Enhancing search and discovery across libraries, archives and museums using text, data and audio analytics, using a single OneSearch service

1. About the Case Organization
The National Library Board of Singapore (NLB) manages the National Library, the 26 Public Libraries and the National Archives of Singapore (NAS). NLB promotes reading, learning and information literacy by providing a trusted, accessible and globally-connected library and information service through the National Library and a comprehensive network of Public Libraries. NAS oversees the collection, preservation and management of public and private archival records, including government files, private memoirs, maps, photographs, oral history interviews and audio-visual materials.

2. About the Challenge
Libraries and archives collect and provide access to digital resources of significant national, cultural and heritage values. Their collections are painstakingly curated, described and made accessible.

A seismic shift in the information seeking behaviour has taken shape. With the ever-expanding reach of the Internet, users are accustomed to quick and easy access to content, and are expecting an online experience that is rich and instant. How can libraries and archives continue to reach out to their users in this era of shorten attention span?

Every year, NLB users collectively contribute to tens of millions of e-retrievals. We see every single one of these interactions as a golden opportunity to ‘push’ relevant content to the user. The success of the ‘customers who bought this item also bought’ recommendation feature at amazon.com is a clear testament of the power of pushing relevant recommendations. However, to effectively connect people to content, we need to connect content to content first.

3. What We Did
To manually associate related content for any sizeable collection will be very labour-intensive. Text and data mining (TDM) technologies can automatically, efficiently and accurately identify these associations. TDM is akin to sieving through all the content, and applying well-established information retrieval theories to identify similar content.

NLB implemented TDM on its digital content, and identified over 1 billion similarity associations. We can now bring together the full spectrum of resources available in NLB regardless of institutions, formats and language and present them in a way that encourages a deeper understanding of the precious memories of the nation.

It is also what knowledge seekers have been dreaming of. They can spend less time gathering the pieces, and focus on digesting and analysing the ‘dossier’ of relevant information to derive new insights.

4. Challenges and Lessons Learned
The first hurdle that we needed to cross is the technical one. As we planned to use TDM for large data sets, we looked for solutions built on the popular Hadoop big data platform. The expertise in these technologies was lacking locally. We have our own Solutions Architect to work on proof-of-concepts to pick up the skills.

The next challenge came when we need to scale the TDM to process a data set of over 10 million items. It proved too much even after we scaled our Hadoop cluster to 13 servers. We went back to the drawing board, and figured that we could split the data set into smaller clusters of related ones before the TDM processing. This ‘divide-and-conquer’ approach worked very well.

5. Impact and Benefits
With over 1 billion associations identified via TDM, a massively inter-connected network of knowledge has been formed.

The related resources identified via TDM are made easily available to the users. The users are now able to view and explore related resources within and across collections and media formats. Information is no longer viewed in isolation, but seen as a part of the larger context. We call this contextual discovery.

The initial results have been promising. The level of access to the resources has increased.
6. Next Steps
We are exploring several possibilities to bring contextual discovery to the next level:

- **Cross-institution.** With NAS coming under NLB, we now have an expanded collection of digital resources on Singapore. We will be working on linking NAS and the National Library resources.
- **Cross-language.** There are 4 official languages in Singapore: English, Chinese, Malay and Tamil. We are considering the use of machine translation to provide cross-language recommendations.

12) MARK GLIKSON

**Deploying a Semantic Operating System in a small financial services company**

1. About the Case Organization
O-Pulse provides a portal for investment advisors to manage client relationships, compliance, and communications with portfolio managers. Currently in private Beta, O-Pulse was founded by a leading third party money management & financial advisory firm based in North Carolina.

2. About the Challenge
Typical of many conventional Enterprises today, the founders of O-Pulse, in their existing financial advisory firm, recognized the need to transform their traditional offline organization into a modern digital enterprise. Management faced increasing pressure, with a noticeable uptick in client attrition to competitors with even the most basic online client portals. O-Pulse was established to create an online portal better servicing their existing advisor clients, and the wider US-based independent investment advisor market.

The v1 O-Pulse portal was designed to facilitate complex targeted content distribution and syndication between portfolio managers, investment advisors and their clients. News, articles, portfolio commentaries and financial updates could be curated, shared and syndicated through one central dashboard with full compliance reporting. A key requirement was to enable white label deployment of the O-Pulse portal to any number of organizations under a SaaS subscription model.

O-Pulse engaged with multiple development firms only to be faced with exorbitant open-ended quotations. All proposals utilized a common “silo” build methodology, inflating costs throughout the development cycle with limited asset re-use and extensive
Enhancing search and discovery across libraries, archives and museums

OneSearch & Text Analytics

Kia Siang Hock
Deputy Director
Technology & Innovation
OneSearch – Libraries, Archives & Museums

- Find resources across institutions
  - National Library
  - National Archives
  - Public Libraries
  - Public Museums

- ‘Bento-style’ search result display
  - Resources are not ‘hidden’ in one long list of search results

- Responsive web and flat design to cater to all devices

http://search.nlb.gov.sg
OneSearch Content Coverage

**NLB Subscribed Resources**
- eBooks
- eJournals, Databases

**NLB Resources**
- National Library (Digital SG Content)
  - Infopedia
  - BookSG
  - NORA
  - PictureSG
  - NewspaperSG
  - Singapore Memories
  - Web Archives
  - MusicSG

**National Archives**
- Posters
- Oral History Interviews
- Government Records
- Audiovisual Recordings
- Photographs
- Maps & Building Plans
- Straits Settlements, Overseas & Private Records
- Speeches & Press Releases

**NHB Resources**
- Digital records of artifacts and artworks from National Heritage Board museums

**Books, CD, Magazines**
Tan Swie Hian
NLB users retrieved tens of millions of e-content every year

It would be really nice if we could convert every single e-retrieval instance into an enriching discovery experience for every single user every time...
The Cenotaph, located at Esplanade Park along Connaught Drive, is a war memorial which commemorates the sacrifice of the men who perished during World War I and World War II. It was unveiled on 31 March 1922 by the Prince of Wales. On 28 December 2010, it was gazetted as a national monument together with two other structures in...

Lim Bo Seng (Article)
Major-General Lim Bo Seng (b. 27 April 1909, Nan Ann, Fujian, China - d. 29 June...

Dalhousie Obelisk (Article)
Dalhousie Obelisk, landmark, located at Empress Place in the Central Region. The tall...

Gwee Peng Kwee
His daily routine school...
Laying of foundation stone and unveiling of Cenotaph...

Newspaper articles
- Singapore’s War Memorial to the Glorious Dead (11 Nov 1920)
- Singapore’s War Memorial (21 Sep 1921)
- Arrival of the Prince (31 Mar 1922)
- Lest we forget (8 Nov 1953)
- Singapore students learn to care about history (13 Jul 1997)
Using text analytics to automatically identify related content

Text tokenised; tokens parsed and weighted (TF/IDF)

Weighted tokens similarity computed

Text tokenised; tokens parsed and weighted (TF/IDF)

Similarity = 0.295
Using clustering to handle large datasets

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

Mahout K-Means Clustering with Cosine Distance
Implementations of Contextual Discovery – over 1 billion associations identified

*Implemented in Jul 2013*

*Implemented in Nov 2013*

*Implemented in Jul 2014*

*Implemented in Sep 2013*
Increased usage of content

Referrals from Infopedia: 0.14%
Pageviews per month: 37,841
Pageviews per visit: 3.64

Infopedia

PictureSG

10.65% (after 6 months)
Thank You!
kia_siang_hock@nlb.gov.sg