A Family History Knowledge Base in OWL 2

Robert Stevens, Nicolas Matentzoglu, Uli Sattler, and Margaret Stevens

School of Computer Science, University of Manchester, Manchester (United Kingdom)
{robert.stevens|Ulrike.sattler}@manchester.ac.uk

Abstract. This paper presents a challenging Family History Knowledge Base (FHKB) authored in OWL 2 DL. Originally, the FHKB was designed to act as a tool for education, especially about OWL 2's features and the use of automated reasoners. As a result, the FHKB has been constructed to maximise use of inference. For individuals representing people, only genealogical assertions on parentage and sparse assertions of siblinghood are given explicitly. All other genealogical inferences are driven by a rich property hierarchy, property characteristics and subproperty chains. A rich collection of entailments are generated, but reasoners struggle to handle a version with all of Robert's known relatives.

1 Introduction

The Family History Knowledge Base (FHKB) is a challenging OWL 2 DL ontology for genealogy. Genealogy affords many opportunities for logical inference, and, as it is also an accessible domain, it affords a good case study for teaching OWL. Genealogy is an attractive example simply because it involves many complex, well understood relationships between individuals. More importantly, most of these relationships can be inferred from few asserted facts. For example, parent-child relationships between individuals together with their sex are enough to infer all common kin relationships. In addition, the genealogy example is accessible to all users for the simple reason that everyone has a family history (although for some it might be unknown).

To represent kinship relations in OWL 2, we need to decide which properties to use as basic, explicit ones for assertions in our ontology. Alternatively to the above mentioned possibility, we could use isSonOf, isDaughterOf as base relation and then infer all other family relations as well as gender from that. For example, sharing one or both parents implies a half sibling or sibling relationship. Limitations in the expressive power of OWL 2 means that we have difficulties to determine sibling relationships in this way.

The FHKB touches many of the expressive features of OWL 2; it seeks to minimise the amount of explicit assertions necessary, and to maximise the use of inference; it uses many of the expressive features of OWL 2. It also offers a range of modelling options that can be explored *vis a vis* their effects on querying and performance of automated reasoners. The full `Stevens` FHKB contains around 450 family members. However, the FHKB as it currently stands already presents a challenge to many of the current OWL 2 DL reasoners.

2 Materials and Methods

We model the typical family relationships as OWL object properties; the objects being Persons. We are interested in being able to infer as many of these family relationships as possible and, as all family relationships are based on parentage, we use the properties to help us infer these other relationships. We do this by exploiting OWL property characteristics. For example, a person's parents are also that person's ancestors, and by making this 'ancestor' property transitive, we can identify parents of parents (and so on) as ancestors of an individual.

Only ten 'core' classes are used in the FHKB TBox:

1. A class DomainEntity, with three disjoint subclasses Person, Sex, and Partnership;
2. Male and Female, disjoint subclasses of Sex. Sex is covered by Male and Female. We also use a functional object property hasSex to relate a Person with his or her Sex.
3. Man that is defined as a Person that hasSex some Male and
4. Woman that is defined as Person that hasSex some Female;
5. Partnership and a subclass Marriage;

Woman and Man are used as domain and range constraints in the property hierarchy that drives the inferences made about individuals. This property hierarchy is shown in Figure 1. This core FHKB has also been extended to cover occupations of people, their eye colours and their places of birth. The FHKB has many other possible extensions.

We aimed at class-level axioms that allowed us to be as 'sparse' as possible with respect to individual assertions, to make our approach as maintainable and re-usable as possible and to reduce the risk of adding erroneous assertions. We have decided to restrict the explicit required information about an individual to property assertions related to their parentage and siblings, using properties isMotherOf, isFatherOf, isBrotherOf, isSisterOf, isDaughterOf, isSonOf. For each individual, we add as much information as necessary to place them in the family tree and infer their sex which, for women, is often a single isDaughterOf or isSisterOf assertion. Of course, there are cases where, in addition, we need to also assert isMotherOf relations. A similar approach is followed for men.

Marriages or partnerships are relations in which 'partners' participate. In our ontology, we use the restriction hasPartner min 2 Person on the Partnership class to be permissive on the number and sex of participants in a partnership. Subproperties of hasPartner of hasMalePartner and hasFemalePartner allow specifics of a particular partnership to be described.

We add some axioms about the Person class, where hasMother and hasFather are the inverses of isMotherOf and isFatherOf, and subproperties of hasParent:

1. That each and every person hasMother some Woman;
2. that each and every person hasFather some Man;
3. that each and every person hasParent max 2 Person.

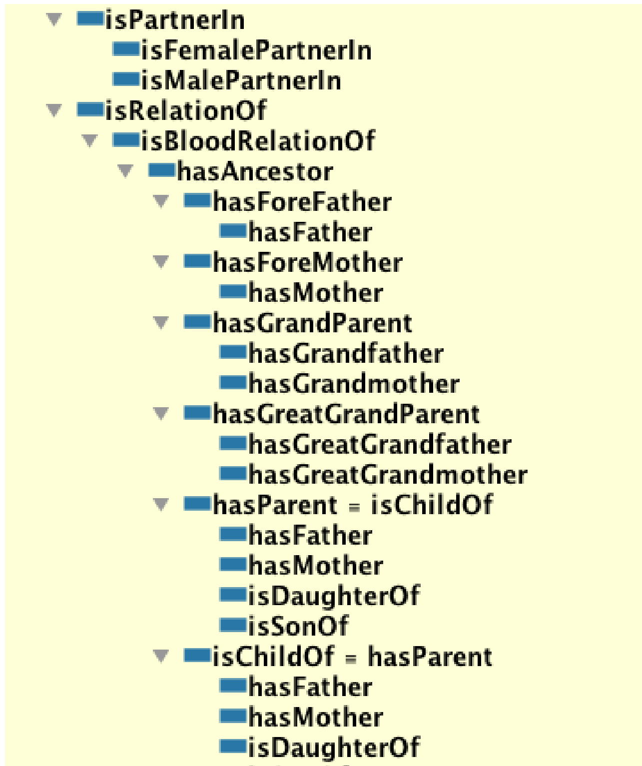

Fig. 1. The property hierarchy from the FHKB.

The last restriction in this list means that there are only two ways of having a parent in the FHKB; this, for instance, means that two siblings's sets of grandparents will be inferred to be identical.

Having asserted parentage and some sibling relationships, subproperty chains were also used to infer other relationships, such as aunts, uncles, grandparents, cousins, spouses, blood relations, etc., none of which are transitive, some of which are symmetric. Moreover, the FHKB contains numerous equivalent class axioms (Figure 2), both to act as queries and tests; most of these are of the form 'Xs are those persons who hold an isXof relationship'. Finally, we use nominals ("value" restrictions) to define the class of Robert's cousins and similar classes.

Of course, this is not the only way of modelling family relations in OWL 2, but it clearly is one that fits our understanding of these relations, and requires only rather few assertions—which makes it less likely to involve conflicting assertions.

3 Results

The full FHKB is in the DL $\mathcal{SROIQ}(\mathcal{D})$ and involves

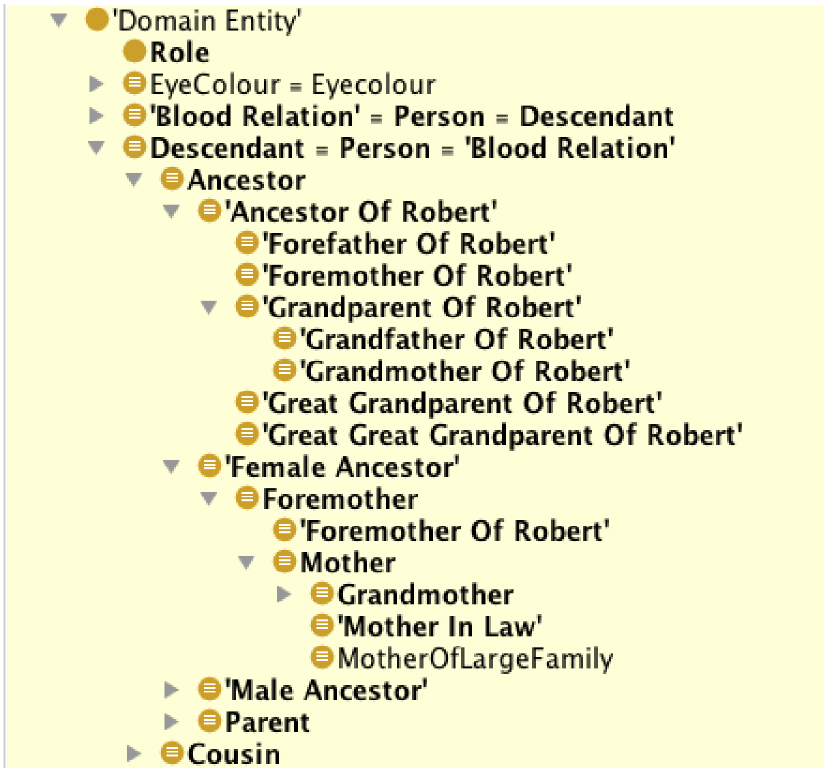

Fig. 2. The Tbox for the FHKB.

- 3,732 logical axioms (and assertions), out of which 514 are tautological (for example, SubClassOf Thing, or instance of Thing);
- a total of 425 TBox axioms with
  - 58 subclass restrictions and 77 equivalent class axioms, involving qualified cardinality restrictions and nominals,
  - 3 disjoint class axioms,
  - 56 domain restrictions and 55 range restrictions,
  - 51 object property inclusions,
  - 28 object property chain inclusions.
  - 33 object property inverse axioms,
  - 5 object property functionality axioms,
  - 7 object property transitivity axioms, and
  - 11 object property symmetry axioms.
- a total of 526 individuals (including 437 instances of Person and 71 instances of Marriage),
- 134 class names,
- 13 data properties and 87 object properties.

So, the TBox itself constitutes only around 11% of the ontology and involves many object property axioms, a rather large number of subproperty chains, nominals, and a few datatypes. The TBox is cyclic and has no explicit GCIs, but does have 2 implicit GCIs.

All reasoning experiments with the FHKB were conducted on a Mac mini with 2.7 GHz Intel Core i7, 16 GB 1333 DDR3 RAM, AMD Radeon HD 6630M 256 MB, Mac OSX Lion 10.7.5. Every reasoning process was executed in a separate isolated virtual machine (Java 7, -Xms2G, -Xmx12G) with a six hour timeout.

The experiment results and the versions of the FHKB used are available at http://owl.cs.manchester.ac.uk/?attachment_id=1173.

The FHKB poses a significant challenge for DL reasoners. Even a single consistency test can cause the reasoner to run out of memory or hit a timeout. We initially evaluated the performance of four reasoners (HermiT 1.3.8 [1], Pellet 2.3.1 [2], JFact 1.2.1[1] and FaCT++ 1.6.3 [4]) on the FHKB, but neither HermiT (due to out of memory exceptions) nor JFact (due to a wrong reasoning result (inconsistency)) managed to successfully deal with it. Our main inclusion criterion for the reasoner was the implementation of the OWL API reasoner interface, but we also made cross-checks with Konclude 0.5.0 [3], which was able to determine consistency efficiently, but failed to produce a classification. With the remaining two reasoners, Pellet and FaCT++, we tried to obtain consistency and classification times. Pellet managed to determine consistency in more than 4 minutes (262 seconds), but generally failed when trying to classify with a timeout. FaCT++ 1.6.3 needed around 2.5 seconds to determine consistency and 47.5 seconds to produce a classification. We were not yet able to generate the full set of ABox entailments, i.e., determine, for each individual, the class names they are entailed to be an instance of. We performed experiments with three versions of the FHKB: the recent one, for which we presented the results, a slightly reduced version that omits some constraints on the TBox (for example a max 2 cardinality restriction on the Person class) and an older version of the FHKB. FaCT++ classifies the slightly reduced version in around 35 seconds and the old version in around 30.5 seconds.

In the FHKB, 1 116 entailments are determined to hold between the individual `robert` and other named individuals, and many new type axioms are also inferred on `robert` from the FHKB's TBox. Most entailments fit well with our knowledge about the `Stevens` family, yet some unexpected relationships were found, mainly due to the way we chose to model isSiblingOf: recall that isSiblingOf being transitive and symmetric means that it is also linking a sibling to itself, and could therefore be called isSiblingOfOrSelf. As a consequence, for example, `robert` is entailed to have two brothers and eleven first cousins—rather than one brother and nine cousins—because he is a brother of himself and all his siblings are also his cousins. There are numerous similar unintended entailments in other degrees and removes of cousin. Again, this is due to our choice of making siblingOf transitive and reflexive, despite the fact that we can not prevent the entailment of reflexive relations. (Other modelling options have similar

[1] https://github.com/owlcs/jfact

problems.) We have accepted this as a compromise because it saved us explicitly adding all sibling relationships.

4 Discussion

The `Stevens` FHKB makes use of a wide range of constructs available in OWL 2 and is designed to lead to a large number of interesting entailments. Admittedly, family relationships could be said to be slightly at odds with the "spirit" of the expressive power of the DLs underlying OWL 2 since they can all be, more or less, expressed in Horn rules. And of course, we have struggled with our modelling, mainly when modelling the sibling relationship: firstly, since we cannot state that isSiblingOf is 'sort of' transitive yet irreflexive, we have to either live with persons being their own siblings (and therefore their siblings being their cousins), or to state all sibling relationships explicitly. We have chosen the former approach, and thus 'too many' inferences are made. Secondly, we were unable to distinguish full- and half-relationships. Only with two parents in common between two male individuals, a full brother relationship should be inferred—if only one parent is in common between two male individuals then half-brotherhood should be inferred. Hence we have chosen to "sparsely seed" sibling relationships rather than inferring them from shared parents.

As we have seen, the FHKB presents challenging problems for OWL 2 reasoners. It uses a rich property hierarchy to infer relationships between individuals. It also deliberately uses challenging expressions such as nominals and qualified cardinality restrictions in its TBox. The FHKB is deliberately set up to maximise inference and this was for educational and exploratory purposes: we have used it in various tutorials of OWL 2. As well as FHKB's educational opportunities, we therefore also offer the FHKB as a challenge for reasoners: by varying the number of individuals involved and further expanding its scope (e.g., geography and occupations), we can potentially turn the FHKB into a suite of benchmark ontologies for OWL 2 reasoners.

Acknowledgements: RS was funded by EPSRC grant EP/ED21352/1. NM was funded by an EPSRC CDT studentship. We would like to thank Simon Jupp, Matthew Horridge, Alan Rector, and Bijan Parsia for their help and advice.

References

1. B. Motik, R. Shearer, and I. Horrocks. Hypertableau Reasoning for Description Logics. *Journal of Artificial Intelligence Research*, 36:165–228, 2009.
2. E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2), 2007.
3. A. Steigmiller, T. Liebig, and B. Glimm. Konclude: System description. *Journal of Web Semantics (JWS)*, 2014.
4. D. Tsarkov and I. Horrocks. FaCT++ description logic reasoner: System description. In *Proc. of the Int. Joint Conf. on Automated Reasoning (IJCAR 2006)*, volume 4130 of *Lecture Notes in Artificial Intelligence*, pages 292–297. Springer, 2006.