranked concepts. Using **wP**, there are three important concepts, namely *Pizza Topping* followed by *Pizza Base* and *Pizza*. *Pizza Topping* and *Pizza Base* are important concepts, since in the original ontology, they are also domains of an object property (*Pizza Topping has Topping of Pizza* and *Pizza Base is base of Pizza*). The same holds for **P** but now *Pizza* is top ranked because it is involved in two object properties whereas the two others are only involved in one object property as a domain class. Using **I,** only *Country* appears on top of the list. Hence this measure alone does not give good hints for main concepts in this ontology. What can now be said on the basis of three of the four measures is that in fact the ontology describes *Pizzas* and how they are made (e.g. made with some *Topping* and some *Pizza Base*). The measures reflect the author's intuitive understanding and expectations of this domain.

Food Ontology. The measures applied on the food ontology return the following results. With the **wNS** the three top ranked concepts are *Meal Course*, *Edible Thing* and *Consumable Thing*. Then some other concepts like *Pasta*, *Seafood*, *Fish*, *Pasta with red sauce* etc. follow. But these concepts did not get that high score. *Meal* and *Meal Course* were also top ranked in the **wP** measure. *Consumable Thing* got the third position in the **P** and **wP** ranking. For **I**, *Oyster Shellfish*, *Sweet Fruit* and some other concepts were top ranked. Once again except of the instance counter (**I**) the results look promising.

Wine Ontology. During determination of the concept statistics one problem arose. The wine ontology imports the food ontology. If also the imported concepts from the food ontology are measured, then the concepts of the food ontology are top ranked and not the concepts of the wine ontology. However if the wine ontology is determined locally, then the following can be said:
According to **wNS** the concept *Wine* is the highest ranked concept. The next concept is *Loire*. Afterwards, *Bordaux*, *Medoc*, *White Wine* follows. But these concepts do not have the same high score like the concept *Wine*. Applying **wP** and **P**, *Wine*, *Region* and *Vintage* were best ranked. Finally with **I,** *Winery*, *Region*, *WineGrape* were top ranked, followed by some other concepts. Here the first time also the concepts ranked by their number of instances make sense.

Photography Ontology. With **wNS**, *Film*, *Equipment*, *Lens*, *Light Sensitive Auto Focus*, *Camera*, *Filter*, *Physical Thing*, *Shutter Speed*, *Test Lens*, *Exposure Parameter* are on top of the list. No values were determined for **wP**, **P** and **I**. The reason is, that the object properties are totally independent in this context. Also, no instances are defined for the specified concepts. However, on the basis of **wNS**, it can be said, that this ontology focuses on the concepts named above.

Simple University Ontology. On the basis of **wNS**, the top ranked concept in the university ontology was the concept *Module*. Then the concept *Academic Rank* follows. Afterwards the concepts *Age_group_simple_VT*, *Module Format*, *Salary Range Value Type*, *Teaching Unit*, *Value Partition* appear in the list. Once again no statistics could be derived for **wP**, **P** and **I**. The reasons are the same as in the photography ontology. Here it was a little bit surprising, that the concept *Module* was on the top of the list. A detailed look on this concept showed that *Module is a Teaching Unit*. Hence the top ranked concepts fits with the idea of this ontology, though it might be expected from the file name "Simple University-01.owl" that the organizational structure of the university is described. The measures showed that this is not the case. In fact concepts for "Teaching" are described. Also those restrictions referring to external concepts defined in another ontology refer to a resource "Teaching-1-01.daml". Thus the top ranked concept reflects the main concepts of the ontology.

Patient Rheuma Ontology. The last examined ontology was the patient/rheuma ontology. Although it focuses on a special medical domain, it was assumed that there are at least some top ranked notions which are also known to non specialists (e.g. "patient", "rheuma"). Surprisingly, the top ranked concepts are not patient or rheuma as it might be expected from the name of the file ("PatientRheuma.owl"). Instead, applying **wNS** the concepts *Ward*, *Number*, *Hospital*, *Medical Organisation*, *Joint Inflammation* and *Physician* are firstly listed. Applying **wP**, *Transport*, *Flight*, *List of Hospitals, Diagnosis* are the four top ranked concepts. With the **P** measure, the ordering of the first four concepts is slightly different: *Transport*, *Flight*, *Organization*, *and Diagnosis* (resp. *Patient* or *Address* which have the same result for **P** as *Diagnosis)*. The concepts *Gene*, *Diagnosis*, *Address* and *Patient* have instances (**I**) but there are not so many. *Gene* has only 2 instances. The others have only 1 instance. Looking at these results the question arose, why they do not match with the expectations. Why is e.g. *Number* at a very good ranking position in the wNS ranking? Why has *Patient* such a bad ranking? What about "rheuma"? The reasons for that can be found inside the ontology. Everything that is a number (e.g. *Booking Number*, *Credit Card Number*, *Flight Number* etc.) was subsumed to *Number*. *Patient* does not have a deeper substructure. Instead it is a subclass of Person and is also a leaf in the taxonomy tree. The concept *Rheuma* does not exist as such. Instead "*Rheumatoid Arthritis*" is mentioned in the ontology. This once again is a leaf in this local ontology. Whereas *Rheumatoid Arthritis* references another external ontology, *Patient* does not have such a reference. Most surprising was the fact, that in the object properties sections, many properties have the domain class *Flight* or *Transport*. The concept *Patient* is involved only in two object properties as a domain class. Even together with its super class (*Person*) the number of object properties was less than the number of object properties defined for the concept flight. The super class *Disease* can be found in only one object property. A look on the number of restrictions for a patient concept showed that *Patient* was involved in two *someValuesFrom* restrictions. Although it was a surprise, in fact the results gave a good picture of the ontology structure, since the concepts "patient" and "rheuma" are not really described in detail in the local context of the file "PatientRheuma.owl". Instead the ontology engineers focused more on the description concepts like ward, hospital, flight, transport etc. Thus if someone would like to (re)use detailed information about patients, this ontology is not the best one for doing it.

3 APPLICATION SCENARIOS

The results of the measures can be the basis for two application scenarios.

These results can be used to generate a natural language abstract (summary) of the ontology. Strategies how to verbalize ontologies are described in (Fuchs et.al. 2005), (Hewlett et. al. 2005) and (Fliedl et. al 2007). In combination with the described measures a summary can be generated if only main concepts are verbalized.

Another application scenario is the mapping of ontology elements to a conceptual database schema or the support of information systems design (Guarino, 1998), (Sugumaran 2006). Strategies for mappings are described in (Vasilecas et.al. 2005) and (Kalibatiene et.al. 2009). With the measures, a selection of appropriate concepts can be made before the mapping is applied on these concepts. Hence this strategy would consider that the ontology and the future database schema have a different scope and focus, though belonging to the same domain.

4 CONLUSIONS

From the examination of the six ontologies, the following can be learned: Except the instance measure (**I**), the three other measures (**wNS**, **wP**, **P**) give a good first impression about the focus of an ontology structure. This can be observed also in the case of the university and patient/rheuma ontology.

The weighted number of successor measure (**wNS**) can be applied more often than the others, since often taxonomies are used. Nevertheless also the weighted property (**wP**) measure as well as the property measure (**P**) are important. Especially for users who want to re-use an ontology for conceptual modeling, knowledge about the "relationships" between concepts is interesting.

In order to give human readers the ability to examine an ontology according to the measures, a prototype was built (see figure 2 for the screenshot). This prototype also allows browsing through the ontology and it is a basis for the two application scenarios.

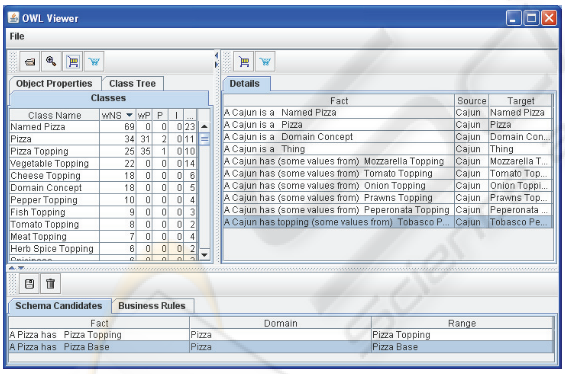

Figure 2: Screenshot of the prototype.

In future also statistics of how often a certain concept appears in a restriction (e.g. *someValuesFrom*, *allValuesFrom* etc.) will be analyzed.

REFERENCES

Bezerra, D., Costa, A., Okada, K., 2009, *SwTO[I] (Software Test Onotlogy Integrated) and its applicaton in Linux Test*. In Proceedings of the 3rd International Workshop on Ontology, Conceptualization for Information Systems, Software Engineering and Service Science, CEUR-WS, Vol 460, http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/, pp. 25 – 36.

Fliedl G., Kop C., Voehringer J., 2007. *From OWL class and property labels to human understandable natural language*. In Kedad Z., Lammari N., Métais E., Meziane F., Rezgui Y. (eds.), Proceedings of the 12th International Conference on Applications of Natural Language to Information Systems, NLDB 2007, Lecture Notes in Computer Science (LNCS), Vol. 4592, Springer Verlag, 2005, pp. 156 – 167.

Fuchs, N.E., Höfler, S., Kaljurand, K., Rinaldi, F. and Schneider, G., 2005. *Attempto Controlled English: A Knowledge Representation Language Readable by Humans and Machines* In Norbert Eisinger N. and Maluszynski, J. (eds.): Reasoning Web, First International Summer School, LNCS 3564, Springer,2005, pp. 213-250.

Guarino, N. 1995. *Formal Ontology, conceptual analysis and Knowledge Representation*. In International Journal of Human-Computer Studies, Vol. 44, Issue. 5-6, 1995, pp. 625 – 640.

Guarino, N., 1998. *Formal Ontology and Information Systems*. In Proceedings of FOIS'98, IOS Press, 1998, pp. 3 – 15.

Hewlett, D., Kalyanpur, A., Kolovski, V., Halaschek-Wiener, C., 2005. *Effective Natural Language Paraphrasing of Ontologies on the Semantic Web*. In End User Semantic Web Interaction Workshop, CEUR-WS Proceedings, Vol. 172, 2005, http://ftp.informatik.rwth-aachen.de/Publications/ CEUR-WS /

Huang, N., Diao, Sh., 2006. *Structure-Based Ontology Evaluation*. In IEEE International Conference on e-Business Engineering (ICEBE06), pp. 1- 6.

Kalibatiene, D., Vasilecas, O., Guizzardi, G., 2009. *Transformation Ontology Axioms to Information Processing Rules – An MDA Based Approach*. In Proceedings of the 3rd International Workshop on Ontology, Conceptualization for Information Systems, Software Engineering and Service Science, CEUR-WS, Vol 460, http://ftp.informatik.rwth-aachen.de/Publications/ CEUR-WS/, pp. 25 – 36.

McGuiness, D.L., van Harmelen F., 2004. *OWL Web Ontology Language Overview*, http://www.w3.org/TR/ owl-features/

Sugumaran, V., Storey, V., 2006. *The Role of Domain Ontologies in Database Design: An Ontology Management and Conceptual Modeling Environment*. In ACM Transaction on Data-base Systems, Vol. 31, No. 3, Sept. 2006, pp. 1064 – 1094.

Vasilecas, O., Bugaite, D., 2005. *Ontology-Based Elicitation of Business Rules*. In A.G. Nilsson, R. Gustas, W.G. Wojtkowski, W. Wojtkowski, S. Wrycza, J. Zupancic, Advances in Information Systems Development: Bridging the Gap between Academia & Industry, Vol. 2, Springer Verlag, Heidelberg, 2005 pp. 795 – 805.