Probabilistic Linkage of Vital Event Records in Scotland using Familial Groups

Özgür Akgün[1], Tom Dalton[1], Alan Dearle[1], Eilidh Garrett[2], Graham Kirby[1]

[1] School of Computer Science, University of St Andrews, UK
{ozgur.akgun, tsd4, alan.dearle, graham.kirby}@st-andrews.ac.uk
[2] University of Essex, UK

As part of the Digitising Scotland project we are attempting to create family trees for the whole Scottish population from 1855. The input to this process are digitised birth, death and marriage vital event records representing approximately eighteen million individuals. The output will be a graph representing relationships between the individuals along with provenance and probability information about the links. This will permit traditional family tree structures to be extracted with any desired confidence level; e.g. conservative trees containing only links with an associated high confidence level. Many automated approaches are based on a technique known as entity resolution in which actors that appear on multiple certificates are unified and established to represent a single individual. We have been experimenting with an alternative automated technique which we call linkage based on familial groups which is akin to the techniques used by demographers when performing linkage by hand. Our process starts by gathering parents and their siblings into bundles with the aim of (as near of possible) partitioning the certificates into familial groups. This may be achieved by bundling marriage and birth certificates according to a signature derived from their attributes. This is similar to but different from blocking used in most entity resolution schemes where certificates of one kind are gathered together. We have experimented with these techniques using hand coded data from an historic Scottish dataset as a gold standard for comparison. In this paper we will report on our techniques and some preliminary results from our experiments.

This work was presented at the following workshop.

Venue The Workshop for the Systematic Linking of Historical Records
Location University of Guelph, Canada
Date May 11-13, 2017
URL http://recordlink.org