The Archiving Method for
Records of Public Sector's Facebook Page

Yun-Young Hwang[1], In-Ho Jang[2], Seung-Jun Cha[2], Kyu-Chul Lee[2,*]

[1] BK21 Next Generation Information Technology Software, 99 Daehak-ro, Yuseong-gu,
Daejeon, Korea
[2] Dept. Computer Engineering, 99 Daehak-ro, Yuseong-gu, Daejeon, Korea
{yyhwang, rolroy, junii, kclee}@cnu.ac.kr

Abstract. All government departments in South Korea used the Facebook Page to inform their own policies and to collect the opinion of the people. The record from the perspective of each government department writings posted on a Facebook page, containing the opinions of people comment on posts, etc. have significant value. In this paper, we proposed long-term preservation method for Facebook page of government departments. In addition, we proposed the long-term preservation format, document preservation format, long-term preservation metadata and document preservation metadata.

Keywords: Web Content Archiving, Facebook, Archiving, Preservation

1   Introduction

In recent years, an increase of smartphone users due to the expansion of wireless Internet services and SNS (Social Network Service) has spread rapidly. SNS has become an effective online solution to create and reinforce social relationships through free communication and information sharing between users. The SNS proliferation brought about change in public service of government departments and public agencies. All government departments in South Korea used the Facebook Page to inform their own policies and to collect the opinion of the people. Record from the perspective of each government department writings posted on a Facebook page, containing the opinions of people comment on posts, etc. have significant value.

The interested in web archiving has spread to the library several countries since National Library of Canada had started EPPP(Electronic Publications Pilot Project) in 1994. Brewster Kahle developed WAIS established the Internet Archive in April 1996. It was a great contribution to promoting the project in each country. The purpose of internet archive is to collect and preserve of public data, and to establish a digital library to support archiving data

Facebook is one of the web pages, but we cannot use the web archiving method for Facebook record archiving. Facebook uses many AJAX requests to navigate the website which is problematic for any automated robot. For example, the "Show More Posts..." link which allows access to older Wall posts issues an asynchronous POST

---

* Corresponding Author

request which returns HTML composing the next listing of friends. A robot cannot simply mine the HTML and JavaScript of the Wall page for a link to produce this content.

There are Facebook archiving related work: NASA Archive-It[1], FacebookArchive[2], ArchiveFacebook[3], and SocialSafe[4]. NASA Archive-It backups aviation agency's social media activity using web archiving. FacebookArchive a formal offer from Facebook in the form of a web page, the archive data is stored on your local disk. ArchiveFacebook provided by Firefox is the archiving tool. It records the Facebook data to local hard disk. SocialSafe based API supports the backup of Facebook records.

They only support the backup of Facebook profile for private user, not a Facebook pages which is used by government agencies.

In this paper, we proposed long-term preservation method for Facebook page of government departments. In addition, we proposed the long-term preservation format, document preservation format, long-term preservation metadata and document preservation metadata.

This paper has the following configuration. Chapter 2 analyzes the Facebook page and Facebook page to be archival data should be selected. Chapter 3 ministries Facebook page explains how to collect data. Finally, Chapter 4 Conclusions and future research will tell.

2   What and How to Archive

There are three types of Facebook, Facebook Profile for private user, Facebook Page for public agencies or government departments, and Facebook Group for small groups. Republic of Korea all 17 government departments are now using Facebook page.

2.1   Facebook Page Data

Facebook page with the following data is stored.

① Post which are produced by the Facebook page as an administrator, and can also be simply writing, pictures or video may be written with the article. The Post can be delivered to all fans of the page. In addition, Facebook friends, including members can comment on each post, post the preference of the "Like" can be expressed by choosing.

② Profile of the page contains the information of the organizations or groups.

③ Application means all apps which are allowed by Facebook. We can use the "Tab" application in Facebook page. This application connects the external web pages by the link.

④ Insights mean the statistical information about Facebook page. It includes the number of fans, visits and posts.

⑤   List of Fans is the array of fans watching the Facebook page. Facebook page of friendship rather than a personal profile of the user "Likes" button to establish ties.
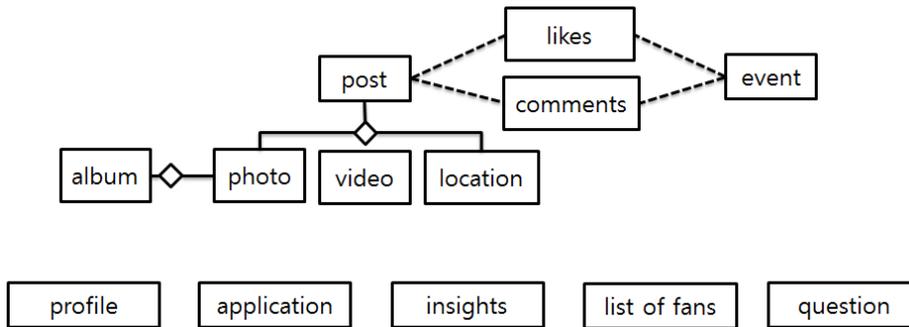
⑥   Question is what about the question and answer.

Figure 1. Facebook Page Data

Facebook manage all data of pages in their databases. Facebook pages were also able to apply the timeline. It can be able we show the posts which are classified by time. Basically, Facebook postings show as much as the length of the schedule, posting a full article to see the "Show More Posts .." you need to click.

2.2   What and How to Archive

Facebook archiving methods are as follows[5]:

①   Enable email notifications from Facebook to archive textual data in an email account.

②   Manually copy and paste screenshots or the text and images from all Facebook screens into a word processor.

③   Use the web browser or browser extension to save complete snapshots of each Facebook screen to disk.

④   Use the Facebook API to mine user data.

⑤   Automate saving snapshots of Facebook data using a web crawler or browser extension.

The case of the personal profile of Facebook, we can receive the email message include the post writing on the Facebook. However, Facebook page cannot support this function, so, option 1 is not good solution to Facebook page. Option 2 is the method that user direct copy the Facebook data for archiving. It is not efficient and not supporting long-term preservation. Option 3 is using the "Save as.." function of web browser. It also has the problem that we cannot support long-term preservation of Facebook page. FQL(Facebook Query Language)[6] and Graph API are freely

available API for writing Facebook applications. The API could be used to archive Facebook Page's content(Option 4). The final option would allow some content to be archived. However, existing web crawler or browser extension which is not able to fully support archiving a Facebook Page's content. The web crawler is not able to collect any content such as "Show More Post.." which needs the user action. Therefore, we collect the content of the Facebook page by using Facebook API. Figure 1 show the target data which has to be collected. We use FQL API for collecting the data, and a listing of Facebook data types and API support for accessing the data is provided in Table 1.

Table 1. FQL API supports for data extracting

Data Type	API Functions
Post	location_post, photo, photo_src, stream, video, video_tag
Comments	comment, comments_info
Likes	like
Event	event, event_member
Question	question, question_option, question_option_votes
Profile	page
Insights	insights
List of Fans	Page_fan

Indicated in Figure 1 Application of Facebook Page's data is not collected in the FQL. This Application is used by Facebook Page Tab sometimes called, shows them connecting an external webpage. Thus, using the existing web archiving is better to archive.

3 Facebook Page Archiving Framework

We have developed an archiving framework for capturing the Facebook page data and creating the long-term preservation format.
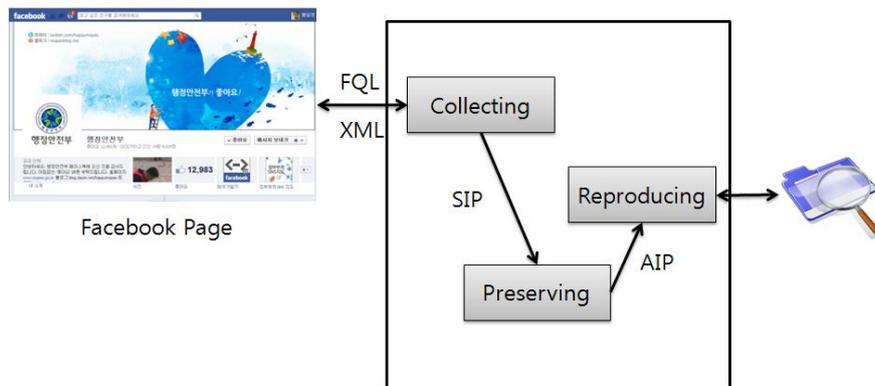

Figure 2. Facebook Page Archiving Framework

We are using FQL API to collect Facebook page data. FQL API query results are returned as an XML file format, the results generated to create the SIP and generate the SIP (Collecting), and then, generate long-term preservation format (Preserving). Figure 3 shows the XML file about Post and Photo through FQL query. Like with Comments on each post a separate query can be obtained through FQL. Therefore, we analyze Post.xml, Comments.xml and Like.xml related to each Post, and generate the one XML file including all of them.

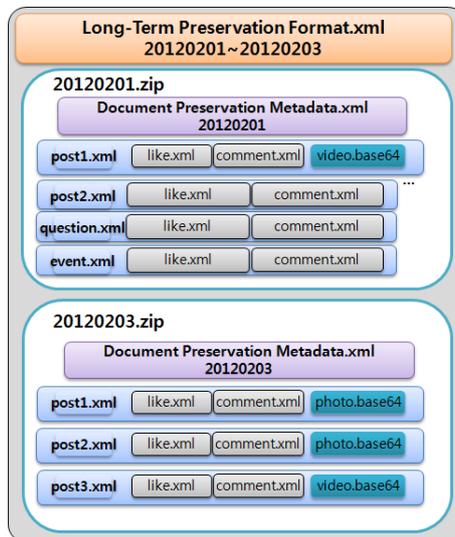**Figure 3.** Collecting Post of Facebook Page

Figure 4. Long-Term Preservation Format for Facebook Page

Long-term preservation format includes the preservation metadata for archiving Facebook page and contents which are represented in SIP. We designed the incremental archiving method by every month. This format consists of the document preservation format classified by daily(Figure 4). We represented the SIP and AIP by XML file format. It is very efficient to present the DIP to users. The users are able to search the archiving data using date or organization.

As shown in Figure 4, we generate an XML file to the AIP, a picture or video is stored on the wire base64 encoded. In units of one month proposed that provides statistical information which includes list of fans, and number of posting and comments. This information is represented in Long-Term Preservation Metadata. The long-term preservation metadata is represented creator, created time, statistical information, structure of contents, and relationships between actors and contents. We also designed the document preservation metadata which includes information about creator of post, create time, last update time, number of comments to post, and number of like to post or comment.

5 Conclusion

In this paper, we proposed the incremental archiving method for Facebook page. When you want to reproduce the Facebook page in special period, we can show the all posts in period. However, after the end of archiving for Post, someone writes a comment at the post, we cannot find the relationship between them. This problem is our future work.

Acknowledge

References

1. NASA,: Archiving NASA: http://words.nasaimages.org/2010/03/05/arvhiving-nasa/ (2010)
2. FacebookArchive: http://www.fbarchive.com/
3. Mat., K., Carlton, N., Hany, S., Michael N., Frank, M.: ArchiveFacebook,: https://addons.mozila.org/ko/firefox/addon/archiveFacebook
4. SocialSafe,: http://www.socialsafe.net/
5. Frank, M., Michael, L., N.,: What Happens When Facebook is Gone?: JCDL'09, June 15–19, Austin, Texas, USA(2009)
6. Facebook Developers: FQL, http://developers.Facebook.com/docs/reference/fql/
7. State of Florida,: ELECTRONIC RECORDS AND RECORDS MANAGEMENT PRACTICES,: (2010)