

A Matarka szerszámosládája

Szeged, 2007

Perlaki Attila

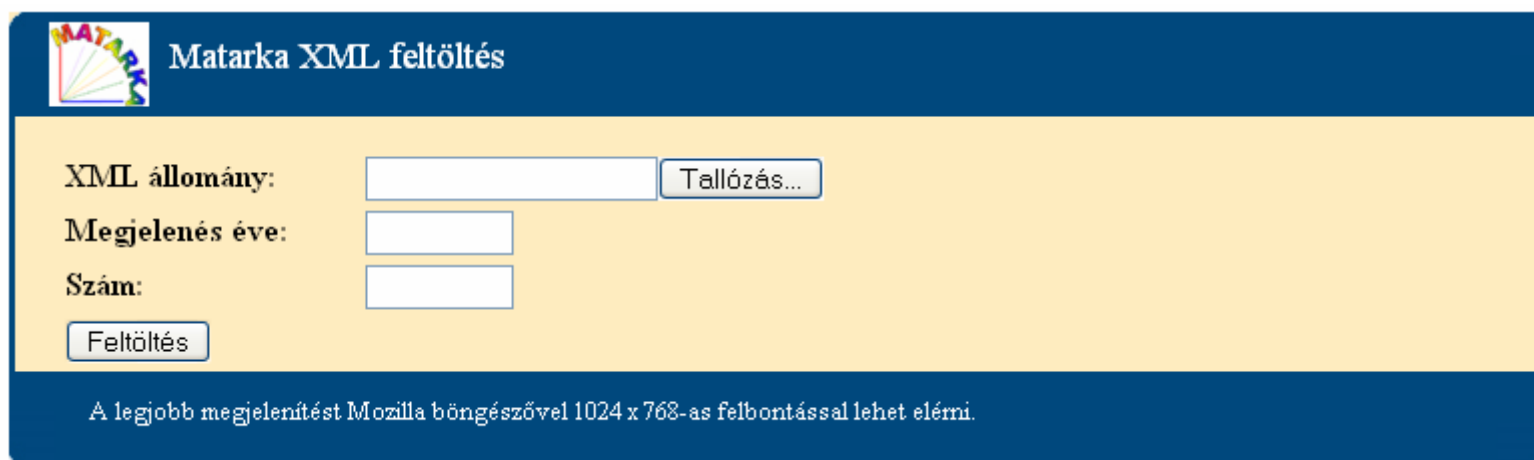
`perlaki@kvtlinux.lib.uni-miskolc.hu`

1. Feltöltés

A Matarka adatbázis feltöltését a közvetlen kézi bevitelen túl XML állományokból is el lehet végezni. A beviteli program web-alapú Java servlet alkalmazás.

Címe: <http://kvt99.lib.uni-miskolc.hu:8080/epa/fel.html>

Működéséhez [Tomcat](#) webserver szükséges és [Java](#) SDK megléte a szerver-gépen, kliensoldalon csupán webböngészőt igényel.



The screenshot shows a web interface titled "Matarka XML feltöltés". It features a logo on the left and a main form area with three input fields: "XML állomány:" with a "Tallózás..." button, "Megjelenés éve:", and "Szám:". A "Feltöltés" button is located below the input fields. A footer note states: "A legjobb megjelenítést Mozilla böngészővel 1024 x 768-as felbontással lehet elérni."

Webfelület

Mindössze három adatmező kitöltését igényli, a feltöltendő XML állományt kell kiválasztani, valamint megadni a megjelenés évét és számát, ha az nincs jelen az XML állományban, vagy attól eltérő jelölésre van szükség. A program ezek után átveszi az XML állományt és feldolgozza, ezt pedig a felhasználónak visszajelzi.

```
- <TOC>
  - <Head>
    <dc_title>Eresz</dc_title>
    <Issue year="2005" number="1">1.évf. 1.sz. (2005.)</Issue>
    <dc_identifier>00811</dc_identifier>
    <URIEPA>epa.oszk.hu/00800/00811</URIEPA>
    <ev_szam_URL>epa.oszk.hu/00800/00811/00001</ev_szam_URL>
  </Head>
  - <Contents>
    - <Article>
      <Link>eresz2005_01.pdf</Link>
```

XML fejléc

```
</Article>
- <Article>
  <Link>eresz2005_05.pdf</Link>
  - <Author>
    <FamilyName>Szabó</FamilyName>
    <GivenName>Etelka</GivenName>
  </Author>
  - <Title>
    József és Aszenáth: a görög szerelmi regény motívumai egy ókeresztény apokrifben
  </Title>
  <Range>5</Range>
</Article>
- <Article>
```

XML cikkeírás

2. Linkellenőr

A linkellenőrző program az eleMEK csomag egyik programjának módosított változata. Feladata a Matarka adatbázisában a linkek ellenőrzése és a hibásan válaszolók kigyűjtése egy listába. A program a listát e-mail-ben továbbítja a paraméterben megadott címre.

A program Java kódban íródott, MySQL adatbázis-kapcsolattal. Az adatbázisból kapott válasz alapján az URL-eket kulcsként használva listát készít és ezeket a HTTP protokollon át „megszólítja”. Ha a válasz bármilyen hibajelzéssel járt, akkor ezt feljegyzi, a helyes válaszokat csupán számolja.

A program futása időigényes, ezért éjszakára érdemes ütemezni.

A folyamat végeztével hibalista készül a következő formában:

```
azonosító> cím  
URL  
hibakód és leírása (angolul)
```

A jellemzőbb hibák:

- Hibásan megadott URL, pl. hiányzik a http:// rész.
- Feleslegesen hosszú cím, pl. session-azonosító beemelése.
- Korlátozottan elérhető cím (pl. tűzfalon belüli) megadása.
- Hálózati, vagy szerverhiba.

358529> Dalol (sie singt) a múlt ... - Dalos György: A körülmetélés; A nagy buli

<http://www.mozgovilag.hu/friss/13rolrol3.htm>

404: Not Found

391162> levélváltás:Mi újság a hajléktalan kutatásban?

www.3sz.hu/bm/Halo/korabbi/2003/Halo03oktober.doc

java.net.MalformedURLException: no protocol: www.3sz.hu/bm/Halo/korabbi/2003/Halo03oktober.doc

392210> Együtt-élés=Co-existence : vándorkiállítás Kelettől Nyugatig

www.3sz.hu/bm/Halo/korabbi/2003/Halo03dec.doc

java.net.MalformedURLException: no protocol: www.3sz.hu/bm/Halo/korabbi/2003/Halo03dec.doc

Összes URL: 59118 (ismétlődők is)

Ellenőrizve: 44450 (minden egyedi lapcím csak egyszer)

Rendben: 43801

Sikertelen: 649

Időigény: 26113 másodperc

3. Újraindexelő

A Matarka adatbázis szótábláinak újraindexelését végző C nyelvű program három modult tartalmaz:

- *Törlés.* Az indexek alapértékre állítása.
- *Cikkindex.* A cikkek címszavainak újraindexelése.
- *Szerzőindex.* A szerzők neveinek újraindexelése.

Az újraindexelés a felvitel során felvitt, de törölt, vagy téves szóalakok kiszűrésével a keresés hatékonyságát javítja. Tipikus karbantartási feladat, amit bizonyos időközönként érdemes elvégezni. A program észrevehetően terheli a szerveret és hosszan fut, ezért éjszakára időzítve érdemes futtatni.

4. NDA OAI kapcsolat

A Matarka teljes szövegre is hivatkozó rekordjai heti frissítéssel átkerülnek az NDA lapjára is, erre a célra az NDA által biztosított OAI szerver szolgál, amelyet helyben szükséges telepíteni és üzemeltetni.

Az adatútvonal:

- A Matarka MySQL tábláiból az *NDA TabExport* Java program JDBC illesztőn át kigyűjti az érvényes URL-t tartalmazó cikkek adatait és a hozzá tartozó szerzőket, majd az NDA „tab” formátumában kiírja.
- Az NDA *TabSeparatedConverter* Java programja a „tab” formátumban kigyűjtött adatokat a *HSQLDB* nevű adatbázisába tölti.
- Az NDA OAI szervere szolgáltatja a külvilág felé az adatokat.
- Az NDA távolról begyűjti és saját adatbázisába felveszi a rekordokat.

A rendszerhez tartozó, illetve szükséges programok:

- Matarka rendszer (helyi fejlesztés):
Apache webserver, PHP, MySQL
- *NDA*TabExport (helyi fejlesztés):
Java JRE, MySQL, MySQL JDBC
- *TabSeparatedConverter* (NDA):
Java JRE
- OAI szerver (NDA):
Tomcat webserver, Java SDK

A webszerver beállításakor módosítani volt szükséges a *server.xml* és a *web.xml* állományokat, valamint létre kellett hozni a *Matarka-OAI-Repository.xml* állományt.

A Matarka jelenlegi adatmegfeleltetései:

- cikk:cikksz → id (fejlécmező!)
- foly:fcim, fuzet:ev, fuzet:vol, fuzet:szam, cikk:elso_oldal
→ relation | isPartOf
- foly:furl → relation | isPartOf
- foly:kozreado → publisher
- foly:issn → identifier
- foly:nyelv → language
- cikk:rovat → title | seriesTitle
- cikk:cikkcim + cikk:cimfoly → title
- szerzo:szerzonev → creator
- cikk:url → identifier
- cikk:filemeret → format | extent
- cikk:fileformat → format
- cikk:datum → date.issued

5. Wiki a könyvtárban

A Wiki hatékony belső információmegosztó eszköz, mellyel a belső dokumentációk könnyű elérhetősége és naprakészen tartása biztosítható. Nincs a „melyik doc-ban is keressem?”, a „most vajon ez a friss verzió?” és a „még nem írtam le, de majd” jól ismert problémája.

Eredetileg teljesen nyílt hozzáféréshez dolgozták ki a Wiki alapprogramját, de egy kiegészítéssel olyan oldalakkal is bővíthető, melyeket csak meghatározott joggal bíró felhasználók olvashatnak, szerkeszthetnek.

PA vitám beállítások figyelőlistám közreműködésem kilépés

szócikk vita módosítás laptörténet **unrestrict** áthelyezés megfigyelés

Dokumentációk

Tartalomjegyzék [elrejt]

- 1 Számítógépek, kiegészítők
- 2 Adatbázisok
- 3 Könyvtári webszerver
- 4 Horizon int. rendszer
- 5 WebPAC
- 6 Matarka
- 7 Midra
- 8 OAI szerver
- 9 Beléptetőrendszer
- 10 Webjogtar
- 11 Windows alkalmazások, szervízprogramok

[szerkesztés]

Számítógépek, kiegészítők

- Szervergépek
- Dolgozói gépek
- Hallgatói gépek
- Hálózati egységek
- Nyomtatók
- **Scannerek**
- **Vonalkódolvasók**

navigáció

- Kezdőlap
- Közösségi lap
- Friss események
- Friss változtatások
- Lap találomra
- Segítség
- Adományok

keresés

Menj Keresés

segédlet

- Mi hivatkozik erre
- Kapcsolódó változtatások
- File felküldése
- Speciális lapok
- Nyomtatható változat
- Állandó ugrópont

Dokumentációk kezdőoldala

PA vitám beállítások figyelőlistám közreműködéseim kilépés

szócikk **vita** módosítás laptörténet **unrestrict** áthelyezés megfigyelés

Matarka programcsomag



A **Matarka** 2 fő része a szolgáltató programok és a segédprogramok. A szolgáltatási rész 3 példányban fut: magyar, teszt (magyar), angol. Két fő része a lekérdező és beviteli (az angolban nincs beviteli felület). Két szerverre van telepítve: kvt99 és kvt66.

A magyar és az angol rész a **tart** adatbázist használja, a **teszt** rész a teszt adatbázist.

A kód nagy része PHP és SQL kódokra épül. A megjelenítés HTML és CSS.

Tartalomjegyzék [\[elrejt\]](#)

- 1 Szolgáltatási rész
 - 1.1 Beviteli felület
 - 1.2 Lekérdező felület
 - 1.2.1 PHP fájlok
 - 1.2.2 Technikai fájlok
 - 1.2.3 Lekérdező felülethez kapcsolódó egyéb PHP fájlok
 - 1.2.4 Képek
- 2 Segédprogramok
 - 2.1 PHP fájlok
 - 2.2 Önálló programok

[szerkesztés]

Szolgáltatási rész

[szerkesztés]

Beviteli felület

[szerkesztés]

navigáció

- Kezdőlap
- Közösségi lap
- Friss események
- Friss változtatások
- Lap találmra
- Segítség
- Adományok

keresés

Menj Keresés


segédlet

- Mi hivatkozik erre
- Kapcsolódó változtatások
- File felküldése
- Speciális lapok
- Nyomtatható változat
- Állandó ugrópont

Matarka dokumentációk lapja

PA vitám beállítások figyelőlistám közreműködésem kilépés

szócikk **vita** módosítás laptörténet **unrestrict** áthelyezés megfigyelés



Tarka.php

navigáció

- Kezdőlap
- Közösségi lap
- Friss események
- Friss változtatások
- Lap találmra
- Segítség
- Adományok

keresés

Menj Keresés

segédlet

- Mi hivatkozik erre
- Kapcsolódó változtatások
- File felküldése
- Speciális lapok
- Nyomtatható változat
- Állandó ugrópont

Tartalomjegyzék [\[elrejt\]](#)

- 1 Inicializálás
 - 1.1 Adatbázis
 - 1.2 Hibakezelés
 - 1.3 Kivétel szavak betöltése
- 2 Függvények

Inicializálás [\[szerkesztés\]](#)

Adatbázis [\[szerkesztés\]](#)

adatbázis kapcsolat felépítése minden fájlban, amelyik tartalmazza a tarka.php-t

Hibakezelés [\[szerkesztés\]](#)

hiba.php betöltése

Kivétel szavak betöltése [\[szerkesztés\]](#)

\$skips[] tömb feltöltése a **kivetelszo** táblából

Függvények [\[szerkesztés\]](#)


- head_print (\$title,\$active,\$dir) HTML fejléc
- head_print_ico (\$title \$active \$dir) HTML fejléc JavaScripttel

Egy kódleírási részlet

PA vitám beállítások figyelőlistám közreműködésem kilépés

szócikk vita módosítás laptörténet **unrestrict** áthelyezés megfigyelés

Words.php



MATARKA - Magyar folyóiratok tartalomjegyzékeinek kereshető adatbázisa

Keresés Készítők Dokumentumok Statisztikák Segítség Vendéglétny Kapcsolat

Egyszerű keresés Haladó keresés **Böngészés** Folyóiratok

Keresés szó- és névalakok között

Szerző: 1850 évől 2007 évig

Keresési példa:
Szerző: bírázat
eredmény: Bírázat Ákosai (1 cikk), Bírázat István(1 cikk), ...

Cikk cím: (csak egy szó legyen be) 1850 évől 2007 évig

Keresési példa:
Cikk címe: bírázat
eredmény: bírázat (25 cikkben), bírázatalkotat (1 cikkben), bírázat (65 cikkben), bírázati (9 cikkben), ...

szócikk

navigáció

- Kezdőlap
- Közösségi lap
- Friss események
- Friss változtatások
- Lap találmra
- Segítség
- Adományok

keresés

segédlet

- Mi hivatkozik erre
- Kapcsolódó változtatások
- File felküldése
- Speciális lapok
- Nyomtatható változat
- Állandó ugrópont

SZÓINDEXEKBEN VALÓ KERESÉS ŪRLAPJA [\[szerkesztés\]](#)

- feldolgozó fájl: [mutato.php](#)

```

$index_szerzo : keresett kifejezés szerzők között
$index_szerzo_ok : szerző indexben való keresést kértek
$index_cim : keresett kifejezés címek között
$index_cim_ok : cím indexben való keresést kértek

```

Egy másik kódleírási részlet

Köszönöm a figyelmet!