Inference Rules

Authors Luther A. Tychonievich

Created 2013-03-15

Title Inference Rules

Version 1.0

Type Technical Solution

Language English

Description Inference rules used in indirect evidence should be codified within the data model.

Keywords inference, indirect evidence, evidence, rule, extension

Inference Rules

Luther A. Tychonievich

15 March 2013

Abstract: *Indirect evidence is an important element in family history research. Many types of indirect evidence can be modelled as inference rules applied to elements of a conclusion to create additional information. Those inference rules should be supported directly in the data model.*

1 Motivation

The proper use of indirect evidence is often cited as an important element of mature genealogical research [1, 2, 3, 4, 5]. Significant portions of indirect evidence can be characterized as an inference rule and a set of supporting beliefs combined to create new information, either positive or negative.

A few examples of common inference rules:

- Age and birth date may be derived from one another via the inference rule, "a person A years old at date D_a is related to the date D_b of their birth via the relationship $D_b + A$ years $\leq D_a < D_b + (A + 1)$ years."
- Depending on the era and location, there is usually some rule providing approximate bounds on the birth date of a mother given the birth dates of her children.
- Negative evidence is a combination of the belief "I didn't find record X" and three individual rules:
  1. "a record looked for and not found doesn't exist where it was sought,"
  2. "a record not existing now was never created," and
  3. "a record of type X is only not created if Y or Z is true,"

  which combine to derive "either Y or Z is true."

Inference rules are also useful in defining other aspects of research. Conflicting evidence might be weighed based on the rule "primary sources are more reliable than secondary sources." A sources might be believed to refer to an existing individual

based on a rule about the number of assertions the source makes that agree with existing information. And so on.

In short, inference rules are widely used, if not widely identified, in genealogical research.

Most inference rules are generalisable. None of the examples I gave above are restricted to one particular person or researcher. Inferences are based on strong trends or patterns in historical facts and are thus, almost by definition, general.

2 A Data Model for Inference Rules

An inference rule may be modelled as a **selector** and a disjunction of **consequents**.

A selector defines the antecedents of the rule: a general template that matches any conclusion that contains the information needed for the rule to apply. These might look something like regular expressions [6], but defined over genealogical conclusions instead of XML or strings. More of my thoughts on selectors can be found in my FHISO requirement document, "Selector Language for Family History Conclusions."

A consequent is of the same form as the data model's normal data except that some parts of it are to be filled in by references to elements matched by the selector when the rule is applied. This derived conclusion might add to, edit, or replace the elements matched by the selector. A comparison might be made to a regular expression replacement string.

In general, a rule might result in a disjunction of consequents, such as "she either died, remarried, or moved out of the country between 1900 and 1910." If such either-or conclusions are supported in the data model they can be handled directly. If not, rules might only be applicable in sets large enough to reduce the conclusion to a single consequent.

The data resulting from the application of an inference rule can be modelled like any other data and cite a "rule application" instead of a source. A rule application is just a data element that references the rule applied and how the selector was matched to the data that existed at the time of the application. Thus, rules can be inserted into any data model that supports some kind of version history and internal citation.

3 Use Cases for Rules in Data

Including rules in data provides for a wide variety of use cases. A few examples are included below.

1. The computer can recognize when changing conclusions invalidate the antecedent of an inference.

2. The "why" of research decisions can be shared in a normalized form.

3. Many existing human-readable comments can be replaced by rules, reducing the translation burden in multilingual collaboration.

4. How-to guides can be made simpler as the rules of a region or era can be encoded as rules and imported directly in the data.

5. Statistical algorithms can verify (or discover, or refute) rules based on databases of records and conclusions.

6. The computer can automatically suggest inference rules that might apply or raise flags for applications that do not appear obviously sound.

7. Researchers can flag particular inferences and see which parts of their conclusion depend on it.

The list could easily go on. Rules bring more of the research process into the realm of computer analysis and assistance.

4 Rules as an Extension

Rules depend only on the data model used for consequents, the query language used in the selectors, and the existence of data history. This should make them simple to add as an extension in many data models.

Since existing tools do not, to my knowledge, use rules they should probably not be *required* by a data standard that wishes for any form of backward compatibility. Thus, they likely ought to be an extension rather than an integral part of a standardised data model.

References

[1] Ancestry Insider, *Genealogical Maturity Model Proposal*. November 2009. `http://ancestryinsider.blogspot.com/2009/11/genealogical-maturity-model.html`

[2] Board for Certification of Genealogists, *BCG Genealogical Standards Manual*. Ancestry Publishing. 2000.

[3] FamilySearch Community, *Genealogical Maturity*. The FamilySearch Research Wiki. `https://familysearch.org/learn/wiki/en/Genealogical_Maturity`

[4] Mills, Elizabeth Shown, *Evidence Explained*. Genealogical Publishing Company, Inc. 2007.

[5] Mills, Elizabeth Shown, *Working With Historical Evidence: Genealogical Principals and Standards*. National Genealogical Society Quarterly 87:3. September 1999.

[6] Perl 5 Porters. *Perlre: Perl regular expressions*. Official documentation for the Perl programming language. `http://perldoc.perl.org/perlre.html`