

**Poster/Table Talk Proposal
2013 ASDAL Conference**

With Full-text, Who Needs an Index?

Abstract

Increased interest in Ellen White's letters and manuscripts appears to justify an effort to make available an electronic index of that literary corpus. In what may appear an exercise in futility, such a project began to be undertaken last summer at Southwestern, even with a knowledge that the White Estate would shortly set a timeframe to make available unauthenticated full-text web access to those records. Nevertheless, while full-text search is expected to be available by the summer of 2015, metadata remains a very useful access point to content and can be expected to aid researchers well into the future. Building upon previous efforts by the White Estate to create various paper-based finding aids and then turning them into electronic format, the project is not an effort to create the index from the ground up. Instead, it consists in bringing a legacy database up to standards in order to make it both available online and easy to migrate to other systems.

**By Alfredo Vergel
Public Services and Special Collections Librarian
817-202-6521 • avergel@swau.edu**

**Southwestern Adventist University
101 W Magnolia St
Keene, Texas 76059**

With Full-text, Who Needs an Index?

Twenty minutes into class time. My eyes traveled around the room scanning for someone who knew what was going on. All I could see were the all-too-familiar blank stares. My technical services professor had a rare teaching style. Her bewildering sink-or-swim approach was paired with a penchant for shockingly blunt responses. Silent rage and adamant shrugs were commonplace among my classmates. I was committed to get the most out of the course though, and did so on many levels. Class sessions were not entirely a waste, and the rigors imposed by an uncompromising instructor led me to take charge of my own learning. I am even able to recall specific class discussions seared into my memory in a way that never happened with my more lenient professors.

One such discussion was on the topic of access points. Using that jargon with the expectation to retrieve a traditional answer, Dr. Rofofsky prodded the class to name the access points of a bibliographic record. Clueless and theretofore unfamiliar with term, or the topic for that matter, I said “the full text.” Heresy! A lecture ensued on my impractical and legally shaky proposition. It wasn’t such a novel idea, mind you. Google had already scanned over five million books. Nevertheless the viability of that project was all but certain as legal challenges mounted on every corner. Fast forward five years. At last count, Google has scanned over 30 million titles, and legal impasses, while not totally gone, are a faint semblance of what they were.

We have become used to searching the full text. We demand it. Ellen White’s literary corpus has not escaped the trend. Not content with unfettered free access to her published writings, both ends of the Adventist ideological spectrum have decried for years the reluctance of the White Estate to release electronic access to the full collection of her letters and manuscripts. As we all know, it took hacking into the electronic vault, so to speak, to get the White Estate to finally commit to it. 2015 is the magic year. So why is the Ellen White Research Center at Southwestern developing an electronic version of the index to Ellen White’s letters and manuscripts? Isn’t it a little too late to care about an obscure bibliographic resource heretofore only available in card catalogs?

I know I’m preaching to the choir, but indulge me as I highlight the continued value of metadata. As recent news reports about the National Security Agency’s surveillance program have illustrated, metadata matters a big deal. The Electronic Frontier Foundation has put it most cleverly:

- They know you rang a phone sex service at 2:24 am and spoke for 18 minutes. But they don’t know what you talked about.
- They know you called the suicide prevention hotline from the Golden Gate Bridge. But the topic of the call remains a secret.
- They know you spoke with an HIV testing service, then your doctor, then your health insurance company in the same hour. But they don’t know what was discussed.
- They know you received a call from the local NRA office while it was having a campaign against gun legislation, and then called your senators and congressional representatives immediately after. But the content of those calls remains safe from government intrusion.

- They know you called a gynecologist, spoke for a half hour, and then called the local Planned Parenthood's number later that day. But nobody knows what you spoke about.

Metadata makes it easier to determine the substance of something without actually handling it. My children know it all too well as their illiterate brains scan the supermarket shelves for the telltale bubble writing that package designers know appeals to them. Our consumer society relies heavily on labels and product descriptions. You don't pick up that box of cereal without first reading its list of ingredients—or at least you shouldn't. Have you looked at the salt content of your veggie sausages lately?

We rely on metadata to shop for a home. We want to know the square footage, number of bedrooms, bathrooms, lot size, etc. before setting foot on any property. At a time when anything can be purchased online, metadata has become even more useful. Product specifications are essential in today's web-based markets. Services of all kinds too are consistently broken down into profiles that help us select what fits our needs and wants. It is no surprise that the federal government's latest attempt to influence the healthcare market involved releasing its huge Medicare provider charge dataset. It is clear that metadata is not going anywhere, and this is true in libraries as much as anywhere else.

To be fair, full-text searching has been a boon to information retrieval, and it too is undeniably here to stay. It has opened new possibilities for research that used to be too labor intensive to be practical. As an example, the on and off discussion of what Ellen White meant in her amalgamation statements would benefit from a look at how her contemporaries used the term. Whether we believe Ellen White to be nothing more than a rather intellectually dependent writers or someone who strived to translate revelation into the language of her day, this should prove to be an instructive exercise. I find it interesting that recently published articles, such as Ron Osborn's 2010 Spectrum piece neglect the opportunity to explore this dimension of the topic.

A simple Google Books search will furnish parallel language by a host of both religious and secular authors. Both similarities and departures would emerge from an analysis of the use of the term that would in turn provide another layer to this hermeneutical discourse. Here are a few interesting instances:

"The sudden amalgamation of all [Christian] denominations would be productive of great evils."--Benjamin Parham Aydelott in *Incidental Benefits of Denominational Division*, 1846 p. 105

"It should be borne in mind, that what we term Caucasian races are not of one origin: they are, on the contrary, an amalgamation of an infinite number of primitive stocks, of different instincts, temperaments, and mental and physical characters." J.C. Nott and G.R. Gliddon in *Types of Mankind*, 1854 p.67

"Let any one who doubts the evils of this mixture of races, and is inclined with a mistaken philanthropy to break down the barriers between them, come to Brazil. He cannot deny

the deterioration consequent upon the amalgamation of races, more wide-spread than in any other country in the world, and which are rapidly effacing the best qualities of the White race, the Negro, and the Indian, leaving the mongrel non-descript type deficient in physical or mental energy." Louis Agassiz in *A Journey in Brazil*, 1868 p.293

"But the leading idea, as well as the hardest task, which Alexander had set before him was the amalgamation of his diverse subjects into one people."--Arthur Mapletoft Curteis in *Rise of the Macedonian Empire*, 1890 p.194

Back to metadata now. Even Google relies on it. It is an important part of its relevancy raking algorithms. We know that a Google search is not an actual search of the web, but rather a search of Google's index of the web, which includes not only the web content itself, but metadata such as referral traffic.

Last week Jim Ford told me about ideas floating around to streamline the assignment of subject headings to records in the Seventh-day Adventist Periodical Index by finding ways to automate or crowd-source the task. I told him that in my opinion subject headings were necessary and that quality human-generated indexing was a strength of the periodical index. Granted, I do from time to time find a full-text search of the General Conference's periodical archives useful, but not necessarily as a substitute to searching the periodical index.

Back to Southwestern's project. It is my sense that with the launch of a web version of the topical index to Ellen White's letters and manuscripts last fall, Southwestern's library filled a gap in online resources on Adventist history.

During the spring of 2012 we saw an exponential increase in the use by students of the topical card index to Ellen White's letters and manuscripts. That led to exploring the possibility of making the index available on the web. It turned out, there was already an electronic version of the index in an outdated file format that, while requiring quite a bit of work to use, obviated the biggest hurdle in completing the project.

The topical index, along with the its biographical and addressee counterparts, is believed to have been created in the 1930s, with additions since. At the time, the index undoubtedly represented an improvement over Marian Davis' scrapbooks, and perhaps it evolved from the work of Ellen White's most prominent assistant.

Having been tested in earnest by students and other patrons, the index has proved useful even beyond Southwestern. As recently as this month there have been repeated requests for making it more easily available to a wider audience, including creating an API that would let other campuses use the dataset in their own web applications. Similar deployments of the biographical and addressee indexes are scheduled for the next two summers.

I do not expect these indexes to get the level of use that the upcoming full-text release will probably get, but I am confident that for students and scholars, they will remain useful access points to Ellen White's literary corpus.