

WEBARCHIVÁLÁS AZ OSZK-BAN

Az Országos Széchényi Könyvtár – sok más ország nemzeti könyvtárához hasonlóan – feladatának tekinti a nyilvános internetes források, ezen belül is elsősorban a webhelyek hosszú távú megőrzését. Ezért az Országos Könyvtári Rendszer kifejlesztésének keretében 2017 tavaszán elindult egy projekt azzal a céllal, hogy megteremtse egy nemzeti szintű magyar webarchívum feltételeit.

MIÉRT?

- hogy legyen múltja is a magyar internetnek, ne csak jelene
- hogy kutathassuk a virtuális világ történetét, valamint a valódi világ elmúlt eseményeinek internetes lenyomatait
- hogy elemezni és ábrázolni lehessen a nagy tömegű digitális információt
- hogy megbízhatóan tudjunk hivatkozni tudományos publikációkban és tananyagokban online forrásokra

MIT?

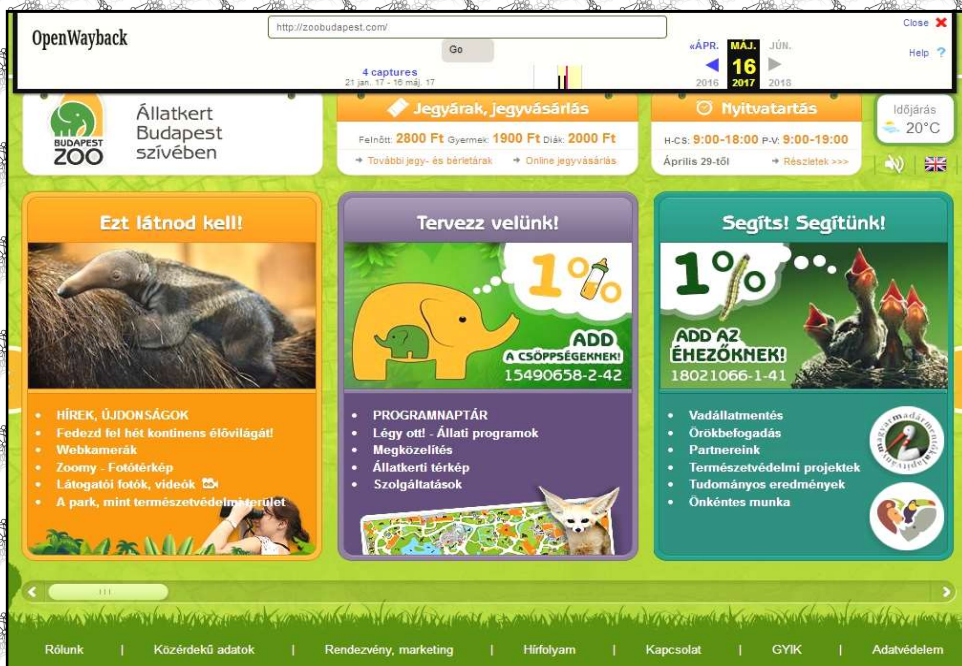
- főként a hazai kulturális, tudományos, oktatási és közéleti témájú honlapokat
- a magyarországi elektronikus újságokat, folyóiratokat, magazinokat, hírportálokat és blogokat
- a külföldi szervereken levő, de magyar vonatkozású és megőrzésre érdemes webhelyeket

HOGYAN?

- témák, műfajok vagy intézmények szerint összeválogatott webhelyeket rendszeresen letöltve
- egyes esetekben engedélyt kérve az archív példányok szolgáltatásához a tartalomgazdától
- kiemelten fontos eseményekről begyűjtve egy ideig a híreket
- évente egy-két alkalommal egy reprezentatív részt aratva a .hu doménról

Néhány információ a projektről (2019 októberi adatok):

- az informatikai háttérrel a KIFÜ és az OSZK biztosítja
- a használt szoftverek: Web Curator Tool, Heritrix, WAIL, Webrecorder, HTTrack, OpenWayback, PyWb, SolrWayback...
- a válogatott webhelyek száma kb. 20 ezer, az archívum összmérete 24 TB
- a nyilvános gyűjteményben 186 honlap, blog és időszaki kiadvány található, továbbá 44-féle OSZK-s online szolgáltatás mentése



A Budapesti Állatkert 2017. május 16-án archivált honlapja az OpenWayback megjelenítőben

Hasznos címek a webhelyek megőrzésével kapcsolatban:

- A projekt ideiglenes weboldala: mekosztaly.oszk.hu/mia
- Internet Archive: web.archive.org (a világ legnagyobb webarchívuma: 387 milliárd lementett weboldal 1996 óta)
- International Internet Preservation Consortium: netpreserve.org (2003-ban alapított nemzetközi szervezet, melynek kb. 45 országból vannak tagjai)
- archive.is (ingyenes, igény szerinti weboldal archiváló szolgáltatás)
- HTTrack: httrack.com (ingyenes webhelyletöltő program Windows-ra, magyarított felülettel)