

# eleMEK – Metaadat-kezelő és szolgáltató digitális gyűjteményekhez

PERLAKI ATTILA

perlaki@kvlinux.lib.uni-miskolc.hu

**Kulcsszavak:** metaadat, digitális dokumentum, XML, platformfüggetlen, Java, GNU GPL, WWW

Az „eleMEK” a digitális dokumentum-gyűjtemények platformfüggetlen, Java- és webalapú moduláris metaadat-kezelő eszköze. Kifejlesztésének alapötlete a Magyar Elektronikus Könyvtár (MEK) digitális dokumentumainak kezelési tapasztalataira alapozva született meg. A mindennapi használat problémái, a helyi nem kompatibilis, nem szakszerű, hiányos és elérhetetlen metaadatokkal szembesülve döntöttek a fejlesztők egy metaadat-szerkezet és annak kezelőjének kifejlesztéséről, a helyi gyűjtemények kezelői számára. Emiatt az „eleMEK” használata és módosítása is engedélyezett a GNU GPL alapján. A cikk a kifejlesztés néhány döntését és az eszköz jelenlegi állapotát ismerteti. Reményeink szerint ezzel új felhasználók és fejlesztők csatlakozhatnak.

## 1. Bevezetés

A metaadatok az egyre növekvő információhalmazban való eligazodásunkat segítik, a kategorizálást, a keresést teszik könnyebbé. Ezek a könyvtárakban váltak tudománnyá, s alakultak ki igen összetett adatstruktúrákká, ám a mindennapi életben is számos esetben alkalmazzuk őket egyszerűbb formáikban. Ahogy azonban egyre több az információ, a megszokott leírók szegényessé, használhatatlanná válnak. Az eleMEK egy hazai fejlesztésű segédeszköz digitális gyűjtemények metaadatainak kezelésére, erre alapuló rendszerek fejlesztéséhez.

A cikkben az eleMEK rövid bemutatására kerül sor az alkalmazás során leggyakrabban felmerülő kérdések alapján, melyek több általános metaadat-kezelési témát is érintenek. A kérdések és válaszok a probléma ismertetésétől a cél meghatározásán és az eszköz kifejlesztését érintő főbb döntések okain át a használatig vezetnek, nem csak a lehetséges felhasználók, hanem a továbbfejlesztés iránt érdeklődők figyelmére is számítva. Az eleMEK ugyanis nem csak szabadon hozzáférhető segédeszköz, de a GPL alapján módosítható is.

## 2. Metaadat-kezelés

### 2.1. Miért segédeszköz?

Az eleMEK nem szokásos szoftvercsomag abban az értelemben, hogy feltelepíti a felhasználó és használja, mint egy „fekete dobozt”. Hasznosabb úgy tekinteni rá, mint egy szabadon továbbfejleszhető szoftvereszköz-készletre, amelyből nem szükséges minden elemet felhasználni, amit pedig felhasználunk, azt saját rendszerünkbe, szolgáltatásunkba építjük bele, hasznos „elemként”. A neve is erre a gondolatra vezethető vissza.

### 2.2. Miért van rá szükség?

Metaadat az is, amit egy operációs rendszer gépünkön őrzött állományainkról szolgáltat, ha listázási parancsot adunk ki. Operációs rendszertől függenek a részletek, de általában legalább két szintről kapunk információt:

- Megtudjuk, hogy melyik meghajtón, könyvtárban (directory, folder) vagyunk, ennek mi a neve, elérési útja, mekkora helyet foglal és esetleg azt is, mikor hoztuk létre.
- Megtudjuk a benne tárolt állományok neveit, méretüket, létrehozási dátumait, jogosultságait, esetleg egy kisebb képi ikont is – vagy közvetlenül, vagy egy-két kattintás után.

A lényeg tehát, hogy nem magát az adatot, hanem az adatról szóló adatokat kaptuk meg (az operációs rendszerek eddigi fejlődési sajátosságai miatt meglehetősen korlátozottakat). Ez alkalmas korlátozott keresésekre (például keresem a 2004-es adóbevallásom tábláit), kategorizálásra (kiválogatom azokat a fotókat, amiket télen készítettem), de általában csak emberi segítséggel megy, automatikusan nem.

Ennek az oka az, hogy a szükséges további metaadatok jó részét a fejünkben tároljuk még mindig (például tudom, hogy a digitális fényképezőgépem DSC betűhármással kezdődő állományneveket állít elő, tehát a sok kép közül valószínű, bár nem biztos, hogy ezek általában készítettek). Alaposan elgondolkodva bizonyára előnyösebb lenne, ha nem nekünk kellene megjegyezni hogy egy-egy állománynév, állománytípus, vagy egy-egy könyvtár mit is azonosít, hanem tetszőleges, vagy legalábbis a mainál sokkal bővebb leírásokkal láthatnánk el anyagainkat.

Ha az állománynév nem mond semmit, ma csak az egyenkénti „belenézegetés” marad lehetőségnek. Ami pedig több tízezer állománynál már reménytelen feladatnak tűnik. Több százmillió gép gépenként több tízezer nyilvánosságnak szánt adatánál biztosan az.

A több tízezer állomány a legutóbbi évekig olyan mennyiség volt, ami még kisvállalatokat sem mindig fenyegetett. Egy magánszemély csak gyűjtőként szembeesült a katalogizálás problémáival. Egy (vállalati) kis-könyvtár néhány száz kötetével is „fejből” elboldogulhatott a sok év rutinjával rendelkező könyvtáros. Ennek azonban már vége, egyetlen digitális fényképezőgép is több száz fotót tud egy kártyán tárolni, amit másodpercek alatt lehet „átönten” a számítógépre, majd onnan akár az internetre. Visszafelé pár nagyságrenddel még több állomány kerülhet a gépre. A „hova is tettem azt a képet, szöveget?” kérdés egyre nehezebben megválaszolható.

### 2.3. Mi az ismert megoldás?

A könyvtárak már jóval a digitális gyűjtemények kora előtt ismerték és használták a metaadatokat, ha nem is éppen így hívták őket. Ha rátekintünk egy katalógus cédulára, azon a legfontosabb információkat általában hamarabb megtaláljuk, mintsem a polcok közti bolyongással – már amennyiben egyáltalán kint van a könyv és nem raktárban. Megtudjuk a szerzőt és a pontos címet, a kiadót, a kiadás dátumát, néhány azonosítókódot, a témakört, néhány kulcsszót és a könyv megtalálását segítő információt. Ha publikációt keresünk, általában annak rövid összefoglaló leírását is olvashatjuk, mely segít dönteni, valóban ezt keressük-e? Ráadásul a katalóguscédula csak az egyik, tulajdonképpen a legegyszerűbb és legszegényebb metaadat a könyvtárak által használtak közül.

Talán érthető, hogy amikor a digitális gyűjtemények metaadat-kezelésének problémáját megoldani kívánták, elsősorban a könyvtárak tapasztalataira építettek, ám ennek mutatkozott néhány árnyoldala is.

Az egyik, hogy a könyvtáraknál kialakult összetett leírási szabályok egyrészt túl bonyolultak voltak más irányú felhasználáshoz, másrészt sok esetben nem voltak megfelelően illeszthetők az adott problémákhoz. Alkalmos szabvány nem lévén a helyi jellegű megoldások

szaporodtak el, és az ezek közti adatcsere gazdaságsan gyakorlatilag kivitelezhetetlen volt. Hozzájárult még ehhez, hogy míg a könyvtárakban szakképzett személyek végezték a leírásokat, addig a digitális gyűjteményeknél elemi hibákkal és hiányosságokkal terhelt metaadat-bázisok születtek.

Ezenközben pedig a digitális tartalmak (nem csak könyvek!) robbanásszerűen szaporodnak, s sajnos vesznek el a digitális kásahegyben, hiába a félelmetesen jónak tűnő és adatbányászati eszközöket is bevető webes keresők egyre elképesztőbb teljesítményei.

### 2.4. Nincs túl késő?

Be kell ismerni, hogy általában véve is az interneten elkésettnek látszik a metaadat-kezelés szabványosított bevezetése. Túl sok leíratlan digitális dokumentum került már fel, melyek közt legfeljebb a teljes szövegű keresőkkel van valami remény megtalálni a keresett információt, rendezettségéről szó sincs. Az internet hőskorában még léteztek elképzelhetetlen munkaigénnyel létrehozott tematikus katalógusok, ezeket azonban, ha csak nem speciális célúak, az adattömeg szó szerint maga alá temette.

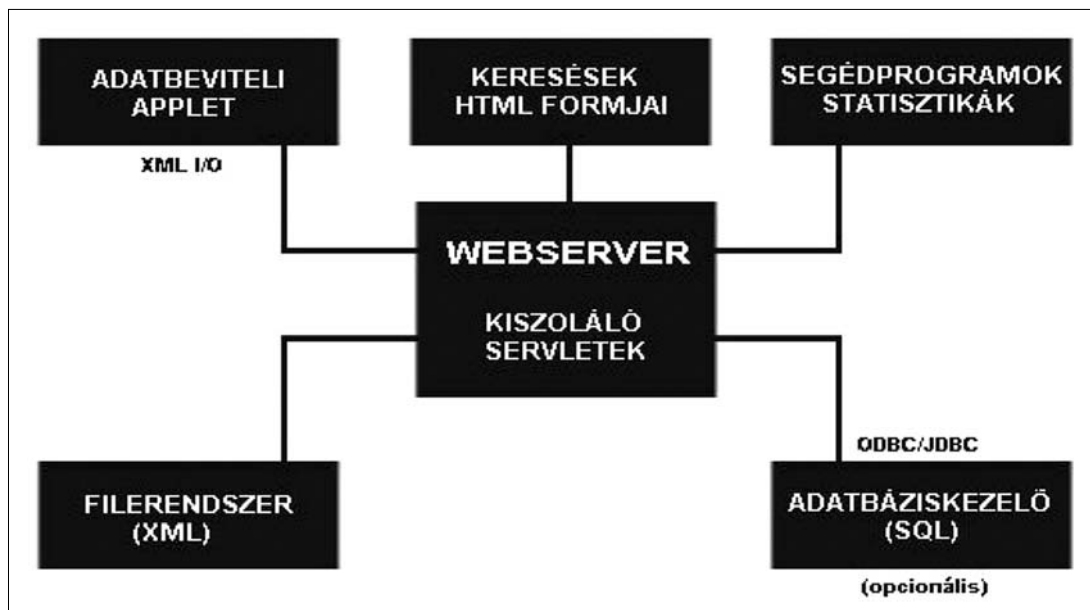
Márpedig az a dokumentum, ami nincs meg, gyakorlatilag nem is létezik, hiába készült el.

## 3. Szoftver

### 3.1. Miért indult el az eleMEK fejlesztése?

Az utólagos katalogizálás az erre váró anyag mennyiségétől függően vagy felesleges erőfeszítésnek tűnik (mert kevés az anyag, „úgy is tudom, mi hol van”), vagy reménytelennek (“ki fog tizenkétezer képet átnézni és egyenként leírni?”). Mégis, valami ilyesmire (és egy kis további rendszeres pluszmunkára) szeretné az eleMEK rávenni az érdeklődőket, cserébe sok felhasználónak sok-sok idejét megtakarítva a kereséskor. Valamint esélyt nyújt arra, hogy a keresés eredménnyel járjon.

1. ábra  
Az eleMEK  
webszerver köré  
épülő eszközei



A Magyar Elektronikus Könyvtár (MEK) 1994 óta gyűjti a tapasztalatot a digitális gyűjtemények kezelésében és szolgáltatásában. Itt merült fel a gondolat, hogy a nagy központi szolgáltatás mellett egy felhasználónál üzemeltethető eszköz is kerüljön megvalósításra. Az egyik fontos cél az, hogy az egymással nem kompatibilis helyi, sokszor ötletszerűen összeállított, különféle formátumokban és rendszerekben tárolt metaadatok „közös nevezőre” legyenek hozhatók, amely nélkül gyakorlatilag nem lehet információt cserélni, keresni.

Az eleMEK legfontosabb alkotó eleme ezért nem valamiféle program, hanem maga az adatszerkezet, amelyhez, mintegy „szerszámosládaként” csatlakozik a programcsomag.

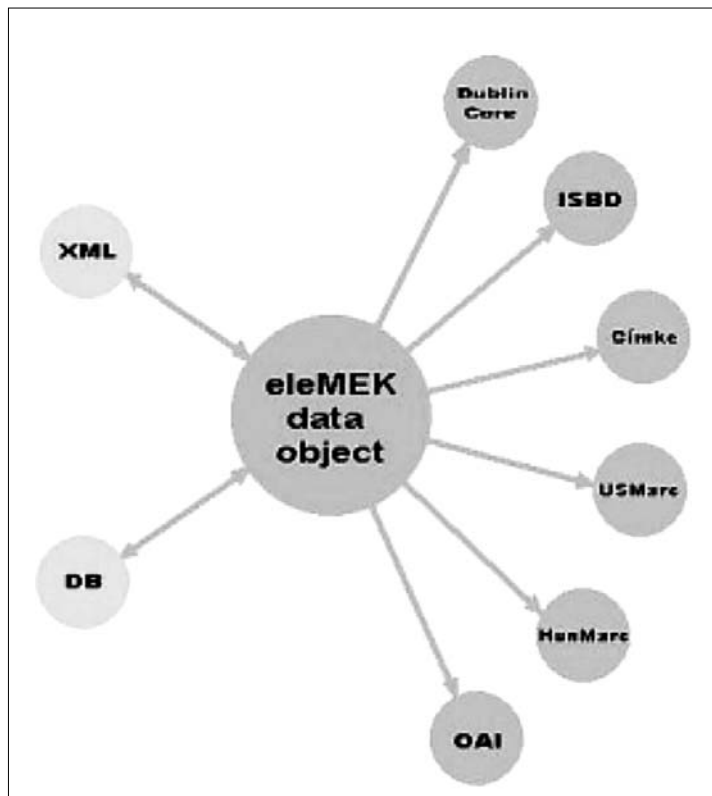
### 3.2. Mit találunk a szerszámosládában?

Az eleMEK szoftvereszközei (1. ábra) a felhasználó szemszögéből a következő részekre bonthatók:

- adatrögzítés és módosítás,
- keresők, listázók,
- import, export és archiváló eszközök,
- ellenőrző, karbantartó és statisztikai eszközök.

A keresők kivételével, amelyek webes felület alá tartoznak, az eszközök lényegében a metaadatokat reprezentáló XML állományok (és igény szerint adatbázistáblák) kezelésére szolgálnak, beleértve a statisztikák készítését is. Az egész csomag felépítése annak figyelembevételével készült, hogy minden eleme igény szerint helyettesíthető legyen, az adatok akár egy szövegszerkesztővel is olvashatók, módosíthatók legyenek.

2. ábra Az eleMEK belső és külső adattípusai



### 3.3. Miért XML?

Az XML szabványú adatrögzítés ma már egyfajta elvárás, amelyet komoly általános célú és ingyenesen is hozzáférhető szoftverek támogatnak, ugyanakkor mentes a korábbi szigorúan kötött szabványos formátumok korlátaitól és az egyedileg megalkotott nem szabványos formátumok hozzáférhetőségét akadályozó tényezőktől. Megköveteli, hogy a célnak megfelelő „nyelvtant” definiáljuk és hozzáférhetővé tegyük, ugyanakkor szerkezeti megkötései alig vannak. Ez különösen szoftver-objektumok tárolásakor hasznos, amelyek összetettségüknél fogva például relációs adatbázis-kezelőkbe nehezen képezhetők le.

Az eleMEK adatszerkezete kétszintű, mindkét szintjén elhagyható és kötelező részekkel, melyek egyediek és ismételhetők is lehetnek. Ez áttekinthető és XML-ben jól megvalósítható. Figyelemmel az eddig felmerült kimeneti formátumokra (2. ábra), a konverziók általában egyszerűek, legfeljebb közepesen bonyolultak, a beépítettek (pl. DublinCore, USMARC) kívül könnyen írhatók hozzá újabbak, akár valamilyen programozási nyelven (C, Java, PHP), akár az XML konverziók saját eszközével, az XSL-lel. Az eleMEK kódja jelenleg Java.

### 3.4. Miért Java?

A cél a platformfüggetlenség volt. A legelterjedtebb operációs rendszerekre a Javát futtató környezet, a JRE ingyenesen letölthető, a programnyelv korai éveiben bíralt erőforrásigény pedig ma már alapszintű kiépítéssel rendelkező PC-k esetén sem számottevő. A kód az 1.4-es verziójú Java képességeire épül, jelen pillanatban már az 1.5-ös tölthető le, valószínűleg az eleMEK elkövetkezendő kiadásai már ehhez alkalmazkodnak az ebben a verzióban bevezetett számos előnyös nyelvi fejlesztés miatt. Ez persze attól is függ, kik hajlandóak a csomagot fejleszteni. A lehetőség adott, mivel a csomag GNU GPL alapján terjeszthető és módosítható.

### 3.5. Miért GPL?

A cél a már fentebb is említett információcsere elősegítése, amit egy teljesen szabadon hozzáférhető és továbbfejleszthető nyílt eszközzel célszerű elérni. Az egyetlen összetevő, amit a kompatibilitás fenntartása érdekében az eredeti fejlesztő kézben szeretne tartani, az adatszerkezet.

A GPL tehát annak, aki nem csak használni, de továbbfejleszteni is szeretné a csomagot, lehetővé teszi, hogy a forráskódot megtekinthesse, módosíthassa, s ezt akár terjeszthesse is.

## 4. Alkalmazás

### 4.1. Mi kell hozzá?

A csomag feltételez néhány összetevőt a működő operációs rendszeren – és szolgáltatási igény esetén – az élő hálózati kapcsolaton túl is. Ezek a következők:

- Java SDK, legalább 1.4-es verzió, 1.5 ajánlott. Java JRE csupán akkor elegendő, ha csak adatrögzítésre és konverziókra használjuk a csomag érintett részeit.
- Tomcat 4.0, vagy magasabb verziójú webservert, amely a keresési szolgáltatásokat támogatja, Java Servlet Pages (JSP) kezelésére képes. Már meglévő webservert mellé is telepíthető és üzemeltethető, ekkor általában a 8080-as porton.
- SQL adatbázis-kezelő, JDBC csatolással. Ez körülbelül 1000-1500 tétel, tehát kis gyűjteményeknél nem alapkövetelmény, a kereső tisztán állományalapú működésben is viszonylag gyors eddig a határig.

A használatbavételt célszerű úgy megkezdeni, hogy a fent említett komponensek már tesztelve felkerültek. A csomaghoz mellékelt dokumentáció kitér mind a Windows, mind a Linux alapú telepítés menetére, és a Tomcat webservert alapbeállításain szükséges változtatásokra is.

Két olyan eleme van a csomagnak, ahol a kezelési felület alapvető fontossággal bír. Az egyik az adatbeviteli eszköz, a másik a keresés.

#### 4.2. Mennyire felhasználóbarát?

Az adatbevitelt egy Java applikáció végzi (3. és 4. ábra), amit szükség esetén appletként is el lehet indítani, amennyiben a felvitelt egy központi gépről kívánjuk felügyelni. Ez egyben azt is jelenti, hogy a program grafikus környezetet igényel, Linux alatt karakteres terminál üzemmódban nem működik.

A feladata lényegében a metaadatok felvitele és a szükséges XML állomány előállítás, első közelítésben egy „felokosított” XML szerkesztő célprogram, azonban a beleépített tudás és az adatstruktúra összetettsége viszonylag bonyolult kódot eredményezett.

A felviteli folyamat egyszerű: ki kell keresni egy már létező metaadatot (ami akár importálással is a rendszerbe kerülhetett) és azt kell szerkeszteni, vagy újat kell létrehozni. Az adatfelvitel lapokra osztva történik, a néhány kötelező adat felvitele és rögzítése után bármikor megszakítható és később folytatható. A felvitel során bármikor a „nézet lapra” lehet kapcsolni és a felvitt adatokat többféle formátumban ellenőrizni.

A program több adatmezőt „listás”, illetve „indexes” meghatározással kezel, ami azt jelenti, hogy ezekhez a beviteli felületen (jobb oldalt) értéklistán kérhetők és a bevétel csupán választás és kattintás. A különbség,

3. ábra  
Az eleMEK adatbevitelének „Megtekintés és mentés” képernyője



hogy a „listás” mezők csak választhatók, ezek értékeit csak az eleMEK adatait karbantartó „adminisztrátor” módosíthatja közvetlen javítással, vagy a mellékelt adminisztrátori eszközzel, míg az indexek a bevétel során automatikusan bővülnek, például új szerzői családnevekkel. Ugyanitt található még egy kicsiny ablak, ahonnan az aktuális adatlaphoz ajánlott teauruszok hívhatók meg, on-line.

Ehhez a programhoz külön súgó tartozik (3. és 4. ábra, felső sor, kérdőjel), amelyet egy szakkönyvtáros állított össze, természetesen a csomagban ez is megtalálható. Ez azt jelenti, hogy a használatot mindenképpen tanulni és gyakorolni kell, ha a felhasználó pontos, jól kezelhető, szakszerű adatokat kíván rögzíteni.

A másik komponens, amelynél fontos a kommunikáció a felhasználóval, s aki ez esetben a szolgáltatást igénybevevő és nem az adatokat rögzítő, a keresés. Ez alapvetően más környezetben és megfontolások alapján került megvalósításra. Mindössze két webes „kérdőív”, illetve az eredménylisták alkotják a látható felületet, melyet az alkalmazó a maga képére alakíthat, s ezért gyakorlatilag a csomagban ennek a résznek a megjelenése alaphelyzetben puritán. A két kérdőív egy

úgynevezett gyorskeresést és egy olvasói, bővebben beállítható, paraméterezzhető keresést (5. ábra) tartalmaz. Az eredményül kapott listán megjelenő adatok gyűjteményenként konfigurálhatók, tilthatók, vagy engedélyezhetők, illetve maga a keresés erre a listára még szűkíthető. Ennek a résznek a testre szabása tehát inkább weblap-készítői tapasztalatokat igényel.

#### 4.3. Hogyan tovább?

Amíg olyan operációs rendszer nem kerül a kezünk alá, amely egyenesen megköveteli a mainál részletesebb metaadatok használatát, addig segédeszközökre szorulunk. Az ígéretet persze már évek óta elhangzanak, de messze vagyunk a hétköznapi, elterjedt használattól. Az eleMEK ebből a szempontból hiánypótló. Sajnos a tavalyi évben a fejlesztés háttérét biztosító NIIF forráshiány miatt kénytelen volt szüneteltetni a támogatást, a tervek és a még el nem készült kiegészítések „fiókban hevernek”. Ezek közül a periodikákat, dokumentum-csomagokat is kezelni képes, metaadatokat hierarchikus módon kezelő kiegészítés és az offline eleMEK emelhető ki, utóbbi alkalmazható például CD-n kiadott digitális gyűjtemény elemeinek keresőrendszereként.

4. ábra  
Az eleMEK adatbevitelének „Szerző” képernyője

## 5. Összefoglalás

A digitális gyűjtemények metaadatainak kezelésére szolgáló eleMEK közkinccs, egy nyílt eszköz-csomag, amit használni és fejleszteni lehet. Célját, hogy a metaadatokat könnyen kicserélhető, könnyen értelmezhető, szabványoknak megfelelő formában tárolja és szolgáltatassa, mai formájában is teljesíti. Elterjedéséhez azonban nagyobb ismertség szükséges és annak felismerése, hogy sem az átlagos felhasználó számára megfizethetetlen és kihasználhatatlan nagy rendszerek, sem az egyéni ötleteken alapuló zárt fejlesztések nem alkalmasak a digitális tárolt tartalmak hatékony megtalálására, rendszerezésére.

## Irodalom

- [1] Drótos László:  
eleMEK – Metaadat-kezelő rendszer digitális gyűjteményekhez,  
TMT 2005/2  
<http://elemek.oszk.hu/ismertetok/elemekek-cikk.htm>
- [2] Drótos László, Perlaki Attila:  
Moduláris könyvtári rendszer elektronikus dokumentumgyűjtemények kezeléséhez: az eleMEK projekt,  
Networkshop, Győr, 2004. április  
<http://elemek.oszk.hu/ismertetok/netwshp2004.htm>
- [3] Drótos László, Perlaki Attila:  
eleMEK rendszerterv,  
NIIF weblap (2004-től folyamatosan aktualizált)  
<http://elemek.niif.hu/rendszerterv/index.html>
- [4] Drótos László, Góczán Andrea:  
eleMEK kitöltési útmutató,  
NIIF weblap (2004-től folyamatosan aktualizált)  
<http://elemek.niif.hu/help/index.html>

5. ábra Az eleMEK weblapú keresőjének „olvasói keresés” kérdőíve

eleMEK olvasói katalógus - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://elemek.niif.hu/action.jsp Search

Home Bookmarks mozilla.org mozillaZine mozdev.org Wikipedia Google JDK 5 Documentation SZTAKI

Rendezés:  
 Cím szerint  Szerző szerint  Időrendben

Cím :		és
Szerző : Név		és
Téma : Alttémakör		és
Tárvszó : Tárvszó / kulcsszó		és
Dokumentumtípus : A típus neve		és

Dátum  
 -tól  -ig

Keresés ékezet nélkül.

Törlés Keresés