

Szoftverhátér

A metaadatok kezelését lehetővé tevő szoftverek ma meglehetősen kezdetleges szinten állnak. A legtöbbjük nem több, mint a felhasználó adatainak korrekt bevitelét szolgáló űrlapok vagy a már strukturált információt megfelelő mezőkbe tördelő programok összessége. Úttörő kezdeményezésnek számít az OCLC CORC projektjében alkalmazott rendszer, amely besorolási állományokkal is rendelkezik, és megkísérli a tartalmi analízist is.

Karbantartás

A metaadatokkal dolgozó rendszereknek megoldást kell találniuk a források esetleges törléséből, áthelyezéséből, átnevezéséből eredő kérdésekre, mivel ezek egyik legfőbb jellemzője az instabilitás. A hagyományos könyvtáratomatizálási rendszerek jelenleg még nem tudják kezelni a problémát, kérdéses tehát, hogyan lehet elérni a megváltozott forrásokat, és ki legyen ennek a felelőse. A CAMEL projekt kettős módszerrel próbálja meg a probléma feloldását. Egyfelől évente egyszer e-mailben kéri fel a rendszerben szereplő források ajánlóit, hogy ellenőrizzék az általuk ajánlott forrásokat, az észlelt változásokat pedig egy listán regisztrálják. Mindez azon a feltételezésen alapszik, hogy az ajánlóknak továbbra is érdekükben áll elektronikus forrásaik elérhetőségének biztosítása. Másfelől a CAMEL rekordban tárolt (OASIS) rekordazonosító alapján szinkronizálják a két adatbázist, biztosítva ezzel az észlelt változások következetes átvezetését.

Összegezés

A projekt e kezdeti stádiumában még korai lenne végső mérleget vonni, de már most elmondhatjuk, hogy az egyetem karai szép számmal együttműködtek az elektronikus források felderítésében. A begyűjtött források karbantartásának sikereiről még nem lehet nyilatkozni. Metaadatokat elsődlegesen a kézi munka csökkentésére alkalmaztak. Ez azonban nem hozta meg a várt eredményt, mert az ajánlók által megadott metaadatok minősége tartalmi, mennyiségi, szintaktikai szempontból nagyon egyenetlen volt.

Az elmúlt években a metaadatok kezelésére alkalmas eszközök és az ezzel kapcsolatos szabályozás nagy fejlődésen ment keresztül. Jóllehet a DC szabvány támogatottsága könyvtáros és azon kívül eső körökben egyaránt növekvőben van, jelenleg még egy szabvány sem érte el felhasználói körében azt a kritikus szintet, amely jövőben betöltött szerepüket valószínűsíthetné. Napjaink könyvtárai még kevés gyakorlattal rendelkeznek az elektronikus források hozzáféréseinek biztosításában, de minden lehetőséget meg kell ragadni a metaadatok alkalmazásával nyerhető előnyök kiaknázására.

/BANERJEE, Kyle: Challenges of using metadata in a library setting: the Collection And Management of Electronic Links (CAMEL) project at Oregon State University. = Library Collections, Acquisitions, & Technical Services, 24. köt. 2. sz. 2000. p. 217–227./

(Lengyel Monika)

Kascade: új nyílt katalógus az internetinformációk eléréséhez

Az információkeresés az interneten is a nagy számok problémája előtt áll. Sem a keresőgépek, sem az üzleti alapú katalógusok nem bírnak a feladattal, de megoldást ígér az önkéntesek tömegeinek együttműködése. A *Kascade* egy interneten alkalmazható katalógus-csereformátumra épülő nyílt rendszer.

Az információkeresés és a dokumentumszám

Az információkeresés folyamatosan megújuló módszereket igényel. A dokumentumok száma mindenféle hordozón rohamosan növekszik, ez

gyorsan elavulttá teheti a legjobb ötleteket is. 1990 körül még győzni lehetett PC-n futó szövegfeldolgozó szoftverekkel. Ám az internet minden korábbi rekordot megdönt, a weblapok száma 3 millió felett van a világon. Ez a szám évente legalább megduplázódik, és az ütem egyre gyorsul. Az új dokumentumok születését csekélyebb sebességgel kíséri a régebbiek eltávolítása. Az információkeresés nagy kihívás előtt áll.

Az információkereső szoftverek fejlődése

1964 óta vannak visszakereshető bibliográfiai adatbázisok. A STAIRS szoftver kudarcát bonyo-

ultsága okozta, de az emberközelséget javító újabb fejlesztések sem tudtak lépést tartani a dokumentumok szaporodásával. Bármely keresővel kiábrándítóan sok téves találatot kapunk, és sok az elmaradt találat; minél nagyobb a dokumentumgyűjtemény, annál rosszabb a visszakeresés eredményessége. Az internet a dokumentumállomány speciális esete: megvan a hatalmas elektronikus adathalmaz, de sokáig tart, és gyenge eredményeket mutat az átfésülése. A problémát a méreten kívül az internet topológiája jelenti, ugyanis a lapok információtartalmát nemcsak a szöveg, hanem a linkek tömege adja. A keresőgépek fejlődése nem képes lépést tartani az igényekkel.

A felhasználók keresési ismeretei

Az információ-visszakereső szoftverek használatához az átlagfelhasználók már a kezdeteknél képzett specialisták segítségét igényelték. Bebizonyosodott ugyan a képzés szükségessége, de ezt a professzionális körön kívül nemigen ismerik el. Olyan módszerek kellene, amibe az átlagember beletanulhat.

Nyílt forrású szoftver

Az önkéntes munkával fejlesztett ingyenes, nyílt forráskódú szoftverek életképességére bizonyíték a Linux operációs rendszer és az Apache keresőgép sikere. Az alkotóknak az oktatás és a termék-támogatás hoz jövedelmet. A filozófia merőben új, olyan, mint a valódi amatőr sport, ahol az elismertség a lényeg.

Nyílt katalógusok

A nyílt katalógus az internetinformációk elérésére fejlesztett, emberi tevékenységgel épített tartalomjegyzék. Az ODP (Open Directory Project) a katalógizálendő szakterületek önként jelentkező szakértőiből építkezik. Az ODP-szoftver módszer-tannal és segédeszközökkel látja el a szerkesztőket. A fő témaköröket elismert főszerkesztők tartják kézben, ők választják ki az új szerkesztőket, szórják ki a szemetet stb. Az e módszerrel épülő DMOZ katalógust az AltaVista és a Google keresőrendszer magába integrálta. A jelszó: „Az emberek jobban csinálják.” Valóban, mára ez lett a világ legnagyobb és legjobb találati arányú webkatalógusa. A módszerrel további nemzeti katalógusok készülnek folyamatosan.

A nyílt katalógusok tervezési szempontjai

1. Elosztott szervezet

Az internethasználók elosztott közösséget alkotnak. Nincs központi irányítás, nincs állami kontroll, mindazonáltal szilárd a rend. A katalógusnak se legyen központosított jellege, hanem amennyire lehet, legyen „kint” az építőknél.

2. Több lehetséges keresési stratégia

A leírók szemantikai hálójá legyen rugalmas, az elektronikus könyvtári tezauruszokhoz hasonlóan. Egy terminushoz legyen megengedve több szinonima, és tartozhasson több generikus fogalomhoz is. Mivel a tudomány és a közérdeklődés szerkezete gyorsan változhat, a taxonómia legyen rugalmas. A keresőeszközök tegyék lehetővé, hogy a felhasználó a saját feje után különféle utakon eljuthasson ugyanahhoz a releváns információhoz.

3. Felhasználóbarát jelleg

A keresőeszközök használata legyen egyszerű; az alkotók is szeressenek vele dolgozni.

Kascade

A Kascade az egyik, a fenti elveken létrehozott nyílt katalóguskezelő eszközrendszer neve. Összetevői:

1. Szerkesztőprogram

A Kascade-ban a szerkesztők az internet rájuk eső részét úgy strukturálják, hogy az ezt leíró katalógusrészletet az ő szerverük tárolja. Az erre kialakított speciális szövegfájlformátum a DII, amelyben lehetnek más DII fájlokra mutató pointererek, ami a rész-katalógusok hálózatba kapcsolását teszi lehetővé. Mind e fájlok, mind kapcsolódásaik bármikor újjáalakíthatók vagy törölhetők, átfedésük sem tilos.

2. A böngésző interfész

Egy kis katalóguskövető ablak, felül egyszerű eszközsorral. A felhasználói interfész építését takarítja meg a katalógusszerkesztőknek.

3. Cache-szerver

A DII fájlok tükrözésének megkönnyítésére szolgál. Ezzel töltheti fel a szerkesztő a művét egy szerverre.

4. Elosztott szervezet

A Kascade a Gnutella módszereiből tanult, ahol az egész interneten elosztott rendszermegosztás áll fenn. A struktúrákat helyileg tartják karban.

/DUFOUR, M. et al.: Kascade: a new Open directory way of access to Internet information. = Information Services & Use, 20. sz. 2000. p. 63–72./

(Góth László)