Dynamic Ontology Evolution in Open Environments

I. Palmisano, V. Tamma
Dept. of Computer Science
University of Liverpool
L69 3BX UK
{I.Palmisano, V.Tamma}
@liverpool.ac.uk

L. Iannone
Dept. of Computer Science
University of Manchester
M13 9PL UK
iannone@cs.man.ac.uk

T. R. Payne, P.Doran
Dept. of Computer Science
University of Liverpool
L69 3BX UK
{T.R.Payne, P.Doran}
@liverpool.ac.uk

Introduction

Interoperation between knowledge-based systems or agents requires common ontologies to facilitate successful information exchange. However, the openness of the Semantic Web means that the notion of there being common domain ontologies sufficient to cater for the requirements of a diverse range of consumers and producers of services has become untenable. In these types of environments it is necessary to consider that no ontology can be expected to remain unchanged throughout its lifetime. However, the dynamism and the large scale of the environment prevent the use of traditional ontology evolution techniques, where changes are mediated by a knowledge engineer [3]. We argue that the ability to estimate the impact of change *a priori*, i.e. before performing the change itself, is crucial, since this estimate can be used to assess the usefulness of the change. We assume that agents are capable of rational behaviour, and that they decide whether to change the ontology they commit to if the cost of the change (in terms of reclassification of knowledge) is offset by the benefits derived from the ability of a system to acquire new capabilities and therefore to achieve new tasks (or answer new queries, in the case of knowledge based systems). However, the agent's decision making process follows the principle of *bounded rationality* [5]: agents operate with limited computational resources, and with partial knowledge of the environment [4]. We present an approach that evaluates the impact of change on an ontology *a priori*, without using reasoning, by estimating which set of axioms in an ontology is impacted by the change.

Change Evaluation

OWL DL ontologies may undergo modifications for many different reasons; the simplest cases are those in which new assertions are added, or new axioms further detailing the domain knowledge. Another important source of modifications for an ontology comes from alignment with another ontology; the problem of ontology alignment has received wide interest in recent years, and many alignment methodologies and systems have been developed [2], together with methods to evaluate the results of different alignment techniques[1].
Not much attention has been devoted to the way a knowledge base should cope with the additions introduced by alignments to another ontology; in an open environment this may cause the knowledge base to grow without control. The way a knowledge base can be affected by alignments to another ontology is illustrated by the following example:
Two agents, A_O and A_K, commit to two different ontologies O and K, which differ on the definition of a specific concept, named C_2 in O, and which is not present in K. By means of an alignment technique, a concept C_1 in K is discovered as the best match for C_2, with the relation between them being an axiom $Ax = C_1 \subseteq C_2$. This enables agent A_O to issue queries against the knowledge base of agent A_K. By asking for instances of C_2, the result will contain all the instances of C_1; this has the clear advantage of returning valid results to A_O, while issuing the same query without previous alignment would return no results. K may be affected by the alignment: let us consider the case in which the definition of C_2 (Def_{C_2}) is added to K, with Ax acting as a connection between Def_{C_2} and the original K, as depicted in Figure 1; the impact this has on the future performance of K cannot be easily foreseen. The simplest heuristic is based on the assumption that $K \sqcup Ax \sqcup Def_{C_2}$ will behave in the same way as its expressivity class. K has expressivity $\mathcal{AL}$; adding Ax and Def_{C_2} changes the expressivity of the knowledge base to $\mathcal{ALC}$. In a larger knowledge base, such an increase in expressivity may be restricted to a portion of the knowledge base. This is true for real world ontologies. The impact of a change is measured using the very simple heuristic of the number of axioms and assertions contained in the section of the knowledge base affected by the change, weighed with the expressivity of the knowledge base. An *axiom* A in the framework represents a DL axiom; let us give an example referring to Ax:
A represents the abstraction over an OWL axiom or assertion, which has a *signature*, and is represented as a node in a directed multigraph whose edges represent *relations* between *axioms*; it has a *signature* s which is the set of named and unnamed concepts and roles mentioned in an *axiom*; A has $s = \{C_1, C_2\}$, and a *main concept* or *main role* m (also called *main node*) which is the concept or role being defined by the *axiom*[2]; m_A is C_1 in the example.
Relations between *axioms* correspond to intersections be-

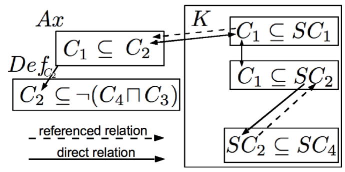

Figure 1: The group rooted at C_1 contains the whole knowledge base K

[1] `http://www.ontologymatching.org/evaluation.html`

[2] An OWL axiom is represented internally as a set of RDF statements; the standard RDF mapping for OWL is defined in `http://www.w3.org/TR/owl-semantics/mapping.html`

Ontology	Expressivity	Groups	Size range
Galen full	$\mathcal{AL}$	72	1/45
# axioms:	$\mathcal{ALC}$	14303	61723/63625
82030	$\mathcal{ALHIF}$	28	13/27
# groups:	$\mathcal{ALR}+\mathcal{HI}$	1	1
37408	$\mathcal{ALR}+\mathcal{HIF}$	313	53/1739
avg size:	$\mathcal{SHIF}$	24711	93/81815
9649			
Galen fragment	$\mathcal{AL}$	552	1/115
# axioms:	$\mathcal{AL}(\mathcal{D})$	441	3/255
9915	$\mathcal{ALC}$	101	3/151
# groups:	$\mathcal{ALC}(\mathcal{D})$	1257	4/7576
4500	$\mathcal{ALCHF}(\mathcal{D})$	95	16/7904
avg size:	$\mathcal{ALCH}(\mathcal{D})$	29	20/7370
8529	$\mathcal{SHF}(\mathcal{D})$	427	47/9889

Table 1: Expressivity Metrics

tween their *signatures* and *main nodes*; they can be: **direct**: a *direct* relation between an *axiom* A and an *axiom* B exists if m_B belongs to the s_A; two *axioms* with the same m have a bidirectional *direct* relation; **indirect**: an *indirect* relation between two *axioms* A and B holds when s_A overlaps with s_B (e.g. $D_1 \subseteq \exists\ R.C$ and $D_2 \subseteq \exists\ R.D$ share a reference to the role R); an *indirect relation* is bidirectional; **referenced**: a *referenced* relation is the inverse of a *direct* relation; such a relation is implicitly defined also for *indirect* relations, in which case it is bidirectional as the *indirect relation*. The relations define three graphs: called O the set of *axioms* in $\mathcal{O}$, O_d is the graph $< O,\ DR >$ where DR is the set of *direct* relations; O_i is the graph $< O,\ IR >$ where IR is the set of *indirect* relations; O_r is the graph $< O,\ RR >$ where RR is the set of *referenced* relations.
We define a *Group G rooted at an axiom A* as the set of *axioms* resulting from the union of the sets of axioms S_d, S_i and S_r explored during the exhaustive visit of O_d, O_i and O_r respectively, starting from A and following the relations.

Evaluation

The grouping framework is implemented in Java; it uses Jena [1] and the SPARQL [3] implementation ARQ[4] to perform the *axiom* extraction. The OWL DL reasoning engine Pellet [6] has been used to check the expressivity of a group. Here the results for the Galen ontology translated to OWL[5] and a fragment of Galen[6], smaller than the original and often used to test reasoners, are reported. The measures taken into account are: the number of axioms in the ontology, the average number of axioms in a group and the expressivity of each group; for each expressivity level, the number of groups (duplicate groups are counted as one) and the size range are reported in Table 1. Table 2 presents in more detail the results obtained on the Galen fragment ontology. Out of 9915 *axioms*, more than 4500 groups were computed, by selecting only *axioms* with named concepts as m. The groups overlap very often, due to the high detail of the ontology, which is has a deep role hierarchy. In the following, Int will be the intersection of the 5 largest groups, labelled G_2 to G_6, and will denote $G_i \setminus Int$ the set difference between one of these groups and Int. Only two groups are significantly larger than Int, G_5 and G_6, by 1923 and 2420 axioms respectively. Table 2 reports the size of Int; for each set, then, the size of the difference $G_i \setminus Int$ and its expressivity is reported. The expressivity of the groups is $\mathcal{ALC}$ for the four smaller groups, and $\mathcal{SHF}$ for the two larger ones; the difference in expressivity is therefore related to roles.

[3] http://www.w3.org/TR/rdf-sparql-query/
[4] http://jena.sourceforge.net/ARQ/
[5] http://www.co-ode.org/galen/
[6] http://www.daml.org/ontologies/400

	# axioms	Expressivity
Int	7469	$\mathcal{ALC}$
$G_2 \setminus Int$	32	$\mathcal{AL}$
$G_3 \setminus Int$	19	$\mathcal{AL}$
$G_4 \setminus Int$	34	$\mathcal{AL}$
$G_5 \setminus Int$	1923	$\mathcal{SHF}$
$G_6 \setminus Int$	2420	$\mathcal{SHF}$

	$I(A)$ impact value
$A \in Int$	47089.25
$A \notin Int$: $A \in G_2$	6086.75
$A \in G_3$	6081.75
$A \in G_4$	6088.75
$A \in G_5$	14252.00
$A \in G_6$	14580.00

Table 2: Expressivity of group overlaps and results for impact computation

If a change to an ontology is proposed (by the agent's decision making process), we can now assess which part of the ontology the change will impact. For example, if the removal of an *axiom* A from the Galen fragment is the proposed change, the agent can assess if this axiom is located at the intersection of the five largest groups, or if it is located within one specific group. In order to estimate the impact of the change, we propose an impact function that is computed considering the size of the distinct groups containing A:

$$I(A) = \sum_{A \in G} size(G) * expr(G)$$

$size()$: Given $\mathcal{G}$ the set of all groups G over $\mathcal{O}$, $size(G) : \mathcal{G} \rightarrow \mathbb{N}$ is the number of *axioms* contained in G.
$expr()$: Given (G) the set of all groups G over $\mathcal{O}$, $expr(G) : \mathcal{G} \rightarrow \mathbb{R}$ is the function that computes ec, the DL expressivity of G, and maps ec into a real numeric value.
The values for $I(A)$ are reported in Table 2; on this basis, the decision making process may choose to accept or reject the change involving A.

Acknowledgements The research presented in this paper is supported by the EPSRC grant EP/D064287/1

1. REFERENCES

[1] J. J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson. Jena: implementing the semantic web recommendations. In *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters, New York, NY, USA, May 17-20*, pages 74–83. ACM, 2004.

[2] J. Euzenat and P. Shvaiko. *Ontology Matching.* Springer-Verlag, Heidelberg (DE), 2007.

[3] D. Fensel and F. van Harmelen. Unifying reasoning and search to web scale. *IEEE Internet Computing*, 11(2):93–95, 2007.

[4] T. R. Payne. Web services from an agent perspective. *IEEE Intelligent Systems*, 23(2):12–14, 2008.

[5] H. Simon. *Models of Bounded Rationality*, volume 3. MIT Press, 1997.

[6] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A practical OWL-DL reasoner. *Journal of Web Semantics*, 5(2), 2007.