

Probabilistic Record Linkage for Genealogical Research

John Lawson, David White, Brenda Price and Ryan Yamagata

Abstract

The most slow and tedious job in genealogical research is searching civil or church records for information about an individual. But, this is an essential step in research. By searching multiple sources such as census records, wills, deeds, birth and death records we can compile a more complete set of information, and potentially the pedigree of an individual. When records are stored electronically modern methods of probabilistic record linkage can combine or link all the information on an individual from various sources in seconds, rather than requiring days or weeks of arduous searching by a genealogist. Researchers in England, Canada and the U.S. Census Bureau developed the theory for probabilistic record linkage to aid in constructing pedigrees of individuals from vital records, in order to track hereditary diseases. However, probabilistic record linkage has yet to be widely applied to most sources of information used for common genealogical research.

This paper is the summary of the results from two Master's Projects in the Department of Statistics at Brigham Young University. Here, we describe the approach to probabilistic record linkage used by the Family History Department of The Church of Jesus Christ of Latter-day Saints in TempleReady™, and demonstrate its application to genealogical research using a set of civil and church records of Quakers in Perquimans and Pasquotank Counties, North Carolina. The results of our study are very promising. Probabilistic record linkage has the potential of dramatically increasing the productivity of genealogical researchers. Although complete automation of genealogical research is a way off, probabilistic record linkage could revolutionize the way research is done. This paper is a report of a work in progress; describing what has been done to the present, and outlining some of the many tasks yet to be addressed.