Devloping a Collection Policy
for a National Archive of Free Israeli Internet Sites

Rivka Shveiky

Abstract

The vast amount of information that exists on the Internet is unstable, frequently disappearing or changing. In order to be rely on free Internet content as a source of knowledge, it must be stored and preserved in the same way that libraries and archives handle other informational formats.

In the 1990's, national libraries and archives turned their attention to the issue of archiving the national body of creative work that was published on the Internet. One of the initial questions considered, and continues to be considered, by web archiving entities is: what should be archived from among the vast amount of information found on the Internet? This study attempts to answer that question by developing a methodical collection policy for a national archive of free Israeli websites and examining the practicability of the policy through a trial collection of Israeli websites.

The study was based on a qualitative approach. The first stage, developing the collection policy, was carried out using the grounded theory method. The second stage, trial implementation, was carried out using the participatory action research method.

The study opened with an examination of the characteristics of national collections and the ways in which peoples and countries document themselves through archival and library collections. The source of data for this part of the study was a sample of official documents from various national collections. Documents were examined regarding the collection policies of five large Israeli collections containing material in different formats, as were the policy documents of 17 national libraries and 25 national web archives from around the world. These documents were analyzed by coding and

---

categorizing their content in order to identify general national collection policy traditions and criteria.

Afterward, the study examined the expectations of various interested parties in regard to a future national archive of Israeli websites. The source of data for this part of the study was in-depth interviews with a sampling of relevant informants. Interviews were conducted with 30 interviewees: 17 potential users of a national website archive, eight information science experts and five position holders whose positions may grant them influence over legislation, budgeting and implementation of a national website archive. The interviewees were asked to outline collection policy for a future archive of Israeli websites. These interviews were also analyzed by coding and categorizing their content in order to identify opinions, needs and expectations regarding the collection policy that is desirable for an archive of Israeli websites.

A collection policy proposal was drafted based on the findings from content analysis of the documents and the interviews, integrating common characteristics and customary traditions of national archives with the expectations and needs of future interested parties. The second stage of the study examined the practicability of the proposed policy. This examination was carried out at the National Library of Israel, by a group of librarians from the library team. The librarians got a list of websites dealing with the elections for the 18th Knesset, that were identified by students through commonly used Internet search engines. The librarians were asked to utilize the proposed policy as an evaluation tool for websites for a trial collection of selected websites dealing with the elections for the 18th Knesset. The librarians made their selections from a list of website which, according to the criteria listed in the proposed policy were suitable for perpetual preservation as part of a national collection. The websites they selected were captured a number of times during the election period – from the government's announcement that elections would be held through the day after the election results were published. These archived websites will be part of the National Library's collections and will be accessible to the public, subject to copyright laws. During the process, the librarians established a discussion group to record the

הספריה ללימודי מידע   Library of Information Science
אוניברסיטת בר-אילן, רמת גן, ישראל   52900   Bar-Ilan University, Ramat-Gan, Israel
טל. 972-3-5318163 .Tel   Email: Ruthi.Tshop@mail.biu.ac.il   Websitewww.is.biu.ac.il/library

ways in which they utilized the proposed policy, as well as their comments and reservations regarding its efficacy as a tool to guide curators in selecting internet resources. These comments and reservations were collected and analyzed and at the end of the process of building the trial collection the entire group gathered to make changes and revisions to the proposed policy. The final product is a policy document suitable for guiding curators of a national collection as far as which internet resources should be archived.

Websites change over time – content is added while other content disappears. Thus, capturing a website only once at any given point in time will not suffice. In order to preserve all of the information appearing on a website it must be captured again periodically. Website curators debate the questions of the desired frequency of capture. If capturing is conducted too infrequently, content that appeared and disappeared between captures might be lost, while overly frequent capturing wastes resources and loads the user with unnecessary information. The final stage of the study tested the appropriate capture frequency for Israeli political websites, based on two collections constructed by the National Library in the past: a collection of websites dealing with the Israeli-Arab conflict that was collected in 2003-2004 and a collection of websites dealing with the elections for the 17th Knesset that was collected in 2006. The websites of the two collections were observed, in order to checked how many of them had completely disappeared from the Internet during the time between the last capture and the check date, in October 2008, how many had changed and what changes had occurred. Based on the findings, recommendations were offered for determining the capture frequency for future collections of Israeli political websites.

הספריה ללימודי מידע   Library of Information Science
אוניברסיטת בר-אילן, רמת גן, ישראל   52900   Bar-Ilan University, Ramat-Gan, Israel
Websitewww.is.biu.ac.il/library Email: Ruthi.Tshop@mail.biu.ac.il Tel. 972-3-5318163 .טל