

Genesis is just one piece of the larger *Distributed Family Tree* project, an effort to connect the many isolated online genealogical databases into one cohesive family tree.

With minimum impact on the existing infrastructure, our goal is to establish “hyperlinks” between online genealogical databases that talk about the same individual. Each such link connects two otherwise independent family trees into one. Subsequent links cluster family trees until eventually they are all connected. While it is hoped that the developers of genealogical systems will extend their models to support these links, they are not required to: we are also developing open source server software to broker these links externally.

Genesis plays the important role of facilitating the creation of links. It does so in a three step process:

- 1. Import genealogical data of interest to the user**

Data is imported using web services for systems that support them, and using techniques developed by the *BYU Data Extraction* research group for those that do not.

- 2. Analyze the data and extend it with conclusions**

Individuals are linked together using the machine learning record linkage technique developed by Burdette Pixton, a recent BYU graduate. The user, who presumably has greater knowledge about the people involved, is given the opportunity to accept, reject, or punt on recommendations made by the system.

- 3. Publish the results back to the Internet**

Results are published to a designated server that supports the linking model.

Genesis can do much more, however! Each of these steps can be extended with plug-ins to support other systems and perform other kinds of analysis. For example, Charla Woodbury, a BYU graduate student, is currently developing an expert system with lexicons and rules that supports research of English and Danish ancestors. This degree of extensibility makes Genesis an ideal open platform for genealogical research assistance.

Find out more by visiting us at:

www.dftproject.org