Record Linkage and Tagging for the BYU Historic Journals Project
(journals.byu.edu)

Douglas J. Kennard   and   Dr. William A. Barrett
(BYU Computer Science Department)
Historic Journals - Introduction

- How would you know?
- Where is it?
- What does it say?

For Repeete Tahiti, about 3 coast one to Kapeka, one to Fakaka, one to Hare e. Went aboard at 10 o'clock P.M. and at 12 M. we left the pier. Mr. Elder Kenneth Stevens of Papeete, Tahiti and George W. Billings of Pearl, Utah and I are in room 36 in second class passenger. We have quite a nice room. The weather has been cloudy and quite cold but not much wind, all well at 7 A.M. On board U.S. Inume, Puengage. Jan. 23, 1920.

Went exploring along this the the water in old. The course goes 327 miles up to noon today. The weather is cloudy and cold. We have had no wind, but not much wind. I am feeling better tonight.
Historic Journals - Introduction

- Who might care?

900 living descendants! (my great-grandfather)
Historic Journals - Introduction

- Who might care?

900 living descendants!

(my great-grandfather)

Descendants of people he wrote about!
Historic Journals - Introduction

My questions:

Which of my ancestors wrote a journal?

Did anyone else write about my ancestors?

How can we share (within limits) diaries?
journals.byu.edu

Get to know your Ancestors
Your ancestors are more than just names, dates, and places. Find out who they really are - how they lived, what they thought, and how they felt - by reading their own personal diaries, letters, and biographies. Read what their friends, relatives, and acquaintances wrote about them, too. Do it quickly and easily here on the Historic Journals website for free. Help other people do the same thing.

Discover forgotten History
Much of what we know about historical people comes from what other people wrote about them. Our approach to linking journal entries (and other materials) to who they are about makes it easier than ever to discover even the most obscure references to the people you are interested in.

Click here to find out how to create an account
Get to know your Ancestors

Your ancestors are more than just names, dates, and places. Find out who they really are - how they lived, what they thought, and how they felt - by reading their own personal diaries, letters, and biographies. Read what their friends, relatives, and acquaintances wrote about them, too. Do it quickly and easily here on the Historic Journals website for free. Help other people do the same thing.

Discover forgotten History

Much of what we know about historical people comes from what other people wrote about them. Our approach to linking journal entries (and other materials) to who they are about makes it easier than ever to discover even the most obscure references to the people you are interested in.

Click here to find out how to create an account
Previously Described Details

JCDL 2009  (Joint Conf. on Digital Libraries)

FHTW 2009  (fht.byu.edu)
Person identifier: KWCF-QGH
BYU Historic Journals

Share / Collaborate
Scanned Journals (images)
Transcriptions
Reference Information
Tag people with PersonIDs

Search for writings by or about ancestors

PersonIDs (FamilySearch)

API
Tagging who is written about
(similar to tagging on social networks, but PIDs)

Person identifier: KWCF-QGH
Historical Social Network
Rosters (implicit connections)

Roster: Brigham Young's Pioneer Company - 1847
Rosters (implicit connections)

Roster: Brigham Young's Pioneer Company - 1847
Rosters (implicit connections)

Military  (ex: Captain Stout’s Army, Revolutionary War)

Community (ex: Bastrop, TX, USA)

Church (ex: Bannockburn Baptist Church, Snowville, AK)

Team (ex: 1980 US Olympic Hockey Team)

Class (ex: Davis High School, Mr. Smith English class, Fall 1983)

Work (ex: Austin, TX, Joe's Grocery Store, 1980-1983)

Other (ex: Brigham Young’s pioneer company, 1847)
Record Linkage

Manual Tagging / Crowd-Sourcing

Davis Bitton “Guide to Mormon Diaries...”

BYU HBLL - Overland Trails

Mormon Missionary Diaries

Automatic / Semi-automatic Record Linkage

F. Esshom “Pioneers and Prominent Men of Utah” (photos)

lds.org “Mormon Overland Travel” pioneer DB
Guide to Mormon Diaries and Autobiographies (Manual Tagging)

Reference book of 2,894 known diaries

Alphabetical (last name)

Where to find the diary

Synopsis of content / bio info
Guide to Mormon Diaries and Autobiographies (Manual Tagging)

Reference book of 2,894 known diaries

Alphabetical (last name)

Where to find the diary

Synopsis of content / bio info

893. Gleason, John Streator
Diaries (1857–1863)
Diaries (1857–1863)
Microfilm of Holograph. 162 pp. HDC [LDS Church Archives] (Ms f 183 item 1)

Guide to Mormon Diaries and Autobiographies (Manual Tagging)

Reference book of 2,894 known diaries

Alphabetical (last name)

Where to find the diary

Synopsis of content / bio info

892. Gillett, John L. [caption]

Arab biography (1857–1858)

Autobiography (1857–1858)

Photocopy of typescript. 147 pp, Princeton Library, Princeton U. Typescript. 134 pp. (Utah His Soc) (Mum A 77)


Not all first-hand observation. One section goes back to tell the history of Mormonism, seen as a fraud. Written about 1903. Comment by Charles Kelly in 1959 suggests that the account, while generally reliable, drew upon "old man's imagination." Repetition. Utah War is gone over a second time, with quotations from government reports. Details on Mormon-Indian cooptinon.

893. Gleason, John Streator

Diaries (1857–1863)

Diaries (1857–1863)

Microfilm of Holograph. 162 pp. HDC [LDS Church Archives] (Ms f 183 item 1)

Guide to Mormon Diaries and Autobiographies (Manual Tagging)

Manual search in new FamilySearch using:

David Bitton's Guide
pioneer DB on lds.org

Mormon Pioneer Overland Travel, 1847–1868

Gleason, John Streator

- Birth Date: 13 Jan. 1819
- Death Date: 21 Dec. 1904
- Gender: Male
- Age: 28
- Company: Brigham Young Pioneer Company (1847)
Guide to Mormon Diaries and Autobiographies
(Manual Tagging)

1,500+ tags (so far)

Description: This is a reference book by Davis Bitton, published by the Mormon Church. It contains autobiographies that are in several libraries. Information about the material. A PDF version of this reference book is available online, just reference information about how to find them. Diary authors would love help tagging the rest. The page numbers listed on the pages are:

- Page 9 #1: Go there (Tag details) James Smith A
- Page 9 #2: Go there (Tag details) James Smith A
- Page 9 #3: Go there (Tag details) Myron Abbott
- Page 10 #4: Go there (Tag details) Myron Alma A
- Page 11 #5: Go there (Tag details) William Elias

1548 Person Tags found in this journal: show/hide
Mormon Missionary Diaries / Overland Trails

433 Diaries

PersonIDs of authors:

used online biographical info to search

PersonIDs of people talked about:

“Crowd-source” the tagging

Needed: bigger crowd to help tag
Pioneers and Prominent Men of Utah
(Frank Esshom, 1913)

5,894 photos

Start: PDF with (poor) OCR
Auto extract photos (3x3 grid)
Manually correct obvious crop/OCR errors
Auto parse:
name, birth, parents' names
API search, store PersonIDs for “close” matches (up to 3)
Pioneers and Prominent Men of Utah
(Frank Esshom, 1913)

5,894 photos

Start: PDF with (poor) OCR

Auto extract photos (3x3 grid)

Manually correct obvious crop/OCR errors

Auto parse:
name, birth, parents' names

API search, store PersonIDs for “close” matches (up to 3)
Pioneers and Prominent Men of Utah
(Frank Esshom, 1913)

5,894 photos

Start: PDF with (poor) OCR
Auto extract photos (3x3 grid)
Manually correct obvious crop/OCR errors
Auto parse:
name, birth, parents' names
API search, store PersonIDs for “close” matches (up to 3)
Pioneers and Prominent Men of Utah (Frank Esshom, 1913)

5,894 photos

Start: PDF with (poor) OCR

Auto extract photos (3x3 grid)

Manually correct obvious crop/OCR errors

Auto parse:
name, birth, parents' names

API search, store PersonIDs for “close” matches (up to 3)
Pioneers and Prominent Men of Utah (Frank Esshom, 1913)

5,894 photos

Start: PDF with (poor) OCR
Auto extract photos (3x3 grid)
Manually correct obvious crop/OCR errors
Auto parse: name, birth, parents' names
API search, store PersonIDs for “close” matches (up to 3)
Mormon Overland Travel pioneer database (on classic.lds.org)

Browse by company
Search by person name (one at a time)
Mormon Overland Travel pioneer database

Goal: index by PersonID, provide hyperlink back to the database

Code to automatically find the PersonIDs

Status: proof of concept, one pioneer company, no links on our site
Crawl the roster to get:
Name, Birth, Death, URLs of Trail Excerpts

Use FamilySearch API to search

Store PersonIDs for “close” matches (up to 3)

Manually verify results
Automatic Record Linkage

Observation: only use best match (not best 3)

Results:

Total People: 148

- Correct: 127
- Unsure: 7
- Incorrect: 9
- Not found: 5
Automatic Record Linkage

Results very promising: auto-link entire database

Unanswered Questions:

First pioneer company (better records?)

Only 3 women, 2 were wrong (maiden vs married?)
Future Work

Investigate auto-linking more document types
- Census
- Birth records
- Death records

Semi-automatic tagging
- find names in diary
- compare to family names, other resources (rosters, city dir., census, news, etc.)
- ranked suggestions in tagging tool
Thank You
Incorrect: 9

3 - sons of person (2 of which were juniors)

2 - Females (maiden vs married name?)

2 - completely wrong

1 - James Cox instead of James Case, but had a James Case as an alternate name

1 - different guy with same 1st / last name and born / died within a year of the same dates