Simplicial Complexes for Description Logics Terminologies

Erika Valencia and Jean-Paul Sansonnet
LIMSI-CNRS, Université Paris-XI
B.P.133, F-91403 Orsay cedex, FRANCE
{erika, jps}@limsi.fr

Abstract

When querying description logics knowledge bases, one may use free variables, called patterns, to refer to the concept or role that we query about [3]. But what if the queries contain also heterogeneous terms ? We propose to solve the so-called *terminological heterogeneity* by using adaptation algorithms in the place of pattern matching algorithms. To apply our adaptation algorithms to the description logics KBases, we explore the possibility of giving an interpretation of a particular terminology in terms of topological structures : simplicial complexes.

Key Words : Description Logics Knowledge Bases, Simplicial Complexes, Adaptation.

1 Introduction

Multi-Agents Background Our study takes place into a project dealing with web pages communication via Internet. We model these pages as agents having their own knowledge representation and reasoning capabilities. The communication between such pages is then assumed to be *dialogical*, where *dialogism* is roughly a simplified subset of natural language. The agents being designed by different users on different purposes a sematical heterogeneity arise among them during an interaction.

Assumptions for a Formalization in Description Logics We model the web pages as DL KBases. Since we consider the dialogical interaction between such Kbases, we model each interaction step as a query managment by the receiver Kbase . In this context, we define what we will call semantical heterogeneity between such terminologies[1]. Note that it appears that Description Logic (DL) is very well suited to model query managment in the database field and for semantical heterogeneity problems [4], [8].

A Problem of Adaptation Since a query to a DL Kbase contains free variable patterns, the classical query management functions make use of more or less sophisticated matching tools. But in the context of semantically heterogeneous agents, the queries may also contain heterogeneous terms. These terms are not explicit patterns and should not be matched using the same tools : we propose to match them using the notion of *adaptation*. More precisely, we use known terms from the local receiver terminology and constraints on the heterogeneous term deduced from the query to adapt the unknown term into a known one, before the classical application of matching tools.

2 Algebraic Topology for Knowledge Representation

Simplicial Complexes are a structure defined in algebraic topology that has already been proposed for representational purposes. Indeed, Atkin proposed in 1977 to use them to represente binary relations, and analyze the resulting representation of a situation, through a process that he called Q-analysis.

2.1 Q-Analysis

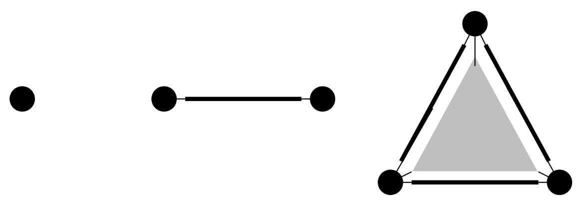
2.1.1 Simplicial Complexes

Simplicial Complexes (SC) are topological abstract structures that generalize the notion of graph. Indeed, all complexes of dimension less than 2 are graphs, and are thus interesting for computationnal purposes. The figure 1 shows the geometrical realization of simplexes of dimension resp. 0, 1 and 2 : they correspond to a vertice (point), an edge (segment) and a face (triangular). The face cannot be reduced to a portion of graph without loosing topological properties, while the simplexes of dimension 0 and 1 can.

An abstract simplicial complex is a couple (V, K) where V is a set of elements called vertices and K is a set of finite parts of V such that if $s \in K$, then all the parts $s' \subseteq s$ belongs also to K. The elements of K are called abstract simplexes.

[1]Article submitted to IBERAMIA 2000.

Figure 1: Simplexes of dimension 0 (point), 1 (line) and 2 (triangular face).



(a) dim 0 (b) dim 1 (c) dim 2

λ	a_1	a_2	a_3
b_1	1	0	0
b_2	0	1	1
b_3	1	1	0

Table 1: Incidence matrix associated with λ.

The dimension of a simplex s is equal to $Card(s) - 1$. The dimension of the complex is the dimension of its biggest simplex.

A p-simplex s is noted : $s = \langle v^0 v^1 ... v^p \rangle$, where $v^i \in V$. A complex is a set of sets closed for the inclusion and the intersection. Thus, simplicial complexes are particularly attractive to generalize semantic networks by keeping the possibility to express hierarchies like in a relational graph [9].

2.1.2 Representation of a Binary Relation

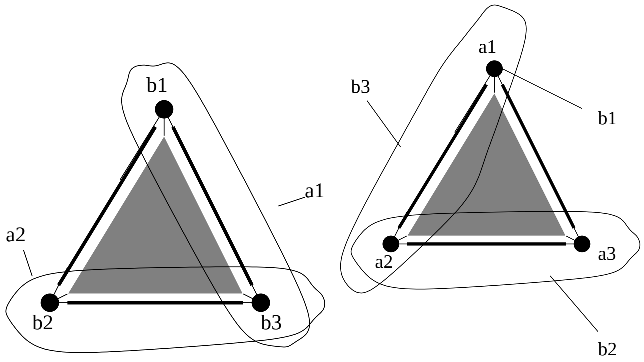
Atkin [2] proposed to represent a binary relation λ between two sets with a simplicial complex.

Let Λ be the incidence matrix of a binary relation $\lambda \subset A \times B$. Let $a \in A$, and the set B_a of $b_i \in B$ such that $(a, b_i) \in \lambda$. The set B_a can be directly read from Λ, as the a-column (see table 1).

We represent the elements b_i of B_a as vertices and a as a simplex build on these vertices. The dimension of the simplex S_a representing a depends on the number of vertices in B_a.

The whole matrix Λ can then be represented as a simplicial complex containing all the simplexes representing each element $a_i \in A$, we note it $K_A(B, \lambda)$ (see figure 2 left).

Figure 2: Simplicial representation of $\lambda : A \to B$ and its dual.



Likewise, we can represent Λ^{-1} with the dual simplicial complex $K_B(A, \lambda^{-1})$ (see figure 2 right). In this case, the elements a_i are taken as vertices and the elements b_i are represented as simplexes. We say that $K_A(B, \lambda)$ and $K_B(A, \lambda^{-1})$ are conjugates, they contain the same information but presented it in a complementary way.

2.1.3 Representation of a Set of Predicates

We extend this method to represent sets of predicates as simplicial complexes too. The idea, which is simple, is to take a set of predicates $P = \{p_1, p_2, ..., p_n\}$ and represent the binary relation $\lambda \subset A \times P$ such that $(a_i, p_j) \in \lambda$ if $p_j(a_i)$ holds.

Two simplexes that have a smaller k-simplex in common are said to share a k-face. In terms of representation, this means that *they have k features in common.* As Freska emphasized it, we call here for the use of discriminating features rather than for precise characterization in terms of universally applicable reference system [5].

We can say that the identity of an element is represented by the features he shares with others and also by the ones that are specific to it [7].

3 Simplicial Semantics for Description Logics Terminologies

3.1 Language Used

We use the syntax of the basic AL description logic, for simplicity since its expressivity is sufficient for the examples we will use. The *syntax* of concepts and roles is briefly presented on table 2. Table 3 also presents the denotational

Table 2: Syntax of AL

Concept	::=
	ConceptName |
	[*all role Concept*] |
	[*atleast role Concept*]
	[*atmost role Concept*]
	[*and Concept Concept*]
	[*not Concept*]
ConceptName	::= DefinedConceptIdentifier |
	PrimitiveConceptIdentifier
Role	::= RoleIdentifier
Ind	::= IndividualIdentifier

Table 3: Semantics of AL

Concept	Denotation
primitive A	A^I
[*all p C*]	$\{d \in D^I \mid p^I(d) \subseteq C^I\}$
[*atleast n C*]	$\{d \in D^I \mid \|p^I(d)\| \geq n\}$
[*atmost n C*]	$\{d \in D^I \mid \|p^I(d)\| \leq n\}$
[*and C D*]	$C^I \cap D^I$
[*not C*]	$\{d \in D^I \mid d \in D^I \setminus C^I$
Top	D^I
Bottom	$\emptyset$

semantics of AL using the notion of valuation I that assigns a subset of a domain D^I to primitive concepts, and a subset of $D^I \times D^I$ to primitive roles.

3.2 Simplicial Semantics

Let D be the domain of interpretation of all the descriptions, the semantics of one of them is then the validity domain of its expression. Theelements of the domain D are usually represented as points : we will represent them as vertices in the CAT formalism.

The constants and variables are usually interpreted as individuals of the world with or without identifiers attached to them, and represented as points. Likewise, in the CAT formalism, they are represented with 0-simplexes (vertices) of the abstract algebraic space.

3.2.1 Quantifications

The Universal Quantification $\forall x P(x)$, denotes the validity domain $D_P = D$ for the predicate P. The modelization of this formula in the CAT formalism corresponds to the maximal simplex including all the vertices of the domain D. We note this maximal simplex S_D.

The Existential Quantification $\exists x P(x)$ denotes a single individual, of undeterminate identifier. Its modelization in terms of CAT corresponds to a single undeterminate vertice.

3.2.2 First Order Predicates

DL concepts are described using n-ary predicates. Let us give an interpretation for n-ary predicates.

Let P be of arbitrary arity n ; its validity domain is a set composed of n-uples. Its representation in terms of CAT is then a complex composed of the simplexes representing r each n-uple of D_P. For example, let $D = \{a, b, c, d\}$, and let the 2-ary predicate R have the validity domain $D_R = \{(a, c), (a, b)\}$, its simplicial representation will then be $S_R = \langle\langle a, b\rangle, \langle a, c\rangle\rangle$, which are two connex edges.

3.2.3 Complementarity degree between two complexes

To represent composed formulae containing FOL constructors, we need to introduce the complementarity operator between simplicial domains.

Let the interpretation domain be $D = \{a, b, c, d\}$ and let the validity domain of the 1-ary predicate P, be $D_P = \{a, b, c\}$. Let us examine the complementary complex $D \setminus D_P$.

First, the maximal complex representing D is given by : $S_D = \langle a, b, c, d\rangle$ with the standard notation (see section 2.1.1) . It means, that the following faces are part of it : $S_D = \langle\langle a, b, c, d\rangle, \langle a, b, c\rangle,\langle b, c, d\rangle, \langle a, b, d\rangle, \langle a, c, d\rangle, \langle a, b\rangle,$

$\langle a, c\rangle, \langle a, d\rangle, \langle b, c\rangle, \langle b, d\rangle,\langle c, d\rangle, \langle a\rangle, \langle b\rangle, \langle c\rangle, \langle d\rangle\rangle$.

First, we can take the complementary simplex by simply removing the face $\langle a, b, c\rangle$ corresponding to the simplex S_P representing D_P (since P is here of arity 1). But, according to the definition of a complex, it has to contain all its faces, so we cannot keep the volume $\langle a, b, c, d\rangle$ if one of its face is missing. In consequence, we also have to remove it from S_D. Thus, the complementary is simply given by :

$$S^1_{\neg P} = \langle\langle b, c, d\rangle, \langle a, b, d\rangle, \langle a, c, d\rangle\rangle$$

Note that this expression is different from the simplex formed by all the vertices of D not belonging to D_P. But if we consider the simplex formed by all the vertices not belonging to D_P, we get a different complementary simplex :

$$S^2_{\neg P} = \langle \langle a, d \rangle, \langle b, d \rangle, \langle c, d \rangle \rangle$$

We say that the two different complementary complexes resp. $S^1_{\neg P}$ and $S^2_{\neg P}$ are the complementary of S_P into S_D at dimension resp. 2 and 1. We can take the complementary of a m-simplex S into a n-simplex S' at any intermediate dimension between the dimension n of S' and the dimension m of S.

3.2.4 Connectors

Conjunction $\wedge$ Let us consider the FOL formula $F(x) = P(x) \wedge Q(x)$, whose validity domain is $D_F = D_P \cap D_Q$. The intersection between two simplexes is given by the *frontier*[2] between them : the shared faces.

Disjunction $\vee$ Let us consider the FOL formula $G(x) = P(x) \vee Q(x)$, whose validity domain is given by $D_G = D_P \cup D_Q$. The union between two simplexes may be the complex composed of the two simplexes or the "big" simplex constructed over the vertices belonging to at least of the simplexes. These two ways of constructing the union are close to the notion of degree of complementarity. They also have different meanings in terms of knowledge representation.

3.3 Simplicial Representation of a TBox

The Kbase modeling an agent is composed of a TBox and a ABox, the terminology being described into the TBox. We try now to give a simplicial interpretation for a finite terminology of a KBase modeling an agent.

3.3.1 The Relations of a TBox

The relations that are used to introduce descriptions in a TBox, can be classified into three families :

1. Type specification : "is-a" . The simplicial representation of this kind of relation is done the same way as for the instantiation relation.

[2] A frontier between two complexes is defined in algebraic topology as the set of all simplexes belonging to both complexes.

2. Concept specification : with the use of attributes. The simplicial representation corresponds to the representation of sets of predicates, and is trivial according to our extension of Q-Analysis given in section 2.1.3.
3. Structure specification : all the "part-of" relations. This kind of relationship plays a crucial role in the representation of the software components that we have to model as agents in InterViews : we study the simplicial representation of these structures in the following section.

3.3.2 The part-of relation

Among the relations that we can represent with the spatial metaphor is the part-of relation. Many different part-of relations have been reported in [1], which classifies them in terms of homeomorphism, functionality and removability of the part.

In any case, all the part-of relations are binary relations and can be represented as simplexes with the Q-Analysis. Note that they correspond to the introduction of special constructors in the DL language (see [1][6]).

3.3.3 Aggregates

In the field of the study of the part-of relations, aggregates should receive special attention [6]. Aggregates result from the explicit naming of a structure build from part-of relations over components. These aggregates present different granularities and may need the introduction of multi-dimensional representations to reflect their structure. The specific structure of an aggregate depends on the relationship between the parts and the whole but also on the relationship between the parts. CAT formalism gives us the opportunity to show these dependencies *all at once*.

We propose to represent the binary relation between the parts P_i of the aggregate and the relations r_i that relates them to the aggregate. If these parts are themselves aggregates, we introduce the relation $p_i \in \{r_i\}$ and consider the links between this part and the elementary parts. By doing this, we can represent composed aggregates (with different granularities). Let us take an example to clarify our proposition, here is an abstract TBox :

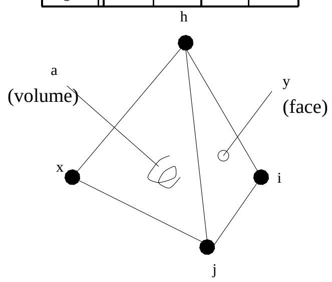
$a := x + y$

$y := h + i + j$

where the $+$ operator symbolizes any of the part-of relations (we call it an abstract part-of relation). To represent this aggregate, we write the incidence

Figure 3: Proposition for a simplicial representation of a composed aggregate (different granularities).

a	h	i	j	x
$+$	1	1	1	1
y	1	1	1	0



matrix of the parts of y and the relations that link them. The aggregate a is itself composed, and represented as a 1-simplex.

Now, the question is how can we represent the compositionality of y on the representation of a ? It cannot be done *at the same level* by simply representing the four vertices h, i, j and x and then the 1-simplex y over it as well as the 4-simplex representing a. Indeed, by doing this, we loose the structural information of y being a sub-aggregate of a.

We propose then to introduce the "relation" y. And in general, each time that a part of a is distinguished and named, we introduce its corresponding "relation", by which all the elementary parts of a that are grouped into a sub-aggregate are linked together. In our example, we will say that "h, i and j are y-linked". This new way of considering a is represented as shown on figure 3.

4 Conclusion

This representation allows us to keep the structural information of the aggregate and also makes easier the direct reference to the named parts. By extension of this method, we could also refer to a sub-aggregate of any aggregate that hasn't been explicitly named. We could imagine that the part-of relations linking the sub-aggregates are not all of the same kind. In this case, the method remains the same, we only have to add labels to the edges of the simplexes (colored simplexes for example, see [9]).

We proposed to split the TBox relations into three families and give a procedure to build a simplicial representation for each family. Thus, we have three complementary and separated simplicial representations of a DL KBase : **Inheritance** ("is-a", relation between concepts of the TBox); **Structure** (relation between concepts and/roles of the TBox);

Note that the correspondence between DL TBoxes and simplicial representations lies on the fact that the denotational semantics is given in set theory, while simplicial complexes are also defined in those terms; the main difference between them is that simplicial complexes only deal with finite sets.

References

[1] Alessandro Artale, Enrico Franconi, Nicola Guarino, and Luca Pazzi. *Data and Knowledge Engineering*, chapter Part-Whole Relation in Object-centered Systems : an Overview. Elsevier, 1996.

[2] R. H. Atkin. *Combinatorial Connectivities in Social Systems*. Verlag, 1977.

[3] Alexander Borgida and Deborah L. McGuiness. Asking queries about frames. In *KR'96*, 1996.

[4] F. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. *Principles of Knowledge Representation*, chapter Reasoning in Description Logics, pages 193–238. Studies in Logic. CSLI Publications, 1996.

[5] Christian Freksa. Spatial and temporal structures in cognitive processes. In C. Freksa, M. Jantzen, and R. Valk, editors, *Foundations of Computer Science*, volume 1337 of *LNCS*. Springer-Verlag, 1997.

[6] Giuseppe De Giacomo and Maurizio Lenzerini. What's in an aggrgate: foundations for description logics with tuples and sets. In *IJCAI95*, pages 801–807. Morgan Kaufmann, 1995.

[7] Kathleen Hornsby and Max J. Egenhofer. Qualitative representation of change. In D.Lukose, H.Delugach, M. Keeler, L. Searle, and J. Sowa, editors, *Spatial Information Theory*, volume 1257 of *LNCS*. Springer-Verlag, 1997.

[8] Bernhard Nebel. *Reasoning and Revision in Hybrid Representation Systems*. Number 422 in LNAI. Springer-Verlag, New York, 2nd edition, 1995.

[9] Erika Valencia and Jean-Louis Giavitto. Modeling diagrammatic reasoning tasks with a topological framework. In *TwD98 Workshop*, dic 1998.