Use of Description Logics Expressive Power in Ontologies

Tomislav Belcar, Sandra Lovrenčić
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
`{tomislav.belcar, sandra.lovrencic}@foi.hr`

Abstract. *Description logics, as a knowledge representation formalism used for ontology development in OWL language, enable very expressive domain modeling. This paper investigates to what extent is this expressive power exploited in practice. Research concentrated on DL languages used as well as on types of classes and number of axioms defined was conducted on selected ontologies. Analysis results and suggestions for further research are presented.*

Keywords. description logics, ontology, knowledge representation

1 Introduction

Ontologies, as one of knowledge representation methods, especially in the Semantic Web development, have been popular in information science for almost two decades [7], although their first mention in the field was in 1980 [10]. First widely used knowledge representation formalisms for ontology development were frames and first order logic [6]. Frames enable easy development of hierarchies and allow subclasses to inherit atributes from superclasses whereas first order logic enables more possibilities in knowledge representation as well as knowledge reasoning.

The most used knowledge representation formalism in ontologies today are description logics (DLs), that are developed from semantic nets and frames [11]. The term is used in plural because there are many description logic (DL) languages that were developed after initial "terminological logic" proposal over 20 years ago [2]. DLs reached their full potential for ontology development as a chosen knowledge representation formalism for their development language OWL – Web Ontology Language (`http://www.w3.org/standards/techs/owl#w3c_all`).

DLs expressive power depends on DL language used, enabling more or less of domain characteristics to be described. Also, knowledge reasoning can be performed according to types of axioms that describe domain knowledge. Although those are assumptions for knowledge reasoning as important as reasoning systems used, research has been conducted primarily about the later – reasoning methods [8] and quality of reasoners [5][9].

Recently, with purpose of obtaining more efficient reasoning, especially for large-scale ontologies, the development DLs is increasingly turning to lightweight DLs with limited expressivity [1]. This fact and little information about practical use of DLs expressive power were motivation for this research. Therefore, the aim of this paper is to show to what extent is DLs expressive power used in ontology development in practice.

2 Description logics in ontologies

Use of DLs as a knowledge modelling language for ontologies is a standard and there are many DLs that have diferent expressive power. First DL language defined is called *AL* (Attributive Language) [13] and the smallest DL proposed to be used for basic knowledge reasoning is *ALC*. The purpose of this paper is not to describe various DL languages; their sintax and semantics are nowadays well known and documented – for example, sintax and semantics of *ALC* can be found in various literature [2][3][6]. Because expressive power of DL languages used in ontologies is a subject of this research, DL operators that construct DL languages as well as class definitions with DL axioms are explained in more detail.

2.1 Primitive and defined ontology classes

Knowledge is in DLs represented with three components: concepts (classes of objects), roles (binary relations between objects) and individuals (concrete objects). Knowledge base is divided into two parts [11]:

- TBox – contains terminological (intensional) knowledge for description of the structure of the domain and therefore consists of concept and role definitions and describes their hierarchical relations (recently was proposed that roles should be defined separately in Rbox [12]);

- ABox – contains assertional or extensional knowledge, specific to individuals of the domain.

Descriptions in TBox consist of two kinds of terminological axioms [4][11] – inclusion (subsumption) axioms and equality axioms.

Inclusion (subsumption) axioms (represented with μ) define necessary conditions for an object to be instance of some concept (class), defining that if a certain individual is a member of a certain class, then it is necessary for this individual to fulfill given conditions. For an object it can not be determined to be an instance of a certain class, unless explicitly stated. Inclusion axioms describe subsumption or is-a relationships while concepts defined in this way are called *primitive*. For example, CμD means that concept C is subsumed by concept D or that C is-a D.

Equality axioms (represented with ≡) define both necessary and sufficient conditions for an object to be instance of some concept (class), defining that if a certain individual is a member of a certain class, then it is necessary for this indivivual to fulfill given conditions, and that for any other individual that fulfills these conditions, this is sufficient for such individual to be considered as a member of that class. Concepts defined with equality axioms are called *defined* and enable performance of more reasoning tasks. For example, C ≡ D + E means that C is equal to union of D and E.

Each class can have several both inclusion and equality axioms. The simplest inclusion axioms are those that define direct subsumption of one class according to another (class hierarchy).

2.2 Description logic languages

After first proposition of *AL* [13] many DL languages have emerged, each of them enabling description of several concept (class) and role (property) characteristics. New languages are developed using various combinations of DL operators. Languages have names that are consisting of letters that describe operators allowed in them [11][12]:

- *AL* – base attributive language that allows atomic negation (negation of concepts on right side of axioms), concept intersection, universal restriction and limited existential quantification;
- FL^- – *AL* sublanguage, without atomic negation;
- FL_0 – FL^- sublanguage, without existential quantification;
- *C* – complex concept negation;
- *S* – *AL* and *C* with transitive properties;
- *H* – role hierarchy;
- *O* – nominals (enumerated classes or object value restrictions);
- *I* – inverse properties;
- *N* – cardinality (number) restrictions;
- *Q* – qualified cardinality (number) restrictions;
- *F* – functional properties;
- *E* – full existential quantification;
- *U* – concept union;
- *R* – limited complex role inclusion axioms, reflexivity and irreflexivity, role disjointness;
- *(D)* – use of datatype properties, data values or datatypes.

Because union and full existential quantification can be obtained using negation, if the language allows all three, writing of *E* and *U* can be omitted, meaning that *ALC* is always written instead of *ALCUE*. As another example, one of most often used languages, *SHOIN* denotes *ALC* with transitive properties, nominals, inverse properties and cardinality restrictions.

3 Description logics expressive power use research

Knowing that word ontology can be used for domain descriptions ranging from simple taxonomies to complex models using formal logic, the research goal was to find out to what extent DLs are used for domain modeling. As it can be seen from previous chapter, main concern was expressive power according to DL operators used and also use of primitive and defined concepts or classes in ontologies. Concretely, the purpose was to discover following for selected set of ontologies:

- DL operators and DL languages used;
- number of defined and primitive classes;
- percentage of defined classes according to total number of classes in ontologies;
- number of primitive classes for each defined class (where applicable, because it was possible that some ontologies do not have defined classes);
- number of axioms for primitive classes in total and for each class.

Protégé tool (`http://protege.stanford.edu`) is widely known and used and represents an excellent platform for ontology development. It also has ontology library with (currently) 93 OWL ontologies for various domains developed by Protégé users. Therefore, it was an ideal ontology collection for this research. Of total number of ontologies 49 could not be loaded into new version of Protégé. Pizza ontology is aimed at learning Protégé and had to be removed as well as general upper Basic Formal Ontology. A certain number of Protégé library ontologies provide also links to multiple ontologies of which only those substantially different from each other were added to research set – in total 21. At the end, 63 ontologies were chosen for testing.

3.1 Use of description logic operators

Ontologies that were tested during research were developed with very broad range of DL languages, starting from basic *AL* language. The most used operator was *(D)*, meaning that almost all ontologies (55) used datatype properties, data values or datatypes. Object properties (those that connect two individuals) in ontologies are used always but it is known that datatype properties (those that connect individuals to data values) are sometimes omitted as unnecessary for domain description. Also, two thirds of ontologies use inverse properties (*I*). 36 ontologies use *AL*, but it is base DL language, and it should not be taken into consideration. 27 ontologies that do not use *AL*, use *S*, as expected, because it is *ALC* with transitive properties. All ontologies should use base language, so it should only be noted that almost half of ontologies use transitive properties. In addition to that, almost half of ontologies also use nominals (*O*) and cardinality restrictions (*N*). The diagram of operator usage can be seen in Fig. 1.

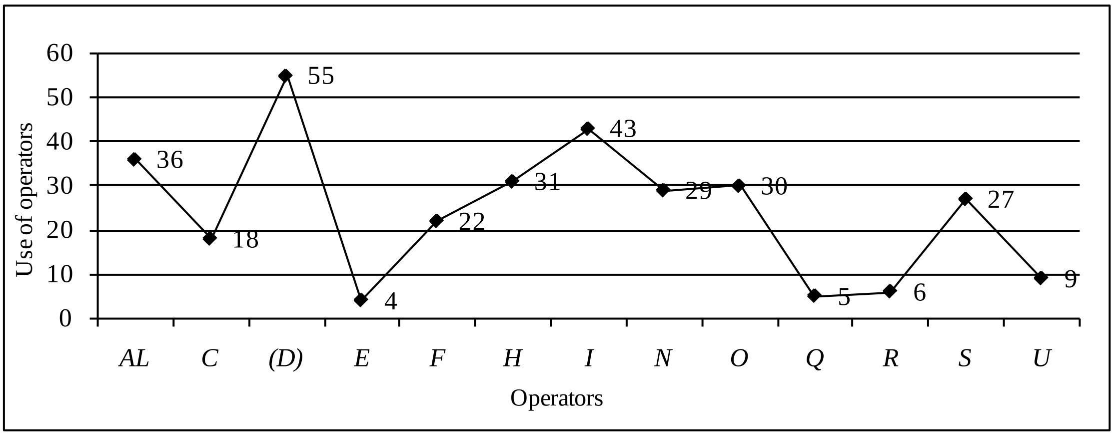

Figure 1. Use of operators in tested ontologies

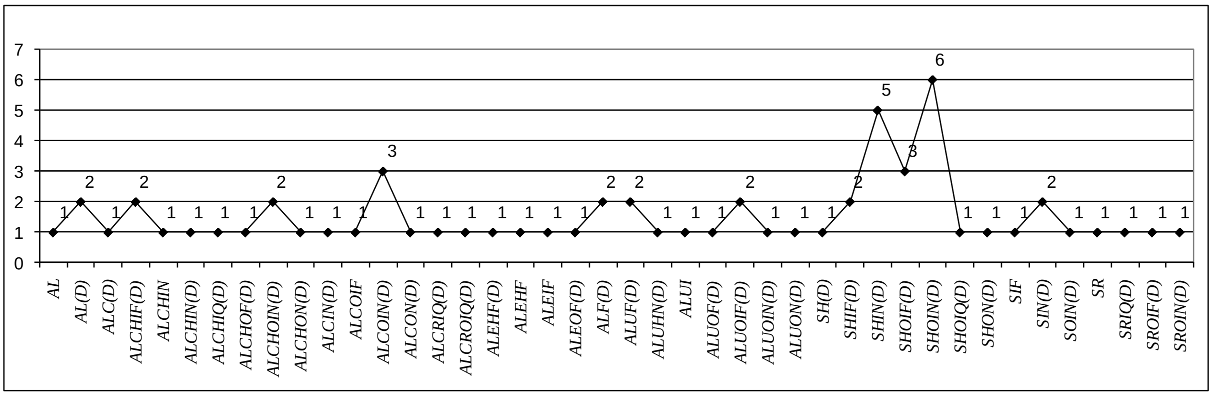

Figure 2. Description logics used in tested ontologies

Table 1. Number of operators used in ontologies

Number of operators	Number of ontologies
0	1
1	2
2	3
3	4
4	6
5	4
6	8
7	10
8	13
9	12

As it can be seen in Fig. 2, DL languages used are widely distributed. This means that ontologies tested actually range from very simple to complex ones that use various DL operators. The most used DL languages are *SHIN(D)* and *SHOIN(D)* and popular reasoning systems are generally based on the later.

In Table 1 it is shown how many operators ontologies use. When operators were counted, *C* was included as three operators (concept negation, concept union and full existential negation) and *S* as four (*C* with transitive properties). Single ontology uses only base language *AL*. Average number of operators for individual ontology is 6,2540, with median 7 and mode 8. It is obvious that at least half of ontologies have 7 or more operators and that 90% of them use at least 3. Standard deviation value of 2,4228 shows dispersion in distribution and skewness of -0,7711 implies that most values are high with asymmetry towards several low values. For better insight into expressive power of DLs used, classes and axioms in ontologies were also analyzed.

3.2 Use of description logic axioms

Although use of defined classes enables knowledge reasoning to its full extent by allowing all deductive reasoning tasks – knowledge base satisfiability, axiom entailment, concept satisfiability, instance retrieval, classification and conjunctive query answering [12] – to be performed, this is not a goal for all ontologies developed. Ontologies can be used only to provide general domain model, without need for further formalization and therefore without need for defined classes. The analysis of tested ontologies showed that 27% percent of ontologies (17) do not have defined classes. In Table 2 analysis of all relevant metrics for use of description logics axioms is presented.

Table 2. Analysis of use of description logic axioms

Criteria	Raw values		Statistical measures				
	Minimum	*Maximum*	*Average*	*Median*	*Mode*	*Standard deviation*	*Skewness*
Total number of classes	2	3956	465,6508	113,0000	11,0000	875,8426	2,7225
Number of defined classes for all ontologies	0	1028	51,5556	5,0000	0,0000	146,4870	5,2979
Number of defined classes only for ontologies with defined classes	1	1028	70,6087	15,5000	3,0000	167,8980	4,5612
Number of primitive classes	1	3745	414,0952	90,0000	11,0000	779,3174	2,7715
% of defiined classes according to total number of classes	0	94,4444	0,1206	0,0533	0,0000	0,1790	2,5359
Number of primitive classes for each defined class (46 ontologies)	0,0588	1929	59,9298	8,8929	15,0000	17,8309	6,7050
Total number of axioms in primitive classes	0	5083	647,1905	188,0000	0,0000	1188,8603	2,6076
Number of axioms for each primitive class in ontology	0	6,1250	1,5948	1,3268	1,4545	1,0651	1,9292

Total number of classes in ontologies ranged from 2 to 3956, number of defined classes from 0 to 1028 (from 1 to 1028 for ontologies with at least one defined class) and number of primitive classes from 1 to 3745, which is a very broad range. Therefore, minimum, maximum and several statistical measures were analyzed according to criteria presented in Table 2. Broad range of total, defined and primitive classes in individual ontologies obviously shows that analyzed values are not distributed evenly. Although maximum values are very high, median and mode show that most of ontologies actually are not large. Substantial differences in values are confirmed with very high values for standard deviation. Skewness results also show asymmetry in distribution, especially for defined classes, both for all ontologies and for only those with at least one defined class.

It is already determined that 27% of ontologies have no defined classes, which is the most often value, according to mode (when analyzing all 63 ontologies). Median shows that there is a small number of defined classes in tested ontologies, which is also confirmed with all statistical measures for percentage of defined classes according to total number of classes. To obtain relevant results only ontologies with at least one defined class were tested separately. This analysis confirmed low number of defined classes in ontologies (with median 15,5 and mode 3), as well as large dispersion and asymmetry. According to actual raw data, only seven ontologies have more than 1000 classes and only eight of them have more than 100 defined classes. Therefore, several extremely high numbers (positive skew) influence on these results.

Number of primitive classes for each defined class (in 46 ontologies that have defined classes) also has large range. Statistical measures show definitely the largest asymmetry in distribution, but in half of ontologies there is 9 or less primitive classes for each that is defined. Total number of axioms in primitive classes shows biggest dispersion of values which is evident from median, mode and especially standard deviation results. When those values are considered for individual ontologies, as average number of axioms for each primitive class, common result is more evenly distributed. Average number of 6 axioms for each primitive class is fairly reduced with median and mode results. It should be taken into consideration that all axioms were counted, meaning that there is a substantial number of the simplest subsumption axioms that define direct subclasses (class hierarchy) along with those additionally created for better class description.

4 Conclusions and future research

Analysis of ontologies from Protégé library according to use of DLs expressive power showed that various

DL languages are used and that ontologies vary greatly in their size and number of defined classes and axioms created. It can be concluded that:

- ontologies use very different combinations of DL operators, ranging from basic *AL* language to languages that use almost all operators (9 as a maximal number);
- most of ontologies use large part of DLs expressive power, because more than half of them use at least 7 operators, two thirds (68%) at least 6 and 90% at least 3;
- 73% of ontologies have defined classes, but their number in ontologies that use them is generally not large, because in half of them their number is at most 15 or 16;
- wide distribution and asymmetry show that in various domains various DL languages with various number of axioms are used, meaning that all research conclusions can be considered only as general guidelines.

According to results, majority of the domains are well modeled and large potential of knowledge reasoning with DLs is enabled. It is obvious that smaller number of ontologies was developed with aim mostly to present hierarchical structure of the domain.

According to research results, several suggestions for further research can be made:

- to repeat research on several other ontology libraries where presumably different application interfaces were used for ontology development;
- to evaluate research criteria and propose improvements that can give more conclusive results which can be evaluated with testing the same set of ontologies;
- to include descriptions of instances created in ontologies for additional insight into DLs expressive power use for concrete practical ontologies;
- to explore whether smaller expressivity is connected with larger ontologies that have lots of data, following increased interest in lightweight DLs .

Although presented research shows positive results of DLs usage when modeling domains with ontologies, more extensive research can be conducted, especially concerning current trends in development of DLs expressive power and their use in practice. A general framework for ontology expressivity with respect to DL languages can be final research step.

References

[1] Baader F: **What's new in Description Logics**, Informatik-Spektrum, Vol. 34, No. 5, Springer-Verlag, Heidelberg, Germany, 2011, pp. 434-442.

[2] Baader F et al.: **Terminological Knowledge Representation: A Proposal for a Terminological Logic**, Proceedings of the International Workshop on Terminological Logics, DFKI-D-91-13, May 6-8, Dagstuhl, Germany, 1991, pp. 120-128.

[3] Baader F, Nutt W: **Basic Description Logics**, In (Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (eds.)): The Description Logic Handbook, 2nd Edition, Cambridge University Press, Cambridge, England, 2007, pp. 43-95.

[4] Calvanese D, Lenzerini M, Nardi D: **Description Logics for Conceptual Data Modeling**, In (J. Chomicki J, Saake G (eds.)): Logics for databases and information systems, Kluwer Academic Publishers, Dordrecht, Netherlands 1998., pp. 229 – 263.

[5] Dentler K, Cornet R, ten Teije A, de Keizer N: **Comparison of reasoners for large ontologies in the OWL 2 EL profile**, Semantic Web, Vol. 2, No. 2, IOS Press, Amsterdam, Netherlands, 2011, pp. 71-87.

[6] Gómez Pérez A, Fernández-López M, Corcho O: **Ontological Engineering**, Springer, Berlin, Germany, 2004.

[7] Gruber T R: **A Translation Approach to Portable Ontology Specifications**, Knowledge Acquisition, Vol. 5, No. 2, Elsevier, Amsterdam, Netherlands, 1993, pp. 199-220.

[8] Hitzler P, van Harmelen F: **A reasonable Semantic Web**, Semantic Web, Vol. 1, No. 1-2, IOS Press, Amsterdam, Netherlands, 2010, pp. 39-44.

[9] Lee C, Park S, Lee D, Lee J-W, Jeong O-R, Lee S-G: **A Comparison of Ontology Reasoning Systems Using Query Sequences**, Proceedings of the 2nd international conference on Ubiquitous information management and communication, January 31- February 1, Suwon, Korea, 2008, pp. 543-546.

[10] Øhrstrøm P, Andersen J, Schärfe H: **What Has Happened to Ontology**, Proceedings of the 13th International Conference on Conceptual Structures - Conceptual Structures: Common Semantics for Sharing Knowledge, July 17-22, Kassel, Germany, 2005, pp. 425-438.

[11] Nardi R, Brachman R J: **An introduction to Description Logics**, In (Baader F, Calvanese D, McGuinness D, Nardi D, Patel-Schneider P (eds.)): The Description Logic Handbook, 2nd Edition, Cambridge University Press, Cambridge, England, 2007, pp. 1-40.

[12] Rudolph S: **Foundations of Description Logics**, In (Polleres A et al. (eds)):, Reasoning Web. Semantic Technologies for the Web of Data - 7th International Summer School 2011, Springer, Berlin, Germany, 2011, pp. 76-136.

[13] Schmidt-Schauß M: **Attributive concept descriptions with complements**, Artificial Intelligence, Vol. 48, No. 1, Elsevier, Amsterdam, Netherlands, 1991., pp. 1-26.