

Ignéczi Lilla

Neumann János Digitális Könyvtár és Multimédia Központ Kht.

Tématérkép belső fejlesztéssel

Vizuális teaurusz a magyar internet katalógusában

A fogyasztói és információszerezési szokásokban, a szolgáltatásokban és a fogyasztók megszerzésében bekövetkezett változások a képi befogadás irányába mutatnak. A mindennapokban is azzal kell szembesülnünk, hogy a vizuális információk tömegével árasztanak el minket: egyszerűbb, a felhasználóra gyorsabban hat a képi formában megjelenő híranyag. Tény, hogy a vizuális megjelenítés a legősibb kifejezési forma, és a bonyolultabb összefüggések, kapcsolatok szemléltetésére a legalkalmasabb. Emellett megfigyeljük, hogy az emberek egyre kevesebbet olvasnak, viszont sokkal többet néznek tévét, járnak moziba, az interneten szívesebben nézegetnek képeket, ábrákat. A Neumann-ház ezeket a szempontokat figyelembe véve fejlesztette 2003-ban, és helyezte a 2004. év elején üzembe a „tématérkép” szabványon alapuló katalógusát.

Történeti előzmények

A magyar internetkatalógus, a *WebKat.hu* fejlesztéséhez 1998-ban kezdtünk hozzá, a rekordok HUNMARC adatcsere formátumú szerkezetén alapuló katalógus online elérhető formában 1999. szeptember elsején jelent meg az interneten.

A katalógusban egyetlen kattintással új böngészőablakban azonnal megjeleníthető a dokumentum, mivel az egyes rekordokhoz hozzárendeljük az URL-eket is. Így az rögtön olvasható. A katalógus legfőbb jellemzője – éppen a dokumentumokhoz való hozzáférés folyamatos biztosítása érdekében – a rendszeres karbantartás, a rugalmas szolgáltatás, a szakmai és a felhasználói igények folyamatos figyelése, és a szolgáltatások ezekhez igazítása.

A formai feldolgozás mellett a tartalom tárgyszavas feltárására is szükség volt. A teaurusz kialakítása azonban csak később kezdődött, mivel 1999-ben a Neumann Könyvtár még számított egy közös, országos szintű új átfogó tárgyszórendszerre. A *WebKat.hu* gyarapodása azonban a tartalmi feltáró munkák megkezdését sürgette, ezért, ha később is, de megkezdődött a teaurusz szerkesztése, amely 2000 nyarán készült el. Szeptemberig folytak a bővítések, és október elsejével megkezdődött a teaurusz lexikai egységeivel végzett tartalmi feltárás (a „tárgyszavazás”). Az integrált rendszer kezelő szoftver bizonyos korlátokat szabott: így nem tudunk a teauruszokra jellemző minden

relációt létrehozni a lexikai egységek között. Mindössze az alá-fölé rendeltségi kapcsolat, a lásd még kapcsolat és a „lásd”-dal jelzett utaló meghatározására volt lehetőség az adott körülmények között.

Az interneten bekövetkezett változások, a böngészésre alkalmas keresőgépek elterjedése miatt 2002-ben úgy döntöttünk, hogy a teauruszból kiindulva fejlettebb keresőrendszert hozunk létre. Megtehetjük, hiszen a rendszerben mindössze 2000 keresésre rendelkezésre álló lexikai egység (tárgyi kifejezés) található, elég kis mennyiség ahhoz, hogy szellemi munkával a szerkezetét át-
dolgozzuk, és a változó igényekhez igazítsuk.

Megtartottuk azt az öt fazettát, amelyek részben a rendszer ETO alapjaira, részben pedig a szempontok szerinti osztályozási rendszerekre utalnak: *tárgyszó, nép/nemzetiség, idő, földrajzi hely, műfaj*. Kiemeltünk viszont a *tárgyszó* fazettából kilenc, legfelsőbb szintű fogalmi kategóriának kijelölt osztályozó fogalmat: egészség, gazdaság, kommunikáció, kormányzat, kultúra, műszaki tudomány, művészet, társadalomtudomány, természettudomány. Forrásként használtuk a Köztauruszt, az AltaVizsla és a Yahoo! katalógusának hierarchikus osztályozási rendszerét, valamint a National Geographic honlapjának fogalmi hierarchiáját. Fogalmi rendszerünk eme kialakításával lényegében felkészültünk arra, hogy a katalógust hierarchikusan kezelni és megjeleníteni tudja az erre a célra alkalmas rendszer, egyben figyelembe vettük az Olib teauruszkezelő moduljának korlátait.

Új hierarchikus keresőrendszerünk – tehát a hierarchikus szerkezetű katalógusunk – kezelő- és vizuális megjelenítő rendszeréül a „tématerképnek” (topic map) nevezett szabványon alapuló szoftvert választottunk. A weben található vizualizált fogalmi struktúrák jelentős része alapul ezen a szabványon. Új katalógusunk, a WebKat.hu tématerképe 2003. december 31-ére készült el, és a Neumann-házban feldolgozott webdokumentumokat szolgáltatja a keresők számára. A tématerképen alapuló vizuális megjelenítés úttörő vállalkozás a hazai könyvtári gyakorlatban, jelenleg a Magyar Elektronikus Könyvtár fejlesztését végzik ugyanilyen alapon. A hierarchikus tématerképpel a tárgyi kifejezések rendszere (a „tárgyszórendszer”) vizuálisan jobban áttekinthető, bejárható, keresésre is alkalmas felülete a téma szerinti navigáció egyik alternatívája.

A fejlesztés célja

Tématerképes katalógusunkkal a WebKat.hu szolgáltatásait, tartalmi keresési lehetőségeit kívántuk bővíteni. Az elmúlt években ugyanis hiába igyekeztünk a rendszerrel szállított WebView felületet sokkal „felhasználóbarátabbá” tenni, tapasztalataink szerint nagyon kevesen értik meg, hogyan lehet az online tezaurusz alapján a tartalom szerinti keresést hatékonyabbá tenni.

Időközben megváltoztak az internetfelhasználók szokásai, nagyon eltérő lett a felkészültségük szintje, sokan előnyben részesítik a vizuális élményeket. Ezért olyan fejlesztésre kértünk és kaptunk is pályázati támogatást, amely a már meglévő tezaurusz relációit, elsősorban a hierarchikusakat vizuális struktúrában jeleníti meg. Fejlesztésünknek az is célja volt, hogy új felhasználókat nyerjünk meg szolgáltatásunknak, mert a pusztán szókereséssel végzett tájékozódás mind kevésbé elégíti ki az igényeket.

A kereséshez használható struktúra grafikus ábrázolása, az összefüggések vizuális megjelenítése egyre fontosabb számukra: ennek a szerkezetnek a segítségével könnyebben megértik a fogalmak közötti kapcsolatokat, illetve a kapcsolatok mentén könnyebben találnak újabb összefüggésekre. A szemantikai összefüggések vizualizáltabb ábrázolásával a tartalom szerinti keresőrendszerünk megértését, átláthatóságát és használatát tettük egyszerűbbé. A létrehozott tématerkép az online elérhető dokumentumok keresését támogatja.

Előzetes felmérés

Kész megoldások tanulmányozása

A tématerképen alapuló kezelő és megjelenítő rendszerek viszonylag új termékek, egyelőre világszerte csak maroknyi cég foglalkozik szabványos tématerkép-szoftverek fejlesztésével, értékesítésével, illetve adaptálásával:

- *Empolis* (korábban Step; ma a német Bertelsmann-csoport tagja),
- *Infoloom* (Egyesült Államok),
- *Mondeca* (Franciaország),
- *Ontopia* (Norvégia),
- *TOTALZOOM* (Magyarország).

Közülük Magyarországon az *Empolis Kft.* foglalkozik a tématerképekkel. A cég Németországban, Lengyelországban, Nagy-Britanniában, Norvégiában és az Egyesült Államokban is működtet leányvállalatot, világviszonylatban körülbelül 400 alkalmazottat foglalkoztat. *K42* elnevezésű tématerkép-szoftverének első változata több mint egy éve jelent meg a piacon. A *K42* teljesen megfelel az XML alapú topic map (XTM) ajánlásnak, és a tématerképeknek nevezett hierarchikus fogalmi rendszerek, s velük katalógusok létrehozására, kezelésére, karbantartására, bejárására és a segítségükkel végzett böngésző keresésre egyaránt módot ad. A keresést a TMQL (Topic Map Query Language) speciális lekérdezőnyelv támogatja. A rendszer része az optimalizált működésű, nagy kapacitású információszervert, amelyet Jávára alapoztak.

A megvalósítás lehetséges módjai

1. Külső szakértő cég megbízása

Előnyei:

- a határidő könnyebben tartható,
- professzionális technológia,
- professzionális kivitelezés.

Hátrányai:

- drága (több millió Ft, vagy több tízezer dollár a tématerkép keresőszoftver – keresőgép),
- a cégnek teljesen át kell látnia a mi rendszerünket – ez időigényes,
- a tudás a kivitelező cégnél marad,
- a további fejlesztés további külső megbízásokkal jár,
- a kész rendszer nem vagy nehezen alakítható teljeskörűen a mi igényeinkhez.

Tématerképekkel két cég foglalkozik ma Magyarországon: Az *Empolis* ajánlata a *K42* szabványos

tématérkép-szoftver. A TOTALZOOM ajánlata a nem szabványos, Flash memórián („nem felejtő” memórián) alapuló megoldás, újak a szakmában, és az általuk forgalmazott termék ugyan nem ki-mondottan tématerképek létrehozására született, de azok megjelenítésére tökéletesen alkalmas. Az árajánlat összege mindkét esetben jelentősen meghaladta az anyagi kereteinket, így erről a lehetőségről is le kellett mondanunk.

Kérdések mindkét esethez:

- Hogyan illeszthető a készen kapott tématerkép a rendszerünkhöz?
- Az Oracle alapú OLIB-bal mennyire tud együttműködni?
- Csak a tématerkép megvalósítását oldja meg, vagy annál többet is nyújt?

2. Saját fejlesztés a szabvány alapján

Előnyei:

- a tudás házon belül marad;
- az „XTM topic map” szabvány elérhető, megtanulható;
- jól ismerjük a rendszerünk felépítését, sajátosságait (OLIB, WebKat.hu);
- könnyebb továbbfejleszteni, ha a kezdetektől mi építjük;
- saját elképzeléseink szerint formálhatjuk.

Hátrányai:

- nem biztos, hogy teljesen kész lesz határidőre (mindössze 4 hónap állt rendelkezésre);
- jelentős kapacitást von el a többi feladattól.

3. Saját fejlesztés az ingyenes, nyílt forráskódú tématerkép keresőszofver átalakításával

Előnyei:

- már több helyen kipróbált rendszer alkalmazása;
- szabványos megoldás;
- költségkímélő megoldás;
- a tudás házon belül marad;
- jól ismerjük a rendszerünk felépítését, sajátosságait (OLIB, WebKat.hu).

Hátrányai:

- sok tanulással és kísérletezgetéssel jár;
- nem biztos, hogy határidőre sikerül alkalmaznunk a technológiát.

Tekintettel arra, hogy egy XTM alapú nyílt forráskódú tématerkép-szoftverre alapozott fejlesztés elvégzésére nem lett volna elegendő az időnk, és a pályázaton nyert összeg sem volt hozzá elég, ezért a lehetőségek mérlegelése után a 2. megoldás mellett döntöttünk, bár később azt is módosítanunk kellett.

Megvalósítás

Saját erőből az XTM technológia elsajátítása időigényesnek bizonyult. A készen vásárolható tématerkép-keresőgépek meglehetősen drágák, ezért saját tématerkép-szoftver kidolgozása lett a cél. Ennek előnye, hogy az egész rendszert úgy alakíthattuk, ahogy elterveztük, a fejlesztés alatt folyamatosan tudtunk tesztelni, a tudás a birtokunkban maradt, ezáltal a későbbi fejlesztéseket is magunk tudjuk majd elvégezni.

Tématérképünk egy MySql adatbázisra épülő, PHP és HTML technológia kombinációjával létrehozott webfelület, amely vizuális böngészést, valamint egyszerű – AND/OR logikai műveleteket tartalmazó – és összetett keresést tesz lehetővé a WebKat.hu katalógusban. A webfelület készítésekor a W3C ajánlásait követtük.

A WebKat.hu tezaurusát és átalakított hierarchikus rendszerét az Oracle adatbázis-kezelőből vettük át. Az így kapott adatokat egy, a tématerkép-hez speciálisan kialakított MySql adatbázisba való betöltéshez az adatbázis szerkezetének megfelelő szövegállománnyá (.txt) alakítottuk át. Minden egyes lexikai egység egyedi rekordazonosítót kapott az új adatbázisban is. A MySql adatbázist az Oracle-ből rendszeres áttöltéssel aktualizáltuk. Alapvető szempont volt a felhasználók igényeinek kielégítése, ezért a tématerképpel nem váltottuk ki a tárgyi kifejezések (témák) szerinti keresést, hanem alternatív lehetőségként ajánljuk fel.

A fejlesztés lépései

1. A tárgyi kifejezések (tezaurusz lexikai egységek, egyéb kifejezések) teljes állományát keresőprogram segítségével az Oracle alapú OLIB integrált könyvtári rendszerből két formátumban vettük át.
2. A tárgyi kifejezéseket minden relációval együtt egy saját fejlesztésű MySql adatbázisba töltöttük.
3. Kialakítottuk a tématerkép webes megjelenítéséhez szükséges PHP/HTML kódokat.
4. Összekapcsoltuk a tématerkép „témáit” egy-egy OLIB lekérdezéssel, hogy konkrét találati halmazokat kaphasson a felhasználó.
5. A fejlesztés közben folyamatosan teszteltük a megoldásokat. A munka befejeztével (még az üzemszerű szolgáltatásba illesztés előtt) minden részletre kiterjedő, több szempontú tesztelést végeztünk. A tesztelés kiterjedt mind a mű-

kódésre és az esetleges programozási hibák feltárására, mind a webes megjelenítésre. A felhasználói szokások figyelembevételével folyamatosan formáltuk a felületet.

6. Az új szolgáltatást beillesztettük a WebKat.hu rendszerébe, amelyben a keresés új lehetőségét képviselte.

A tématerkép a WebKat.hu katalógus menürendszeréből érhető el.

Irodalom

HUNMARC, a bibliográfiai rekordok adatcsere formátuma. Összeáll. Sipos Márta; kész. az Országos Széchényi Könyvtár Fejlesztési Osztályán. Budapest: OSZK, 1993. 129 p. ISBN 963 200 344 6

HUNMARC, a besorolási rekordok adatcsere formátuma. Összeáll. Sipos Márta. 1998. március. 61 p. Tervezet

Köztaurusz. 2003. január 1. <http://www.oszk.hu> – *Útmutató/Tezaurusz* és <http://mek.oszk.hu/adatbazis/thes.htm>

Tématerkép <http://www.neumann.haz.hu/thesaurus> (az interneten megjelenő honlapjának a pontos adatait kell itt megadni)

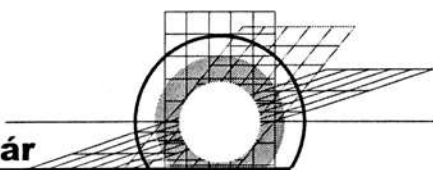
UNGVÁRY Rudolf: Az OSZK tezaurusza és a KÖZTAURUSZ. = Könyvtári Figyelő, Új folyam 11. (47.) köt. 1. sz. 2001. p. 11–40. <http://www.oszk.hu/kiadvany/kf/2001/1/ungvary.html>

WebKat.hu (az interneten megjelenő honlapjának a pontos adatait kell itt megadni)

XTMN Topic-Map Standard, ISO/IEC 13250:2000. XTM TzopicMaps Org. <http://www.topicmaps.org/xtm>

Beérkezett: 2004. V. 20-án.

Rendezvénynaplár



CIVR 2004 nemzetközi konferencia a képek és videók visszakereséséről

Dublin, 2004. július 21–23.

Szervező: Dublin City University,
Centre for Digital Video Processing

URL: <http://www.civr2004.org/>

Magyar Könyvtárosok Egyesülete 36. vándorgyűlése

Miskolc, 2004. július 29–31.

Szervező: Magyar Könyvtárosok Egyesülete
1054 Budapest, Hold u. 6.

Tel./fax: 311-8634

E-mail: mke@szk.hu

II. Rákóczi Ferenc Megyei Könyvtár
3530 Miskolc, Görgey Artúr út 11.

E-mail: rakoczi@rfmlib.hu

vandorgyules@rfmlib.hu

URL: <http://www.rfmlib.hu>

6. országos konferencia a digitális könyvtárakról

Pushchino (Oroszország), 2004. szeptember 29–október 1.

Szervező: Institute of Mathematical Problems of
Biology RAS

Institute Street, 4.

Pushchino, Moscow Region, 142290

Tel.: +7-0967 732408

Fax: +7-0967 330570

E-mail: rcdl2004@impb.psn.ru

URL: <http://www.impb.ru/~rcdl2004/en/>

Könyvtárak a 21. században

Salzburg, 2004. október 23–30.

Szervező: Salzburg Seminar
Leopoldkronstrasse 56-58., Box 129.

A-5010 Salzburg, Austria

Tel.: +43 662 83 9 830

Fax: +43 662 83 9 837

E-mail: info@salzburgseminar.org

URL: <http://www.salzburgseminar.org>

Ontológiák, adatbázisok és szemantikai alkalmazások nagy információs rendszerekben

3. ODBASE nemzetközi konferencia

Larnaca (Cyprus), 2004. október 25–29.

Szervező: University of Cyprus

URL: <http://www.cs.rmit.edu.au/fedconf/>

HISEC 2004. Biztonság az informatikában és a távközlésben. Nemzeti adatvédelmi és adatbiztonsági konferencia

Budapest, 2004. október 26–27.

Szervező: HISEC Konferencia Titkárság

Kugler Anita

Conference Tours Kft.

1055 Budapest, Kossuth tér 6–8.

Tel./fax: 353-0025

E-mail: conftour@mtesz.hu