Példák, szoftverek

Az alábbi példákban látható, hogy különféle könyvtárak milyen módon használták fel a crowdsourcing és a civil sciences módszereket állományuk leírásának fejlesztésére.

Magyar példák

Nemzetközi példák

Szoftverek

Magyarországi példák

- **Cloudsourcing blog**
  A Cloudsourcing blog [1] egy magyar blog, mely a crowdsourcing és valamelyest a citizen science fogalmakat igyekszik körüljárni, bemutatni. Néhány példát is találunk a blogbejegyzések között. A blog nagy hátránya és hibája, hogy az utolsó bejegyzések 2011-re datálódnak, láthatóan régen nem frissítették. Azonban az akkor összeszedett, a fogalmakat körüljáró írások miatt fontosnak tartottuk feltüntetni a példák között.
- **Nyugat-magyarországi Egyetem Savaria Egyetemi Központ Könyvtára**
  A Nyugat-magyarországi Egyetem Savaria Egyetemi Központ Könyvtárában [2]elindult egy kezdeményezés az intézménytörténeti képtár feldolgozása során. A Google képtárába [3] feltöltött képekhez a nagyközönség "Megjegyzések" hozzáfűzésével tud plusz információkat adni, főként a képen szereplő személyekről. Jobb oldalon láthatók a *Legutóbbi megjegyzések*, ahol főként a könyvtár korábbi munkatársai segítik a régi képek mind részletesebb leírását.
- **Hagyomány és múltidéző**
  A történelmi tudástárat [4], amely a közösségi tudásmegosztáson és tudásépítésen alapszik, néhány lelkes önkéntes kezdte el működtetni. Az oldalt azzal a céllal hozták létre, hogy történelmi hagyományőrző körökön belül segítsék a leleteken alapuló, hiteles történelmi rekonstrukciók készítését, terjedését, terjesztését. A segítség egyrészről a kezdők felé irányul, hogy jobban megismerjék az általuk választott korszakot, és hogy a felszerelésüket (viselet, eszközök, fegyverek) a megfelelő irányba tudják fejleszteni. Másrészről szeretnének segítséget nyújtani a témával régebb óta foglalkozó embereknek is, hogy tökéletesíthessék, kiegészíthessék a meglévő felszerelésüket, egyben további rekonstrukcióra bíztatják őket. Céljukat úgy szeretnék elérni, hogy az internetről, a könyvtárakból származó információkat, illetve a leletek rekonstruálásából eredő tapasztalataikat összegyűjtik, és rendszerezve azokat megosztják az érdeklődőkkel. A megvalósításban bárki részt vehet, a beküldött anyagokat a beküldő nevén teszik közzé. Az oldal építése és üzemeltetése során ingyenes technikai megoldásokat használnak fel.
- **Arcanum Adatbázis Kft.** – Képeslapok [5] (Postcards) szolgáltatása
  Jelenleg öt intézmény, a Zempléni Múzeum, az Országos Széchényi Könyvtár, a Magyar Kereskedelmi és Vendéglátóipari Múzeum, a Magyar Mezőgazdasági Múzeum és a Damjanich János Múzeum közel 300 000 db képeslapja kereshető és tekinthető meg az oldalon. A képeslapokat a szolgáltató térképen is elhelyezte, azonban sok esetben csupán a város van

beazonosítva, a képen ábrázolt épület, tér vagy utca nem. Regisztráció és bejelentkezés után a civileknek, az érdeklődőknek is van lehetőségüket önkéntes alapon javítani a képeslapokon ábrázolt közterület nevét, illetve pontosítani a megadott adatokat.

- **Felolvasó blog**: Az oldalra [6] olyan önkéntes felolvasásokat gyűjtenek, amelyek klasszikusok és kortársak novelláit, meséit, verseit vagy egyéb más rövidebb műveit szólaltatják meg magyar nyelven. Amennyiben a felolvasott mű kortárs szerzőtől való, akkor a szerző és/vagy fordító nyilatkozatának megléte szükséges ahhoz, hogy az oldalra felkerüljön a hanganyag. Erről minden esetben a felolvasónak magának kell gondoskodnia. Az oldal fizikailag nem tartalmaz hanganyagokat, gyűjtőoldal, amely beágyazva jeleníti meg és teszi lehetővé a hangzóanyagok meghallgatását. Ehhez a SoundCloud [7] hangmegosztó oldalát, vagy az Internet Archive [8] oldalát ajánlják. A hanganyagokat a felolvasóknak maguknak kell feltölteniük a hangmegosztó oldalra, az oldal szerkesztői csak a posztoláshoz szükséges linket, és a felolvasás szövegének pár ízelítő sorát kérik. *Forrás* [9]
- **Fortepan**
  A Fortepan [10] egy online fotóarchívum jelenleg kb. 44.000 képpel, ahol minden képre a Creative Commons CC-BY-SA-3.0 [11] licenc érvényes, s így azok szabadon felhasználhatók. A Fortepan önkéntes szerkesztők és segítők munkájára épül. Támogatásként csak magánadományokat fogadnak el.
  Az önkéntesek bővíthetik a képgyűjteményt, illetve a saját maguk és mások által feltöltött fotókhoz kulcsszavakat írhatnak. A fórumban [12] lehetőség van egy-egy kép részletes leírására is. Hasonló céllal készült a Fortepan Wiki [13], amelyben lehetőség van a képek megfejtésére, majd az ahhoz kapcsolódó fontosabb adatok megosztására, mint pl. a kép készítésének helye, ideje, részletes leírására, térképen való elhelyezése. A Fortepan tökéletes példája a crowdsourcing [14]-nak, amikor a közösség/közönség épít, dolgoz fel és gazdagít adatokkal egy dokumentumgyűjteményt. Forrás *[15]*
- **Világháborús emlékek felkutatása –** Siófok, BRTK Könyvtár **[16]**
  Siófok és környékének I. és II. világháborús katonáiról, hősi halottairól, katonai áldozatairól fennmaradt emlékek, fényképek után indított kutatást a BRTK Könyvtár. A gyűjtőmunka – dr. Kovács Emőke támogatásával és a BRTK Könyvtár koordinálásával – Pusztai Fruzsina könyvtáros kezdeményezésére indult.
  100 éve éppen javában zajlott az I. világháború, 70. éve pedig már véget is ért a II. világháború. Ehhez kötődik a BRTK Könyvtár felhívása, amelynek során várják a helyi és környékbeli lakosok világháborús katonákról, áldozatokról szóló emlékeit, történeteit, fényképeket, naplókat, leveleket. Dr. Oláh Vilmos és Kenedy Ferenc foglalkozott a korábbiakban ezzel az embert próbáló feladattal, Siófok és környékének I. és II. világháborús katonáival, hősi halottaival, katonai és polgári áldozataival, felkutatták, amit akkoriban csak lehetett, és kitűnő munkát végeztek, de a digitalizált világ, az internet új lehetőségeket rejt magában és felgyorsítja a keresést. Pusztai Fruzsina és a könyvtár egyeztetni, kiegészíteni, pontosítani szeretné az ő listáikat, és szeretnének arcokat társítani a nevek mellé. Forrás: Siófokinapilap.hu [17]

Nemzetközi példák

- **Metadata Games**: Az oldalon [18] található, okostelefonokra letölthető alkalmazások segítségével játékos módon segíthetik az önkéntesek a könyvtárak, múzeumok és egyetemek feldolgozó munkáit világszerte. A projekt lényege, hogy az önkéntesek „tag”-ekkel, magyarul címkékkel, könyvtáros szaknyelven tárgyszavakkal látnak el különböző képeket, videókat, hangfájlokat. A projekt szervezői szerint e segítség nélkül némelyik dokumentum akár örökre is eltűnhetne, feldolgozatlanul maradna. Az alap, kiinduló mobil alkalmazás forráskódja ingyenesen elérhető a projekt honlapján keresztül, az egész kezdeményezés „open source”, azaz nyílt hozzáférésű.
- **New York Public Library –** Whats on menu? [19] (Étlapok feltárása)
  A projekt célja feldolgozni, digitálisan rögzíteni New York történelmi éttermeinek régi étlapjait a múltból, ehhez kérik a civilek segítségét. Ezáltal próbálják meg azt is feltárni, hogy az előző századokban milyen ételeket fogyasztottak New York City lakói és vendégei. Az oldalon egy adott ételre illetve étteremre is lehet keresni. A Geotagger alkalmazás segítségével az önkéntesek egy térképen is elhelyezhetik az adott étlapokat, menüket, ezáltal segítve a földrajzi szempontból történő visszakeresést. A projektben eddig több, mint 1.300.000 étel

került leírásra és feldolgozásra valamivel több, mint 17.500 étlap alapján.

- **New York Public Library –** Building Inspector [20] (Régi épületek és térképek azonosítása)
  A projekt célja New York City történelmi térképein, földmérők által készített helyszínrajzain azonosítani régi épületeket, azok elhezkedését, címeket, területeket az állampolgárok segítségével, bevonásával. A weboldalon bemutató videók segítik az állampolgárok munkáját a helyes feldolgozás érdekében.
- **University of California –** California Digital Newspaper Collection [21] (Digitalizált újságok szövegeinek javítása)
  A digitalizálási munkák 2005-ben kezdődtek. A CDNC honlapja 2007 októbere óta működik, 2009-től kezdve a Veridian [22] szoftverrel. 2011 augusztusában vált lehetővé, hogy a regisztrált felhasználók módosíthatják az archívumban található szövegeket (user text correction). A néhány hónap alatt regisztrált több mint 1300 főből mintegy hatszázan több mint egymillió szöveges sort javítottak ki. A honlapon kezdettől fogva megtalálható a Google Analytics, amivel a honlapon tartózkodás átlagos idejét lehet mérni. Forrás *[23]*
- **National Library of Australia** – Trove [24] (Digitalizált újságok szövegeinek javítása)
  A könyvtári és levéltári szférában a civil archiválást, az állampolgárok bevonását elsőként 2009-ben, az Ausztrál Nemzeti Könyvtárban alkalmazták a Trove a nemzeti történelmi dokumentumok interaktív portáljának építéséhez. Egy speciális felületen segítségével bárki írhat szöveget, tehet fel fényképeket, képeket az oldalra. A civileknek 8 millió oldal leírásának javítására van lehetőségük. A Trove projektben az önkéntesek munkájának értéke meghaladta a 2000 USA dollárt. Forrás *[23]*
- **Cambridge Public Library** – Historic Cambridge Newspaper Collection [25] (Digitalizált újságok szövegeinek javítása)
  A Cambridgei-i Városi Könyvtárban a Trove porjekthez módszert követték, 2013-ban ünnepelték a történelmi hírlapgyűjtemény (Historic Cambridge Newspaper Collection [26]) egyéves fennállását. A DL Consulting céggel és új-zélandi szakemberekkel együttműködve a könyvtár digitalizálta és nyílt hozzáférésűvé tette mindazokat a hírlapokat, amelyeket a szerzői jog nem érint, köztük az USA legrégebbi hetilapját, a Cambridge Chronicle-t. Az eltelt év alatt a felhasználók 40 ezer olvashatatlan sort korrigáltak, a teljesítmény heti ezer sorra nőtt. A feldolgozás ebben a projektben is a Veridian szoftverrel történt/történik. Forrás *[23]*
- **University of Oxford** – Ancient lives [27] (Kéziratok, papiruszok szövegének átírása)
  A brit Ancient Lives (ókori életek) projekt is kihasználja azt, hogy sokan hajlandóak kéziratos dokumentum átírásával vesződni – a program gazdája az Oxfordi Egyetem –, és ezt a tevékenységformát a citizen science fogalmával jellemzik. Ennek keretében egyetemi tanárok és civilek (citizen scientists) foglalkoznak olyan papiruszok szövegének átírásával, melyek az Egypt Exploration Society és más intézmények birtokában vannak. (A papiruszlapok képeit feltöltik az internetre, és bárki megfejtheti a hieroglifákat.)
  2011–2012-ben másfél millió leírás készült így, amelyekből több mint százat sikerült azonosítani (többek között Plutarkhosz és Euripidész alkotásait). A közös metaadatok nem rosszabbak, mint amit a szakemberek készítenek, sőt, a kettő kiegészíti egymást. Forrás 1 *[28]*, forrás 2 *[28]*
- **British Library** – UK Soundmap [29] (Közösen épített hanggyűjtemény)
  Az Egyesült Királyság hangtérképe. A projekt során 350 önkéntes mindenféle hangzó anyagot küldött be a BL Sound and Vision részlegébe 2010 júliusa és 2011 júliusa között. Az egyéves periódus alatt mindenféle hangminta bekerült a BL gyűjteményébe: madárhangok, a szél hangja egy adott helyen, gőzmozdony zaja, az utca zsongása egy adott városban, stb. A feltöltött hangokat - a projekt nevéhez híven - a Google Maps-et alkalmazva az Egyesült Királyság térképén is elhelyezték, így földrajzilag is lehet válogatni a különböző minták között. *Forrás* [30]
- **National Library of Finnland** – DigiTalkoot [31] (Digitalizált újságok szövegeinek javítása)
  A Finn Nemzeti Könyvtár [32] és a Microtask [33]által működtetett projekt, melynek célja az szoftverrel felismertetett szövegek kijavítása önkéntesek, civil résztvevők által. A projektben 110 ezren több mint 8 millió szót javítottak ki. Az efféle munka időigényes, és anyagi ráfordítással is jár, ezért körültekintően kell megtervezni, kinek, mit, hogyan lehet felkínálni, illetve ki és hogyan ellenőrzi az elvégzett feladatot. A siker érdekében felosztják a munkát, továbbképzik a résztvevőket, rendszeresen ellenőrzik kompetenciáikat, gyakorlati képzést nyújtanak, és konkrét példákat elemeznek (pl. egy munka többféle megoldását). *Forrás* [28] [28]
- **Zooniverse** – Old Weather [34] (Meteorológiai dokumentumok javítása)

A projekt részvevőit arra kérték, hogy a XIX. század közepétől a XX. század elejéig írják le az angol tengerészet kéziratos meteorológiai feljegyzéseit, hogy szakadatbázisukat kiegészíthessék. Az így létrejövő eredmények segítenek a tudósoknak feltérképezni az akkori klímamodellt és környezeti feltételeket, kondíciókat. Ezeket az adatokat később történészek is felhasználják a hajó- és flottamozgások kutatása során.

Szoftverek

- **Veridian szoftver**
  A szoftver [22] kimondottan könyvtárak és kulturális örökség gondozásáért felelős intézmények számára készült, főként a nem jogvédett, régi hírlapok digitalizálására fókuszál. A szoftver és a szolgáltatás keretein belül a kulturális intézmények számára lényegében minden munkafolyam elérhető, így a digitalizáció megtervezése, az anyag szkennelése, a szöveg automatizált felismertetése (OCR), metaadatok elkészítése METS/ALTO XML-ben, a digitalizált anyag szolgáltatása és hosszútávú megőrzése. A szoftver Civic Epistemologies-projekt szempontjából legfontosabb összetevője a „User Text Correction”, azaz a felhasználók, önkéntesek, civilek által végzett (végezhető) szövegkorrektúra, az OCR-ezett szövegben a szoftver automatizmusa miatt előforduló hibák kijavítása. Számos, a nemzetközi példák között olvasható projekt erre az önkéntes segítségre alapoz, sok helyen több millió szót önkéntesek javítottak ki a hírlapok tömör soraiban a Veridian szoftver segítségével. *Forrás* [35]
- **PyBossa**
  A PyBossa Project [36] egy nyílt forráskódú és hozzáférésű keretrendszer, mely weblapokba beépíthető alkalmazásokat, „toolkit”-okat ajánl és tesz elérhetővé a felhasználók számára. Használata teljesen ingyenes, a honlapjukon az egyes szakterültekhez készült alkalmazásoknál közzéteszik azok forráskódját, amelynek köszönhetően egy hozzáértő webszerkesztő könnyedén beépítheti az adott projekt honlapjába a feldolgozáshoz szükséges alkalmazást. A PyBossa keretében hat feldolgozási területhez szükséges alkalmazás érhető el: képfelismerés, hangminták felismerése, PDF dokumentumok átírása, földrajzi elhelyezés, videóminták felismerése, okostelefonok által begyűjtött adatok rendszerezése. Az alkalmazásokat már több projekt is használja (MicroPasts [37] – régészettel kapcsolatos; ForestWatchers [38] – erdőkkel kapcsolatos monitorozás; Feynman’s Flowers [39] – fizikai, mágneses molekulákkal kapcsolatos projekt).

2015/01/07 - 11:20

Forrás webcím: https://oszk.hu/civic/peldak

Hivatkozások:
[1] http://cloudsourcing.blog.hu/
[2] http://sek.nyme.hu/konyvtar/
[3] http://picasaweb.google.com/sekkonyvtar
[4] http://sites.google.com/site/hagyomanyesmultidezo/
[5] http://postcards.arcanum.hu/hu/
[6] https://felolvaso.wordpress.com/
[7] http://soundcloud.com/
[8] http://archive.org/
[9] https://felolvaso.wordpress.com/mi-ez/
[10] http://www.fortepan.hu/
[11] https://creativecommons.org/licenses/by-sa/3.0/
[12] http://forum.index.hu/Article/showArticle?t=9204653
[13] http://fortepan.wikia.com/wiki/Fortepan_Wiki
[14] http://www.oszk.hu/hirek/civic
[15] http://www.fortepan.hu
[16] http://www.konyvtar-siofok.hu/?p=hirek&id=140
[17] http://www.siofokinapilap.hu/cikk/Vilaghaborus_emlekek_utan_kutat_a_siofoki_konyvtar/184880/
[18] http://www.metadatagames.org/

[19] http://menus.nypl.org/
[20] http://buildinginspector.nypl.org/
[21] http://cdnc.ucr.edu/cgi-bin/cdnc
[22] http://www.veridiansoftware.com/
[23] http://humanus.bibl.u-szeged.hu/human/cikk-mokka?marcid=human/cikk/491340
[24] http://help.nla.gov.au/trove/using-trove/digitised-newspapers/text-correction
[25] http://cambridge.dlconsulting.com/cgi-bin/cambridge?a=p&p=home&e=-------en-20--1--txt-txIN------
[26] http://cambridge.dlconsulting.com
[27] http://www.ancientlives.org/
[28] http://humanus.bibl.u-szeged.hu/human/cikk-mokka?marcid=human/cikk/475345
[29] http://sounds.bl.uk/sound-maps/
[30] http://humanus.bibl.u-szeged.hu/human/cikk-mokka?marcid=human/cikk/486098
[31] http://www.digitalkoot.fi/
[32] http://www.kansalliskirjasto.fi/
[33] http://www.microtask.com/
[34] http://www.oldweather.org/
[35] http://www.veridiansoftware.com/veridian-newspaper-digitization-services/
[36] http://pybossa.com/
[37] http://crowdsourced.micropasts.org/
[38] http://forestwatchers.net/
[39] http://crowdcrafting.org/app/feynmanflowers/