A Family History Knowledge Base Using OWL 2

Robert Stevens and Margaret Stevens

School of Computer Science, University of Manchester, UK
`robert.stevens@manchester.ac.uk`

1 Introduction

We present a case study in the use of OWL 2 in the form of a family history knowledge base (FHKB). Genealogy affords many opportunities for logical inference and thus it has been used many times as an example system and not only for description logics. OWL 2 offers greater opportunities for inference than OWL-DL and we test those on a knowledge base of the Stevens family history.

Genealogy is an attractive example simply because it involves many complex, deterministic relationships between individuals. What is more, most of these relationships can be inferred from very few asserted facts. Ideally, motherhood, fatherhood and gender should be enough from which 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).

The particular feature of OWL 2 that is attractive in this example is that of sub-property chains. For example, that 'My parent's brothers are my uncles' fits naturally in to this paradigm. The property `hasParent` chained with the property `hasBrother` implies the property `hasUncle`; so any individual with a `hasParent` relationship to an individual with a `hasBrother` relationship to a third individual also holds the implied `hasUncle` with that third individual.

Using a representation such as OWL 2, it should be possible to infer the majority of desirable family relationships, including the sex of individuals, from assertion of parentage (`fatherOf`, `motherOf`, `sonOf`, `daughterOf`, ...) and siblings.

In addition, it would be useful to be able to have a sparse assertion of sibling relationships and infer brothers and sisters. An example of sparse assertion is that if David is the brother of John and David is the brother of Peter and we know that both John and Peter are men, then John must be the brother of Peter and *vice versa*. Similarly, that individuals are people of a particular sex should be inferred from the fact they are, for example, a mother, father, son or daughter. Maximising inference from asserted facts in this way reduces both data gathering effort and scope for error. One asserted fact may cause many incorrect inferences, but those many errors are fixed by one change rather than many.

The aim of the Stevens FHKB was to see how complete a set of inferences could be made and how correct such inferences would be using OWL 2.

2 Materials and Methods

Only five primitive classes were described in the FHKB TBox:

1. `Person`, a subclass of `DomainEntity`; `Sex`, a subclass of `DomainEntity`; `Person` and `Sex` are disjoint;
2. `Male` and `Female`, disjoint subclasses of `Sex`.

All other classes are fully defined and mainly used for making queries against the FHKB. Two important defined classes are:

1. `Man` that is defined as a `Person that hasSex SOME Male`;
2. `woman` that is defined as `Person that hasSex Female`;
3. `hasSex` is functional.

`Man` and `Woman` are used as domain and range constraints in the property hierarchy that drives the inference made about individuals. Only assertions on parentage and siblings were made; all other kin relationships, including the type of individual persons, was to be driven through inference.

1. The root of the property hierarchy is `hasrelation`. It is both transitive and symmetric: All my relatives are also my relatives and in turn have myself as a relative etc.
2. `parentOf` subsumes `fatherOf`, together with `motherOf`, with the inverses treated equivalently. Domain and range constraints of `Man` and `Woman` allow the sex of individuals to be inferred from assertions of parentage. Declaring `motherOf` and `fatherOf` to be functional constrains a person to have only one mother and only one father.
3. `ancestorOf` subsumes `parentOf`, with the inverses treated similarly. Where the parentage properties are intransitive, `ancestorOf` is transitive. As `hparentOf` implies `ancestorOf`, asserting parentage will also allow all ancestors and descendants to be gathered.
4. `relationof`, `ancestorOf` and `parentOf` all have domain and range of `Person`. 'Sexed' properties all use domain and range constraints of either `Man` or `Woman` in the expected manner. The intention is that domain and range should be used to infer that an individual is at least a `Person` and then either a `Man` or `woman`.
5. Similarly, `brotherOf`, `sisterOf` and their inverses (with appropriate domain and range constraints) are sub-properties of `siblingOf`. The latter is both transitive and symmetric. Where `brotherOf` cannot be transitive (I'm the brother of my sister, but her sister is not my brother), any sibling is my sibling. Ideally, making these sibling properties irreflexive would prevent my being my own brother.
6. Having asserted parentage and some sibling relationships, sub-property chains were also used to infer many relationships:
   (a) `uncleOf` is given by the sub-property chain `brotherOf o parentOf`. Similarly for `auntOf`. This only gives blood relationships; aunts and uncles by marriage are not encompassed in this pattern though the extension is trivial.

(b) Great aunts and uncles can be inferred with sub-property chains such as `hasParent o hasUncle` as appropriate.
(c) `firstCousinOf` is made with a sub-property chain of `hasParent o siblingOf o parentOf`. That is, the children of my parent's siblings are my first cousins. `firstCousinOf` has the characteristic of being symmetric—I am the first cousin of my first cousins.
(d) Further degrees of cousins (second and third cousins etc., can be built in a similar manner. First cousins share a grandparent, but not parents; second cousins share a great-grandparent, but not grandparents; third cousins share great-great-grandparents, but not great-grandparents. So, for my second cousins, I want my grandparents sibling's grandchildren. Other degrees are done with the same kind of pattern.
(e) 'Removes' of cousins refer to generational gaps; so my first cousins children are once removed (my first cousins being the same generation as myself). So, a sub-property chain of `firstCousinOf o parentOf` will imply `firstCousinOnceRemovedOf`. This will make `robert_david_bright_1965` a first cousin once removed of my father's first cousins (and *vice versa*); so the relationships, like all cousin relationships, are symmetric.
(f) `hasGrandfather` is a sub-property chain of `hasParent o hasFather`—my parent's fathers are my grandfathers. `hasGreatGrandfather` can be made with `hasParent o hasGrandfather` and so on. Naturally, the same goes for grandmothers. Having these as sub-properties of `hasGrandParent` allows sex-neutral grand- and great- relationships to be found. All these family relationships are ultimately sub-properties of `hasRelation`.

Four hundred and six members of the extended Stevens family were entered as individuals into the FHKB. The earliest family member has a birth year of 1726[1].

3 Discussion

An FHKB is or should be a compelling example for OWL, especially with the extensions available in OWL 2. The Stevens FHKB demonstrates that much can be done with automated inference. There remain, however, several points that stop an FHKB being the compelling example that it should:

1. The inability to use, as required, the irreflexive property characteristic when the property is used in sub-property chains means, that too many inferences are made. For example, all the cousin relationships need `siblingOf` to be irreflexive. Without this property characteristic my father becomes his own

[1] A de-identified version of the Stevens FHKB is available at `http://owl.cs.manchester.ac.uk/repository/download?ontology=http://www.co-ode.org/roberts/family-tree.owl&version=0&format=RDF/XML`

sibling and thus my brother and myself become our own first cousins. consequently, at present, many entailments in the Stevens FHKB are incorrect.

2. The inability to state, for example, that if `John` is inferred to be a `siblingOf` `David` and that `John` is a `Man` is known, then it should imply that a `brotherOf` relationship holds between `John` and `David`. David is asserted to be the brother of John. the gender of John is known as his parentage of his children is asserted. That John `hasBrother` David can be inferred (from the inverse of the asserted `brotherOf` relationship from David), but at present not that John is `brotherOf` David; thus John is not inferred to be my uncle. For this inference we need that John `brotherOf` David to be inferred from the facts currently known from assertion and inference.

   There is thought to be a "work around' for this problem. it involves asserting the type `isMan some Self` using an object property `isMan` on male individuals in the FHKB. This acts as a 'flag' on the individual signalling its maleness. Then a sub-property chain such as `isMan o siblingOf` implies `brotherOf` could be used. The `isMan` is asserted; the `siblingOf` inferred and thus `brotherOf` can be inferred. This does work, but would mean that the `siblingOf` property hierarchy could not be used without pushing the FHKB into OWL 2 full[2]. without this property hierarchy, other inferences would be lost. It also impinges on the requirement for sparse assertion. That is, gender is, in effect, asserted.
3. The inability to infer full- or half-relationships. With two parents in common between two male individuals, then a full brother relationship can be inferred. If, however, only one parent is in common between two male individuals then half-brotherhood should be inferred.

Rules could be used to fix the problems encountered in the use of OWL 2. In fact, sex, parentage and rules are sufficient to find all the kin relationships needed. Defined classes can also find the correct set of, for instance first cousins, but the task set here was to use only OWL 2. The FHKB will be further extended with information about marriages. Making inferences of in-law' relationships would allow other relationships by marriage to be inferred. Marriage, re-marriage, and all that entails such as step-parenthood would also be useful. Information exists, for example, on eye-colour of Stevens ancestors back into the mid-1800s. Working out probability of carrying genes for brown and blue eye colour should be possible. Naturally, other information such as roles played by individuals can also be added to the FHKB that would allow a wider range of queries to be asked. Moving outside OWL 2, the addition of rules will allow further inferences to be made. Development of the Stevens FHKB offers an OWL 2 version of a traditional example in the field.

Acknowledgements: I would like to thank Matthew Horridge, Simon Jupp, Alan Rector, Uli Sattler and Bijan Parsia for help and advice during this exercise.

[2] See, for example, `http://lists.w3.org/Archives/Public/public-owl-dev/2007OctDec/0004.html` and `http://www.w3.org/2007/OWL/tracker/issues/22?changelog`