Mining $\mathcal{EL}^{\bot}$ Bases with Adaptable Role Depth (Extended Abstract)

Ricardo Guimarães, Ana Ozaki, Cosimo Persia,[1] and Barış Sertkaya[2]

[1] Department of Informatics, University of Bergen, Norway
Ricardo.Guimaraes@uib.no, Ana.Ozaki@uib.no, Cosimo.Persia@uib.no
[2] Frankfurt University of Applied Sciences, Germany
sertkaya@fb2.fra-uas.de

CEUR-WS.org/Vol-2954/abstract-19.pdf

This extended abstract gives an overview of our work [9], published at AAAI 2021. We propose an approach for automatically extracting concept inclusions (CIs) from data. This data can be, for instance a collection of facts from a database or from a knowledge graph. For example, in the DBpedia knowledge graph [14], one can represent the relationship between a city 'a' and the region 'b' it belongs to with the facts $\mathsf{city}(\mathsf{a})$, $\mathsf{region}(\mathsf{b})$, $\mathsf{partof}(\mathsf{a},\mathsf{b})$, and $\mathsf{capital}(\mathsf{b},\mathsf{a})$. One can then mine a CI expressing that a capital is a city that is part of a region.

For mining such CIs, we apply methods from Formal Concept Analysis (FCA) [8]. Several other works in the literature have applied FCA for similar purposes. In [2, 7] the authors use FCA for mining an $\mathcal{EL}^{\bot}_{gfp}$-base of the CIs that hold in an interpretation. A base is a set of CIs that entail every valid implication of the interpretation. Fixpoint semantics of $\mathcal{EL}^{\bot}_{gfp}$ is used here to overcome the difficulty of mining CIs from cyclic relationships in the data, which however comes with practical drawbacks. Based on the approach by Distel et al. [7], Borchmann et al. proposed a more practical method for mining $\mathcal{EL}^{\bot}$-bases with a predefined and fixed role depth for concept expressions [4]. As a consequence, this base is sound and complete only w.r.t. the CIs containing concepts with bounded role depth. For work on applying FCA methods in the DL context see [1, 3, 18–20]. For more work on mining CIs using FCA see [5, 6, 12, 13, 15], for work on learning DL ontologies see [11, 17] and for a survey comparing different approaches for learning DL ontologies see [16].

Our work [9] brings the best of the approaches by Distel et al. [7] and Borchmann et al. [4] together: we *directly* compute a finite $\mathcal{EL}_{\bot}$-base (without a detour over $\mathcal{EL}^{\bot}_{gfp}$) that captures the *whole* language (not only until a certain role depth). In particular, we present a new approach that computes an adaptable role depth based only on the objects that are considered during the computation of CIs. To define an $\mathcal{EL}^{\bot}$ base, we use the notion of a model-based most specific concept (MMSC) up to a certain role depth [4, 7]. The MMSC plays a key rôle in the computation of a base from a given finite interpretation.

Definition 1. *An $\mathcal{EL}^{\bot}$ concept expression C is a model-based most specific concept of $X \subseteq \Delta^{\mathcal{I}}$ with role depth $d \geq 0$ iff (1) $X \subseteq C^{\mathcal{I}}$, (2) C has role depth at*

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

most d, and (3) for all $\mathcal{EL}^\bot$ concept expressions D with role depth at most d, if $X \subseteq D^\mathcal{I}$ then $\emptyset \models C \sqsubseteq D$.

A straightforward (and inefficient) way of computing $\mathsf{mmsc}\left(\{X\}, \mathcal{I}, d\right)$, for a fixed d, would be conjoining every $\mathcal{EL}^\bot$ concept expression C (over $\mathsf{N_C} \cup \mathsf{N_R}$) such that $X \subseteq C^\mathcal{I}$ and the depth of C is bounded by d. A more elegant method for computing MMSCs is based on the product of description graphs and unravelling cyclic concept expressions up to a sufficient role depth. The interesting challenge is how to identify the smallest d that satisfies the property: if $x \in \mathsf{mmsc}\left(X, \mathcal{I}, d\right)^\mathcal{I}$, then $x \in \mathsf{mmsc}\left(X, \mathcal{I}, k\right)^\mathcal{I}$ for every $k > d$.

We have developed a method for computing MMSCs with a role depth that is suitable for building an $\mathcal{EL}^\bot$ base of the given interpretation [9]. This method is based on the notion of *maximum vertices from* (MVF), which measures the maximum number of distinct vertices that a walk with a fixed starting point can visit in the graph. There we prove that in a product of description graphs, even for the vertices that are parts of cycles there is a certain depth of unravellings, which we call a fixpoint, that is guaranteed to be an upper bound. We use this result for defining a finite set of concept expressions $M_\mathcal{I}$ for building a base of the CIs valid in $\mathcal{I}$, which is a definition adapted from the work by Distel et al. [7].

Definition 2. *Let $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ be a finite interpretation. The set $M_\mathcal{I}$ is the union of $\{\bot\} \cup \mathsf{N_C}$ and*

$$\{\exists r.\mathsf{mmsc}\left(X, \mathcal{I}\right) \mid r \in \mathsf{N_R} \text{ and } X \subseteq \Delta^\mathcal{I}, X \neq \emptyset\}$$

We also define $\Lambda_\mathcal{I} = \{\bigsqcap U \mid U \subseteq M_\mathcal{I}\}$.

Building the base mostly relies on the fact that, given a finite interpretation $\mathcal{I}$, for any $\mathcal{EL}^\bot$ concept expression C, there is a concept expression $D \in \Lambda_\mathcal{I}$ such that $C^\mathcal{I} = D^\mathcal{I}$.

Theorem 3. *Let $\mathcal{I}$ be a finite interpretation and let $\Lambda_\mathcal{I}$ be as above. Then,*

$$\begin{aligned}\mathcal{B}(\mathcal{I}) = &\{C \equiv \mathsf{mmsc}\left(C^\mathcal{I}, \mathcal{I}\right) \mid C \in \Lambda_\mathcal{I}\} \cup \\ &\{C \sqsubseteq D \mid C, D \in \Lambda_\mathcal{I} \text{ and } \mathcal{I} \models C \sqsubseteq D\}\end{aligned}$$

is a finite $\mathcal{EL}^\bot$ base for $\mathcal{I}$.

The base computed this way is complete for all CIs of any depth holding in $\mathcal{I}$. Unfortunately, it is not guaranteed to have minimal cardinality. In plain FCA and in the works [4, 7], the computed bases fulfill this minimality condition. However, in our setting which does not make use of the fixpoint semantics, the base needs to contain CIs of the form given in the second line of the definition. This makes the base loose the minimality condition. For full proofs of the results see the long version [10].

As future work, we are going to work on constructing a base that satisfies the minimality condition. Additionally, we are going to investigate the problem of mining CIs in the presence of noise in the dataset by making use of the support and confidence measures from association rule mining.

Bibliography

[1] Baader, F.: Computing a minimal representation of the subsumption lattice of all conjunctions of concepts defined in a terminology. In: Proc. Intl. KRUSE Symposium. pp. 168–178 (1995)

[2] Baader, F., Distel, F.: A finite basis for the set of $\mathcal{EL}$-implications holding in a finite model. In: Medina, R., Obiedkov, S. (eds.) ICFCA 2008. pp. 46–61. Springer-Verlag (2008)

[3] Baader, F., Ganter, B., Sertkaya, B., Sattler, U.: Completing description logic knowledge bases using formal concept analysis. In: Veloso, M.M. (ed.) IJCAI. pp. 230–235. AAAI Press (2007)

[4] Borchmann, D., Distel, F., Kriegel, F.: Axiomatisation of general concept inclusions from finite interpretations. Journal of Applied Non-Classical Logics **26**(1), 1–46 (jan 2016)

[5] Borchmann, D.: Learning Terminological Knowledge with High Confidence from Erroneous Data. Ph.D. thesis, Dresden University of Technology (2014)

[6] Borchmann, D., Distel, F.: Mining of el-gcis. In: Spiliopoulou, M., Wang, H., Cook, D.J., Pei, J., Wang, W., Zaïane, O.R., Wu, X. (eds.) Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on. pp. 1083–1090. IEEE Computer Society (2011)

[7] Distel, F.: Learning description logic knowledge bases from data using methods from formal concept analysis. Ph.D. thesis, Dresden University of Technology (2011)

[8] Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg (1999)

[9] Guimarães, R., Ozaki, A., Persia, C., Sertkaya, B.: Mining EL bases with adaptable role depth. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021. pp. 6367–6374. AAAI Press (2021)

[10] Guimarães, R., Ozaki, A., Persia, C., Sertkaya, B.: Mining $\mathcal{EL}^{\perp}$ bases with adaptable role depth. Tech. rep. (2020), https://arxiv.org/abs/2102.10689

[11] Konev, B., Lutz, C., Ozaki, A., Wolter, F.: Exact learning of lightweight description logic ontologies. J. Mach. Learn. Res. **18**, 201:1–201:63 (2017)

[12] Kriegel, F.: Constructing and Extending Description Logic Ontologies using Methods of Formal Concept Analysis. Ph.D. thesis, Technische Universität Dresden, Dresden, Germany (2019)

[13] Kriegel, F.: Learning Description Logic Axioms from Discrete Probability Distributions over Description Graphs. In: Calimeri, F., Leone, N., Manna, M. (eds.) JELIA 2019. pp. 399–417. Springer (2019)

[14] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: Dbpedia - A large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web **6**(2), 167–195 (2015)

[15] Monnin, P., Lezoche, M., Napoli, A., Coulet, A.: Using formal concept analysis for checking the structure of an ontology in LOD: the example of dbpedia. In: ISMIS. pp. 674–683. Springer (2017)
[16] Ozaki, A.: Learning description logic ontologies: Five approaches. where do they stand? KI - Künstliche Intelligenz (04 2020)
[17] Ozaki, A.: On the complexity of learning description logic ontologies. In: Manna, M., Pieris, A. (eds.) Reasoning Web. Declarative Artificial Intelligence - 16th International Summer School 2020, Oslo, Norway, June 24-26, 2020, Tutorial Lectures. Lecture Notes in Computer Science, vol. 12258, pp. 36–52. Springer (2020)
[18] Rudolph, S.: Exploring relational structures via $\mathcal{FLE}$. In: Wolff, K.E., Pfeiffer, H.D., Delugach, H.S. (eds.) ICCS. pp. 196–212. Springer-Verlag (2004)
[19] Rudolph, S.: Relational exploration: Combining Description Logics and Formal Concept Analysis for knowledge specification. Ph.D. thesis, Fakultät Mathematik und Naturwissenschaften, TU Dresden, Germany (2006)
[20] Sertkaya, B.: A survey on how description logic ontologies benefit from formal concept analysis. In: Kryszkiewicz, M., Obiedkov, S. (eds.) CLA. CEUR Workshop Proceedings, vol. 672, pp. 2–21 (2010)