Partial and Informative Common Subsumers in Description Logics

Simona Colucci[1 2] and **Eugenio Di Sciascio**[1] and **Francesco Maria Donini** [3] and **Eufemia Tinelli** [4]

Abstract. Least Common Subsumers in Description Logics have shown their usefulness for discovering commonalities among all concepts of a collection. Several applications are nevertheless focused on searching for properties shared by significant portions of a collection rather than by the collection as a whole. Actually, this is an issue we faced in a real case scenario that provided initial motivation for this study, namely the process of Core Competence extraction in knowledge intensive companies. The paper defines four reasoning services for the identification of meaningful common subsumers describing partial commonalities in a collection. In particular Common Subsumers adding informative content to the Least Common Subsumer are investigated, with reference to different DLs.

1 Introduction

Least Common Subsumers(LCSs) were originally proposed by Cohen, Borgida and Hirsh [5] as novel reasoning service for the Description Logic underlying Classic [4]. By definition, for a collection of concept descriptions, their LCS represents the most specific concept description subsuming all of the elements of the collection. The usefulness of such inference task has been shown in several application classes, varying from learning from examples [6, 7, 10], to similarity-based Information Retrieval [12, 13] and bottom-up construction of knowledge bases [1].

Nevertheless, there are some problems where the computation of LCS does not provide solutions. The LCS in fact intuitively represents properties shared by *all* the elements of a given collection. In several applications, instead, such a sharing is not required to be full: in other words we could be interested in finding a concept description subsuming a portion of the elements in the collection. Different perspectives on the introduced problem may be taken: if the LCS of the collection is the universal concept, we can determine the concept description subsuming a number m of concept descriptions in the collection, where m is the maximum cardinality of subsets of the collection for which a common subsumer non-equivalent to the universal concept exists. We give the name *Best Common Subsumer* to such a concept description, in analogy with LCS. Alternatively, we could be interested in determining a concept description subsuming at least k elements in the collection, where k is a threshold value established a priori on the basis of a decisional process dependent on the application domain. We give such a different concept description the name k*-Common Subsumer* (k-CS). In particular, the search should revert on those k-CSs adding informative content to LCS: we call *Informative* k*-Common Subsumer* (IkCS) a k-CS more specific than the LCS of the collection. We here define the k-*CS*, the *IkCS*, the *BCS* and one more specific service (*Best Informative Common Subsumer*) and give some computation results in different DLs, namely $\mathcal{ALN}$, $\mathcal{EL}$ and $\mathcal{ALE}$.

[1] SisInfLab–Politecnico di Bari, Bari, Italy
[2] D.O.O.M. s.r.l., Matera, Italy
[3] Università della Tuscia, Viterbo, Italy
[4] Università di Bari, Bari, Italy

2 Definitions

The definition of the four novel services relies on Least Common Subsumer definition, which we recall in the following.

Definition 1 (LCS, [7]) *Let $C_1, \ldots, C_n$ be n concepts in a DL $\mathcal{L}$. An LCS of $C_1, \ldots, C_n$, denoted by $LCS(C_1, \ldots, C_n)$, is a concept E in $\mathcal{L}$ such that the following conditions hold:(i) $C_i \sqsubseteq E$ for $i = 1, \ldots, n$; (ii) E is the least $\mathcal{L}$-concept satisfying (i),i.e., , if E' is an $\mathcal{L}$-concept satisfying $C_i \sqsubseteq E'$ for all $i = 1, \ldots, n$, then $E \sqsubseteq E'$.*

We define in the following a new concept, which represents the commonalities of k concepts out of the n in a collection of DL concepts.

Definition 2 (k-CS) *Let $C_1, \ldots, C_n$ be n concepts in a DL $\mathcal{L}$, and let be $k < n$. A* k-Common Subsumer *(k-CS) of $C_1, \ldots, C_n$ is a concept D such that D is an LCS of k concepts among $C_1, \ldots, C_n$.*

Among k-Common Subsumers we distinguish concepts adding informative content to the LCS of the investigated collection.

Definition 3 (IkCS) *Let $C_1, \ldots, C_n$ be n concepts in a DL $\mathcal{L}$, and let $k < n$. An* Informative k-Common Subsumer *(IkCS) of $C_1, \ldots, C_n$ is a k-CS E such that E is strictly subsumed by $LCS(C_1, \ldots, C_n)$.*

Some Informative k-Common Subsumers are peculiar for subsuming the maximum number of concepts in the collection, with such a maximum less than the cardinality n of the collection. We therefore define in what follows:

Definition 4 (BICS) *Let $C_1, \ldots, C_n$ be n concepts in a DL $\mathcal{L}$. A* Best Informative Common Subsumer *(BICS) of $C_1, \ldots, C_n$ is a concept B such that B is an Informative k-CS for $C_1, \ldots, C_n$, and for every $k < j \leq n$ every j-CS is not informative.*

For collections whose LCS is equivalent to the universal concept the following definition makes also sense:

Definition 5 (BCS) *Let $C_1, \ldots, C_n$ be n concepts in a DL $\mathcal{L}$. A* Best Common Subsumer *(BCS) of $C_1, \ldots, C_n$ is a concept S such that S is a k-CS for $C_1, \ldots, C_n$, and for every $k < j \leq n$ every j-CS $\equiv \top$.*

Proposition 1 *If $LCS(C_1, \ldots, C_n) \equiv \top$, every BCS is a BICS.*

Even though the services defined above may appear quite similar to each other at a first sight, it has to be underlined that they deal with different problems:
k-CS: can be computed for every collection of elements and finds least common subsumers of k elements among the n belonging to the collection;
IkCS: describes those k-CSs adding an informative content to the one provided by LCS, *i.e.,* more specific than LCS. Observe that IkCS does not exist when every subset of k concepts has the same LCS as the one of all $C_1, \ldots, C_n$;
BICS: describes IkCSs subsuming h concepts, such that h is the maximum cardinality of subsets of the collection for which an IkCS exists. A BICS does not exist if and only if $C_i \equiv C_j$ for all $i, j = 1, \ldots, n$;
BCS: may be computed only for collections admitting only LCS equivalent to the universal concept; it finds k-CSs such that k is the maximum cardinality of subsets of the collection for which an LCS not equivalent to $\top$ exists.

3 Computation

The complexity of computing the common subsumers defined in Section 2 depends on the specific DL in which the collection is represented. We will therefore separate the results for three different DLs in the following. Nevertheless, some results are common to every DL, like the following theorem, which deals with the cardinality of the set of k-CSs, given a collection of concepts in a DL $\mathcal{L}$.

Theorem 1 *For some sets of n concepts $C_1, \ldots, C_n$ in a DL $\mathcal{L}$, and for some $k < n$, there are exponentially many kCS of $C_1, \ldots, C_n$.*

The following theorem, instead, focuses on the complexity for finding a BCS w.r.t. to the one for computing an LCS.

Theorem 2 *Let m be the sum of the sizes of $C_1, \ldots, C_n$. Then finding a BCS of $C_1, \ldots, C_n$ amounts to the computation of $O(m^2)$ subsumption tests in $\mathcal{L}$, plus the computation of one LCS.*

Both theorems are proved in [8]. Hereafter, regardless of the DL employed for the representation of concepts, we will refer to the solution sets for the introduced reasoning services by the names: B for the set of BCSs, BI for the set of BICSs, I_k for the set of IkCSs, given $k < n$ and L_k for the set of k-CSs, given $k < n$. For a collection of concept descriptions in $\mathcal{ALN}$, an algorithm can be defined computing the solution sets [8]. Complexity results for this algorithm are claimed in the following theorem.

Theorem 3 *Let $C_1, \ldots, C_n$, $\mathcal{T}$ be n concepts and a simple Tbox in $\mathcal{ALN}$, let m be the sum of the sizes of $C_1, \ldots, C_n$, and let $S(s)$ be a monotone function bounding the cost of deciding $C \sqsubseteq_{\mathcal{T}} D$ in $\mathcal{ALN}$, whose argument s is $|C| + |D| + |\mathcal{T}|$. The computation of the solution sets B, BI, L_k, I_k for a collection of concept descriptions in $\mathcal{ALN}$ is then a problem in $O(m^2 + (S(m))^2)$.*

Baader et al. [2] showed that, by taking into account existential restriction, the n-ary LCS operation is exponential, even for the small DL $\mathcal{EL}$, and even shortening possible repetitions by using a TBox [3]. The computation results for the determination of the solution sets of a concept collection in $\mathcal{EL}$ and $\mathcal{ALE}$ are affected by results for LCS:

Theorem 4 *The computation of the solution sets B, BI, L_k, I_k for a collection of concept descriptions in $\mathcal{EL}$ or $\mathcal{ALE}$ may be reduced to the problem of computing the LCS of the subsets of the collection and may then grow exponential in the size of the collection.*

For computing L_k it is sufficient to compute for every subset $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ the concept $LCS(C_{i_1}, \ldots, C_{i_k})$. The same holds for I_k, excluding those $LCS(C_1, \ldots, \mathcal{C}_k)$ which are equivalent to $LCS(C_1, \ldots, \mathcal{C}_n)$. For the computation of the sets B and BI, instead, an algorithm can be defined[8], based on the one proposed by Kusters and Molitor [11] for LCS computation.

4 Conclusions

Motivated by a real-world application need —finding Core Competence in knowledge-intensive companies— we defined and investigated novel reasoning services finding commonalities among portions in a collection of concepts in $\mathcal{ALN}$, $\mathcal{EL}$ and $\mathcal{ALE}$. In all of the three studied languages a computation algorithm has been designed. The computation algorithm for $\mathcal{ALN}$ has been also implemented in the framework of IMPAKT, a novel and optimized knowledge-based system for competences and skills management[9], which will be released late this year by D.O.O.M. s.r.l.

Acknowledgment

We thank Franz Baader for helpful discussions. This work has been supported in part by projects EU-FP6-IST-26896, PE_013 *Innovative models for customer profiling* and PS_092 *DIPIS*.

REFERENCES

[1] F. Baader and R. Küsters. Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions. In *Proc. of KI-98*, volume 1504 of *LNCS*, pages 129–140, Bremen, Germany, 1998. Springer–Verlag.

[2] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumer in description logics with existential restriction. Technical Report LTCS-Report 98-09, RWTH Aachen, 1998.

[3] F. Baader and A.-Y. Turhan. On the problem of computing small representations of least common subsumers. In *Proc. of KI 2002*, volume 2479 of *LNAI*, Aachen, Germany, 2002. Springer-Verlag.

[4] A. Borgida, R.J. Brachman, D. L. McGuinness, and L. Alperin Resnick. CLASSIC: A structural data model for objects. In *Proc. of ACM SIGMOD*, pages 59–67, 1989.

[5] W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In *Proc. of AAAI-92*, pages 754–760. AAAIP, 1992.

[6] W. Cohen and H. Hirsh. The learnability of description logics with equality constraints. *Machine Learning*, 17(2-3):169–199, 1994.

[7] W. Cohen and H. Hirsh. Learning the CLASSIC description logics: Theorethical and experimental results. In *Proc. of KR'94*, pages 121–133, 1994.

[8] S. Colucci, E. Di Sciascio, and F.M. Donini. Partial and informative common subsumers of concept collections in description logics. In *Proc. of DL 2008*, 2008.

[9] S. Colucci, T. Di Noia, E. Di Sciascio, F.M. Donini, and A. Ragone. Semantic-based skill management for automated task assignment and courseware composition. *Journal of Universal Computer Science*, 13(9):1184–1212, 2007.

[10] M. Frazier and L. Pitt. CLASSIC learning. In *Proc. of the 7th Annual ACM Conference on Computational Learning Theory*, pages 23–34, New Brunswick, New Jersey, 1994. ACM Press and Addison Wesley.

[11] R. Küsters and R. Molitor. Structural Subsumption and Least Common Subsumers in a Description Logic with Existential and Number Restrictions. *Studia Logica*, 81:227–259, 2005.

[12] T. Mantay, R. Moller, and A. Kaplunova. Computing probabilistic least common subsumers in description logics. In *KI - Kunstliche Intelligenz*, volume 1701 of *LNCS*, pages 89–100. Springer-Verlag, 1999.

[13] R. Möller, V. Haarslev, and B. Neumann. Semantics-based information retrieval. In *Proc. of IT&KNOWS-98*, Vienna, Budapest, 1998.