Optimizing Your Site for Search Engines

RootsTech Presentation Syllabus

- The Problem
  - Only a small percentage of genealogical content is indexed by search engines
  - While Google does not release internal data, mocavo.com claims “While Google is the go-to search engine for most of us on a daily basis, they only index about 5% of the web’s genealogical content.”
  - Further, even content that is indexed might not be scored as highly relevant to the search being performed
  - Therefore, your genealogical content may never appear in search results

- Why It Matters
  - Most information searches on the web happen through search engines such as Google
  - Search engines drive traffic to your site—if you appear in the search results
  - Those new to genealogy especially will not know of specialized genealogy sites
  - People don’t search for “New Zealand Family History” or “Genealogy Indexes”; they search for specific people, places, and dates
  - If your site doesn’t rank well for such searches, you will miss a lot of traffic

- How Search Engines Work
  - Crawling and Indexing:
    - Crawl the Web by following links
      - Only public links are followed
      - Only public pages are visited
      - No link means no crawl! (e.g., forms can be a barrier to crawling)
    - Index each page (map words to pages)
    - Analyze page, detect spam, determine content value
  - When user searches:
    - Match search terms to pages using the index
    - Rank results based on relevancy and value
    - Display top results

- Ways This Can Fail for Your Genealogy Site
  - Site cannot be crawled
  - Pages are not indexed
  - Pages receive low ranking
  - Site/Pages not deemed relevant or valuable to search terms
● Why Your Genealogy Site Does Not Get Crawled
  ○ Explicit exclusion: paywall, login, robots.txt
  ○ Unintended exclusion: search form, broken links

● Why Your Genealogy Pages Do Not Get Indexed
  ○ Indexes expensive to build and maintain so heuristics are used to avoid indexing pages not deemed “valuable”
  ○ Content appears to be spam, especially problematic for auto-generated pages
  ○ Low information content (e.g., many names or links on an index page)
  ○ Poor internal site linkages (pages don’t reference each other)

● Why Your Genealogy Pages Are Ranked Low
  ○ Few outside links to specific pages
    - How many people will link to “Robert Gardner, Jr. born 1828”??
  ○ Too many links, poor use of highlighting, poor URLs, etc.
  ○ Low information content (e.g., most of page is navigation, ads, etc.) or important content not prominent in the page
  ○ Few links from outside pages; especially problematic for uncommon names
    - Impossible if pages require login

● Why Your Genealogy Pages Have Low Relevancy
  ○ Search engine doesn’t recognize search as a genealogy search
  ○ You are competing with current events, news, etc.
  ○ Solution: User needs to indicate intention in search (e.g., add ~genealogy)

● How to Improve Search Ranking
  ○ Improve Crawling, Indexing, Perceived Relevance
    - see webmaster tools
      ● robots.txt, broken links, provide static links and cross-links, link important content from higher-level pages, focus on unique and relevant data, make relevant content prominent, make it easy for users to link to your pages, spammy pages, too many links, keywords, etc.
  ○ Genealogy-specific
    - How to work around the paywall or login
      ● some data must be public
      ● consider First Click Free, used by many news services
    - How to work around the search form
      ● Sitemaps
- Sitemaps link to public page with minimal information
- Public page links to private page with full information
- This exposes data to search engine, but don’t forget internal linkages!
- Add genealogy microdata tags
  - Tags indicate site is genealogy-related and signal to the search engine presence of interesting content to index.
  - Indicates relevancy to genealogy searches, so search engine can make a better match for such searches

- Announce historical-data.org
  - Proposed genealogy microdata standard based on schema.org
    - HistoricalPerson, HistoricalFamily, HistoricalEvent, HistoricalRecord
  - Google and other companies can develop tools to make use of this data, enhancing the value of your content
  - Sample markup:
    <div itemscope itemtype="http://historical-data.org/HistoricalPerson.html">
      <img itemprop="image" src="http://example.com/freeman.jpg"/>
      <meta itemprop="url" content="http://example.com/people/Morgan-Freeman"/>
      <meta itemprop="gender" content="m"/>
      <span itemprop="birth">
        itemscope itemtype="http://historical-data.org/HistoricalEvent.html"
        itemref="birth_date birth_location"</span>
      <span itemprop="name">Morgan Freeman</span>
    </div>
    ...
    <time id="birth_date" itemprop="startDate" datetime="1937-06-01">June 1, 1937</time>
    <td id="birth_location" itemprop="location" itemscope itemtype="http://schema.org/Place">
      <span itemprop="name">Memphis, TN, USA</span>
      <span itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
        <meta itemprop="addressLocality" content="Memphis"/>
        <meta itemprop="addressRegion" content="TN"/>
        <meta itemprop="addressCountry" content="USA"/>
      </span>
      <span itemprop="geo" itemscope itemtype="http://schema.org/GeoCoordinates">
        <meta itemprop="latitude" content="35.1495343"/>
        <meta itemprop="longitude" content="-90.0489801"/>
      </span>
    </td>
- Announce Sample Chrome Historical-Data plug-in
- Demo
- Debugging Tool
- Example of ways historical-data tags can be used in applications