WEB ARCHIVING AT THE NATIONAL SZÉCHÉNYI LIBRARY

The Hungarian National Széchényi Library – as national libraries in many countries – has a responsibility for long term preservation of publicly available internet resources, especially websites. By the development framework of the National Library Platform a pilot project has started in spring 2017 in order to establish the professional conditions of a permanent Hungarian web archive.

WHY?

- the internet has a lively present, but its past needs to be demonstrated also
- the history of the virtual world must be researchable, the online imprints of the current events in the real world must be preserved
- the large volume of digital information must be analysable and visualizable
- there is a need for stable references to online sources in education materials and scientific publications
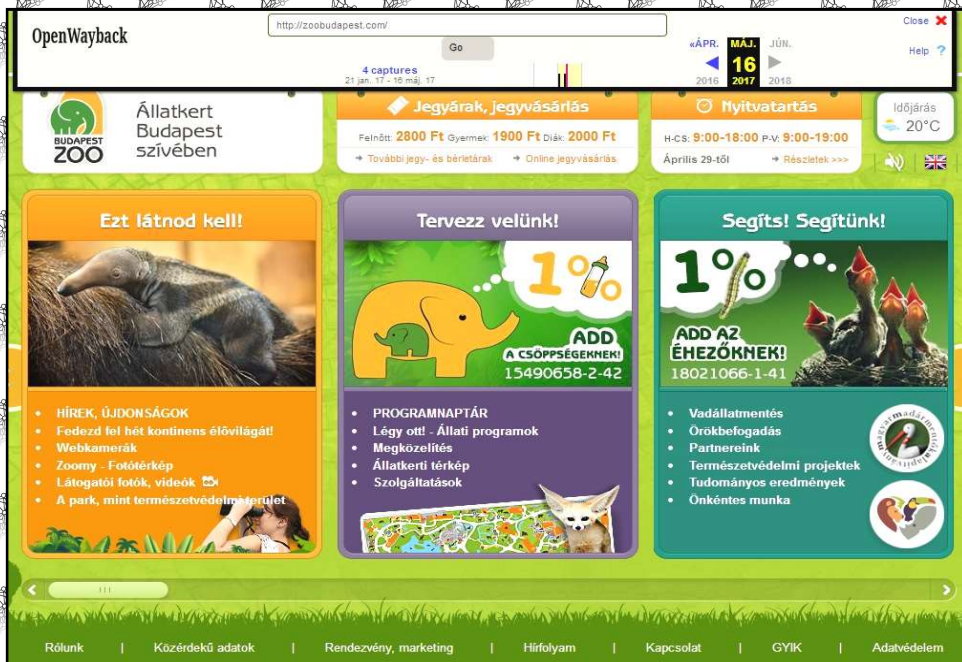
WHAT?

- mainly Hungarian websites focusing on culture, science, education or public events
- Hungarian electronic journals, periodicals, magazines, news portals and blogs
- relevant contents related to Hungary on foreign servers

HOW?

- thematic collections by topic, genre or institution have crawled periodically
- in some cases copyright agreements are requested form the content owners to publish the archived copies publicly
- special event-related temporal harvests
- representative crawls of the .hu domain once or twice a year

Some information about the project (based on data from October 2019):

- IT infrastructure provided by Government Agency for IT Development and by National Széchényi Library
- software background: Web Curator Tool, Heritrix, WAIL, Webrecorder, HTTrack, OpenWayback, PyWb, SolrWayback...
- number of selected websites ca. 20,000, total size of the archive: 24 TB
- the public collection contains archived copies of 186 websites, blogs or periodicals, and 44 online content services from the national library

Homepage of the Budapest Zoo archived in 16 May 2017 and displayed in Open Wayback

Useful links related to web archiving:

- Temporary project website: mekosztaly.oszk.hu/mia
- Internet Archive: web.archive.org (the biggest web archive of the world: 387 billion archived web pages since 1996)
- International Internet Preservation Consortium: netpreserve.org (founded in 2003, members from around 45 countries)
- archive.is (free on-demand web archiving service)
- HTTrack: httrack.com (free web archiving software for Windows)