Genealogies as a Digital Library Index

SILS 490-161
Fall 2013
Friday 9:00-11:45 AM
Manning - 304

Reagan W. Moore
Director, Data Intensive Cyber-Environments Center
Professor, School of Information and Library Science
22 Manning
919 962 9548
rwmoore@renci.org

Office Hours: Friday 1-2 PM
Other Times by Appointment

Course Description

This course has several goals:
- Develop an understanding of how to generate a genealogy index into a collection of material (Digital Library or Archive).
- Develop an understanding of the properties of genealogies
- Develop an understanding of the basic algorithms used to analyze genealogy properties
- Evaluate core genealogies that can be used as the basis for a genealogy index.

The course materials leverage 20 years of genealogy research and comprise:
- "Core Genealogies" book that explains genealogy concepts
- 241,826 person research genealogy that covers US presidents, royal lines of Europe, and cultural groups of Europe
- CoreGen2 computer program that generates core genealogies from the research genealogy, and analyzes properties of genealogies
- Lineages of historical persons listed in Palmer's "History of the Modern World", which will be used as an initial digital library index.

Course Structure

Genealogies as a Digital Library Index
This course will be structured around a half-semester class in which we will investigate the feasibility of using a genealogy as an index for a digital library. The starting point for the index is a genealogy that links 300 historical persons (from the textbook, "A History of the Modern World" by Palmer and Colton) to a core genealogy.

You will also complete a project to link:

- the ancestry of two significant persons (politician, actor, businessman, judge) to a core genealogy

You will be able to also look at your family ancestry and explore links to a core genealogy.

Guidelines for the Class

This goal of this class is to explore the feasibility of using a genealogy as an index into a digital library. A previous analysis has shown that roughly 42% (300) of the historical persons listed in the Modern European History Book can be linked to a common research genealogy. We will use web and historical resources to explore relationships between the remaining historical figures. A core genealogy that links historically important persons can serve as the basis for a digital library index. Research questions include:

- What fraction of the persons mentioned in a digital library can be linked to a common core genealogy
- How large is the core genealogy (number of persons) that is the basis for the digital library index?
- What fraction of the class is able to link their ancestry to the same core genealogy?

The course material will be shared through Sakai and the SILS LifeTime Library and will be accessible through a web browser. Note that the same data grid technology used in the LifeTime Library is also being used to support national data grids, national archives, and national libraries.

Sources for genealogy information range from Wikipedia, to Google, to the Family History Library maintained by the Church of the Latter Day Saints, to popular genealogies that have traced royal descents for Americans, to material specific to your family such as family bibles, marriage certificates, and family traditions. Other sources include census records, social security records, history books, public records, NARA archives, and royal genealogies of Europe. Examples of the latter include Europaishe Stammtafeln, Cockayne's Peerage of Great Britain, Collins Peerage of England, and Richardson's Plantagenet Ancestry. Genealogy information from all four sources have already been ingested into the research genealogy and can be used to create a digital library index.

The best way to learn about genealogies is to construct one. By building upon the 241,826-person research genealogy, properties about genealogies can be discovered and evaluated. We will seek additional lineages that can be tied into the research genealogy. We will discuss possible index completeness criteria. Is a digital library index complete if it:
- Links prominent people to persons mentioned in the books in a digital library?
- Identifies relationships between the authors of the books in a digital library?

- Identifies relationships between your ancestry and the books in a digital library?

We will examine whether a genealogy can be used as an index into historical events, and thus serve as an index into records that document events, or equivalently serve as an index to an archives, or as an index to a digital library.

Using our Class Time

Each weekly class will start with a question on the previously presented material, followed by up to 60 minutes discussion of the reading material. We will then apply the remaining time to constructing a genealogy index, Tasks include generating and evaluating genealogies for personal ancestry, or lineages for historically important persons, or lineages for different cultures. Our goal is to identify criteria for evaluating the use of a genealogy as a digital library index, based on practical exerience in creating an index.

Notable lineages that are discovered will be presented during the class. A good example is a lineage that links Christopher Columbus to the research genealogy. Columbus was an in-law of a descendent of the Sinclair family of the Orkney Islands. The Sinclair family tradition records that Henry Sinclair took a fleet of 12 boats to Rhode Island in 1398. The supposition is that Columbus knew where he was going, based on family tradition. Similar interactions between family tradition and historical events are of interest.

Individual participation is strongly encouraged. This can include discussion of problems in selecting sources, or questions on genealogy properties. Each student will present the results of their research efforts during the class.

Preparation:

Each student is expected to invest time in researching their genealogy outside of class. All results will be provided as a gedcom 5.5 file for integration with the research genealogy. The Gedcom file will be deposited into a shared collection that is managed by the SILS LifeTime Library. One expectation is that the class members of European descent will be able to demonstrate that they are at least 20th cousins. A second expectation is that multiple sources will be found for each person and analyzed to determine reliability of the information. All sources of information will be documented through annotations on each person and each date.

Attendance:

Class participation is important. I will take roll each week. I strongly encourage each student to ask questions, present results, and discuss challenges encountered in doing the research. The discussions can inform other class members of approaches that either worked or did not work.

Grade Determination

Class presentation on genealogy index   30%

Family and / or significant person links to core genealogy	30%
Final Exam	20%
Class participation	20%

Course Materials:
Text - Reagan Moore, "Core Genealogies", provided as a pdf file.
Research genealogy of 241,826 persons, provided as a Gedcom 5.5 file.
CoreGen2 program for analyzing gedcom files, provided as a Windows
executable. This uses a Windows Visual C++ .NET GUI.
You will need to select a genealogy program that can import Gedcom files.
Examples include Reunion, Family Tree Maker.

Class Schedule
Please note: this schedule is subject to change.

Friday August 23
Course introduction – Core Genealogies Book
Introduction to SILS LifeTime Library (digital library for course material)
Introduction to genealogy analysis program – CoreGen2
Examples of the use of a genealogy as an index to a digital library
Initial selection of historical persons to research

Friday August 30
Reading: Core Genealogies, Chapters 1-6.
Analyzing properties of large genealogies
Access to SILS LifeTime Library for posting class material (bring laptop)
Example digital library index based on Palmer's History of the Modern World
Using the CoreGen2 Program to analyze genealogies (bring laptop)
Error analyses and statistical analyses
Finding ancestors or a Root Person
Class research session on finding your links to a core genealogy

Friday September 6
Reading: Core Genealogies, Chapter 7-10
Finding descendents of an Ancestral person
Finding ascents from a Root Person
Finding descents from an Ancestral person
Relatives of a Root Person
Class research session on finding your links to a core genealogy

Friday September 13
Reading: Core Genealogies, Chapter 11-13
Lineages between a Root Person and an Ancestral Person

Sources for a Root Person
Titles for Ancestors of a Root Person
Class presentations on lineages identified to core genealogy

Friday September 20
Reading: Core Genealogies, Chapters 14-16
Core Genealogy based on Royal Lines
Core Genealogy based on Magna Charta Barons
Core Genealogy based on U.S. Presidents
Class presentations on lineages identified to core genealogy

Friday September 27
Reading: Core Genealogies, Chapters 17-18
Genealogy as an Index, and building your family genealogy
Comparison of Core genealogies
Using a Genealogy as an index to a digital library
Class presentations on lineages identified to core genealogy

Friday October 4
Final Exam covering core genealogy concepts. A set of questions will be published at the start of the course, from which the final exam questions will be selected.