

Thematic & Selective Web Archiving Efforts in Singapore

- Study objectives:
 - Practice of selective and thematic Web Archiving in Singapore - focusing on harvesting/collection issues.
 - Assess need and potential of collaboration
- Background:
 - National Library Board (NLB) selective & thematic Web archiving
 - Asian Tsunami Web archiving by IA/NTU/NYIT/UW*
- Methodology: Survey Questionnaire
- Status: first three test cases
 - S. Rajaratnam School of International Relation (RSIS)
 - Lee Kuan Yew (LKY) School of Public Administration
 - National University of Singapore

* Wu, H. J. & Heok, A K. Documenting Online Collaboration between Researchers and Information Professionals: The Case of the Asian Tsunami Web Archive. Singapore Journal of Library and Information Management. 34, 75-90.

Survey Questionnaire (I)

- Q1. In determining which website may be of interest, is any of the following methods used?
 - A. Social Network Analysis/Link Analysis (of the actors)
 - B. Expert/Researcher Network, who knows the domain
 - C. Institutional/Community Voices, known institutional and community Websites
 - D. Link/Text mining from existing resources
 - E. Other Methods _____

If multiple methods are used, which one is the dominant one: _____

Survey Questionnaire (II)

- Q2. How often are the materials collected during the collection process?
 - A. As often as required
 - B. Upon some triggering events
 - C. Periodically, the period is _____
 - D. Other scheduling: _____

If multiple methods are used, which one is the dominant one: _____

- Q3. How is the expiration of the collection period determined?
 - A. Never
 - B. Depending on the update frequency of the websites of concern
 - C. A fixed duration _____

Survey Questionnaire (III)

- Q4. Are there any technologies or systems (such as, Web Curator Tool) being used in managing the curation process? If not, what technologies or systems would be desirable to better your collection methods?
- Q5. How would you like the above methods to be improved?
- Q6. What is the estimated size of the current collection?

In terms of number of websites: _____

In terms of bytes: We did not collect data on this.

Response as of now

- Sent out three test cases, one responded so far: LKY School of Public Administration – Study on Singapore’s Political Bloggersphere
- Q1: All selection methods are used: social network/link analysis, expert network, institutional/community voices, but mainly using text/link mining from existing resources:
- Q2. They will collect websites upon some triggering events
- Q3. Depending on the update frequency of websites, they will terminate the collection.
- Q4. They have not explored using any tools or technologies
- Q5. They believe their methods so far is sufficient
- Q6. The collection size of around 70 websites.

Observation & Further Thoughts

- Survey is still on-going – welcome your comment to refine it.
- Based on personal contacts with researchers (not yet reflected in the survey), Singapore researchers are building their own web archives for their research, mostly in traditional ways – manually
- Internationally, some researchers are starting to use the collection from culture institution.
 - There is an international community called Association of Internet Researchers (AOIR)
- Thus, should relationship with researchers be promoted, e.g., researcher use-case for collection development in addition to that for access?
 - IIPC to (re-)start Researcher WG?