

A kollektív hiperdokumentum

A szerző néhány hónapja a HUNGARNET Egyesület Könyvtáros Szakosztálya elnökségének felvetette egy, a digitális könyvtár témakörét felölelő hiperdokumentum elkészítésének lehetőségét. A projekt célja kettős: hasznos lenne közösen megismerni a témára vonatkozó kutatási és fejlesztési eredményeket, másrészt a projekt lehetőséget adna a hálózati közösségnek arra, hogy gyakorlatot szerezzen egy olyan technikában, amely láthatóan programmá válik a következő években.

A (világ-)Háló, angolul World Wide Web vitathatatlanul új korszakot nyitott az informatika történetében. A hipertextre (hipermédiára) alapozott szabad mozgás a számítógépes világhálózaton (az Interneten) hallatlan mértékben megkönnyítette az ismeretekhez való hozzáférést.

A folyamatot jól jellemzi Lesley West és Peter Murray-Rust közelmúltban megjelent cikkének címe (egyrészt a *Managing Information* 1997. májusi számában [1], másrészt a Hálón: www.venus.co.uk/vhg/aslib97.html#02): Steps towards the global linking of knowledge, azaz lépések az ismeretek globális összekapcsolása felé. Ezeket a lépéseket az épülőben lévő nagy szövegarchívumok kezelése is megköveteli. Akár ezeket tekintjük, akár a Hálón található információ- vagy dokumentummennyiséget, mi szakmabeliek mindannyian egyetérthetünk mind az említett közleményben leírt T. S. Elliot idézettel: „Ahol a bölcsesség található, elveszünk az ismeretekben, ahol az ismeret, ott elveszünk az információban” (A sziklák. 1938), mind John Naisbitt sokat idézett gondolatával arról, hogy szomjazunk az ismeretekre, és fuldoklunk az információkban.

A problémák megoldására valók a keresőrobotok és a hálózati ágensek, de mindannyian tudjuk, hogy ezek hatékonysága még nem megfelelő. Tisztában kell lennünk azzal, hogy a Háló és protokollja elsősorban a (multimédia) dokumentumok közötti bolyongást támogatja, és nem a tartalmi kapcsolatok felépítését, amely a hipertext- vagy hipermédia-kutatás tárgya.

Jelen cikkben a kollektív hiperdokumentum készítésének folyamatát és módszereit vizsgálom. Az első részben bemutatok a szakirodalomból néhány, a témára vonatkozó és a továbbiak alapjául szolgáló kezdeményezést. A második részben leírom a hiperdokumentum készítésnek egy lehetséges módszerét. Végül javaslatot teszek egy konkrét akcióra.

1. Kísérletek hiperdokumentum készítésére

1.1 A kollaboratív hiperglosszárúrium

A Lesley West és Peter Murray-Rust nevével fémjelzett *Virtual HyperGlossary (VHG)* projekt azt célozza, hogy eszközt adjon elosztott információ kezelésére sok szakterületre érvényesen, úgy, hogy több glosszárúriumot, azaz terminológiai adatbázist épít fel egyedileg képzett egységekből. Az egyes szakterületi glosszárúriumok függetlenül épülnek fel, de kapcsolatuk van egymással.

A VHG fejlesztése korábban a proteinszerkezet kutatásának szakterületén indult, és ennek kiterjesztéseként hirdették meg a projektet. A hálózathoz méltóan szabadon, kismértékű kuratori felügyelet alatt fejleszthető adatbázisok önmagukban is érdekesek, de témánk szempontjából az a fontos, hogy a glosszárúriumhoz dokumentumok is csatlakoztathatók.

A projekt leírása több címen megtalálható:

<http://gserv1.dl.ac.uk/CBMT/glossary/>,
<http://www.cryst.bbk.ac.uk/glossary/>,
<http://www.venus.co.uk/vhg/vhgnews.html>,
<http://www.seqnet.dl.ac.uk/CBMT/glossary/>.

A hiperglosszárúrium építését jól megérthetjük Murray-Rust – magyarra fordított – példájából, lásd árnyékolt rész.

Hogyan működik a HiperGlosszárúrium?

Bemutatunk egy vázlatot arról, hogy elvben miként működik a hiperglosszárúrium.

Tételezzünk fel hat szerzőt, akik sorban A,B,C,D,E,F és Z (a kurátor), és akik nem kell, hogy kommunikáljanak egymással.

Lássunk egy feltételezett történetet.

A megír egy glosszárúriumtételt az *alphahelix*ről.

B megír egy glosszárúriumtételt a *betasheet*ről.

E megírja saját közleményét a *helixek*ről mindenfajta jelölés (markup) nélkül.

F megírja saját közleményét a *globinokról* mindenfajta jelölés (markup) nélkül.

Eddig nincsen semmiféle kereszthivatkozás. Ezután különböző szerzők hozzáadnak ezekhez az elemekhez URL-eket, vagy a hipercikkek belsejében, vagy a glosszárümtételek URL-fejezetében.

D megírja saját közleményét az előrejelzésről, és benne hivatkozik mind az *alphahelixre*, mind a *betasheetre*.

C ír egy cikket a *Ramachandran plotról*, és ebbe beleír

hivatkozásokat a *alphahelix*től és a *betasheet*től a *ramachandran* felé,

hivatkozásokat a *alphahelix* és a *betasheet* felé a *ramachandrantól*.

Z (a kurátor) nem írt semmit, de úgy gondolja, hogy az *alphahelix* glosszárümtétel olvasója számára hasznos lehet azt tudni, hogy mi a *helix*. Ezért az *alphahelix* tételbe beír egy URL-t, amely a *helixre* mutat. Megjegyzendő, hogy erről sem A-nak, sem E-nek nem szükséges tudnia.

F rátaál a jelölő- (markup) szkriptre, és átküldi rajta minden hipercikkét. A szkript észleli, hogy a *globinokról* szóban van két hivatkozás az *alphahelix* felé, és automatikusan hozzáadja a csatolópontokat az *alphahelix*hez.

A úgy látja, hogy a két glosszárümtételt javítaná, ha egymásra hivatkoznának, és (anélkül, hogy meg kellene kérdeznie a kurátort) hozzáadja az oda-vissza hivatkozásokat az *alphahelix* és *betasheet* közé.

Ha az összes szerző 'ésszerű' döntéseket hozott, akkor nincs szükség arra, hogy a kurátor bármilyen szerkesztői változtatást tegyen.

Peter Murray-Rust

1.2 HYPERLINE – hipertextkörnyezet nagy, szöveges adatbázisok kezelésére

M. Agosti és munkatársai a hipertext típusú információfeldolgozás kérdéseivel foglalkoznak [2].

Az általuk javasolt információs architektúrát dokumentumgyűjtemények tartalmának jellemzésére, és a gyűjtemény szervezésére dolgozták ki elvi keretként és prototípusként. A referenciamodell kifejezi a dokumentumok szemantikus reprezentációját, és a felhasználó kérdését automatikusan megoldja. A struktúra többszintű. A legalsó szinten helyezkednek el a dokumentumok, amelyek gyűjteményéből a keresést a második szint strukturált fogalom (concept) halmaza – egyfajta strukturált index – könnyíti meg. Lehetőséget látnak a fogalmak további osztályba sorolására is, ez lenne egy további felső szint, de ezzel ebben a közleményben nem foglalkoznak. A második szint segédadatstruktúráját előállításának módja szerint osztályozzák, azaz lehet

- automatikusan a dokumentumokból előállított fogalmak halmaza,
- egy előre megadott tezauruszerű indexhalmaz,

és végül az olvasó által létrehozott leíró kifejezések gyűjteménye.

Ezek között hierarchikus rokonsági kapcsolatokat kell, illetve lehet kiépíteni.

A hipertextkörnyezet különböző funkciójú keresési, illetve olvasási helyzetekben teszi lehetővé a két sík közötti átlépést, így a fogalmak síkján elindulva közvetlenül eljuthatunk a keresett dokumentumhoz, majd vissza.

1.3 A kapcsolópontok kérdése

P. Thistlewaite az ausztrál parlament információs rendszerének (amely egy *nagy nyitott web*) automatikus menedzselésével kapcsolatban részletesen vizsgálja a kapcsolópontok kérdését [3]. Az egyik probléma ezek dinamikus viselkedése, azaz a kapcsolópontok menedzselése, frissítése. A másik, hogy a valós életben a kapcsolatok egy a többhöz jellegűek és kétirányúak. Ezeket a jelenlegi HTML nem támogatja automatikusan. A harmadik kérdés a kapcsolatok típusainak a problémája. Különböző taxonómiák léteznek, egyesek 27, illetve 80 típust is tartalmaznak. Ilyen sok típus kezelése illuzórikusnak tűnik. A szerző négy típust vizsgál részletesen: a strukturális, a hivatkozási (referencia), a szemantikus típust, valamint az esetleges relációkat, és az ezeknek megfelelő kapcsolópontokat.

Érdekes gondolat az, hogy a dokumentumokat nem kötelező HTML formátumban tárolni. A mai módszerekkel megoldható, hogy a kliens kérésére a szerver a dokumentumból előállítja a HTML változatot, és ebbe beilleszti a külön tárolt kapcsolópontokat. A tárolási stratégia is többféle lehet: a szövegben elhelyezett „*horgony*”-okkal, extenzionálisan (a karakterszámhoz kapcsoltnak), végül intenzionálisan (egy szóhoz, fogalomhoz, kifejezéshez kapcsoltnak). Mindegyik megoldásnak vannak előnyei és hátrányai.

A tipizált kapcsolópontok használatát vizsgálták kísérletileg Lisa Baron és munkatársai [4]. Egyetemi hallgatókkal végzett kísérleteik a hatékonyságot tekintve az áttekintő (browsing) szituációban nem mutattak ki szignifikáns különbséget, míg keresésnél (query) a jelzett, tartalomra utaló kapcsolópontok javították a keresési teljesítményt. A szerzők a kapcsolati típusokat szervező és tartalomalapú csoportra osztották, és az utóbbiban szemantikus, retorikus és pragmatikus alcsoportot különböztettek meg.

A szervező típusú kapcsolatok a tartalomjegyzékekben és a normál Háló-oldalokon található, alacsonyabb szinten mint a *dokumentum végére*, *következő oldal* stb. kapcsolópontok, amelyek a dokumentum felületi, szintaktikus makrostruktúráját jelzik.

A tartalomalapú kapcsolatok 12 elemet tartalmaztak az 1. táblázat szerint.

1. táblázat

Szemantikus	Retorikus	Pragmatikus
Hasonló	Meghatározás	Figyelmeztetés
Ellentétes	Magyarázat	Előfeltétel
Része/tagja	Folytatás	Felhasználás
	Illusztráció	Példa
	Összegzés	

A táblázatból látható, hogy a retorikus és pragmatikus kapcsolattípusok a tanuló típusú olvasást segíthetik, míg a szemantikus típusok a szabad bolyongásban lehetnek hasznosak.

1.4 Más megközelítések

Az irodalom – a nyomtatott és az elektronikus egyaránt – igen bőséges a jelzett témakörben. Az előző pontokban leírt projekteket és közleményeket azért mutattam be, mert számunkra érdekes elveket tartalmaztak. A hiperdokumentum készítése széles spektrumának két szélső megközelítését említtem még meg. Az egyik, amelyik digitális könyvtárban, archívumban gondolkodik. Ilyen projekt nagy számban található. A másik az, amikor egy dokumentum közös megírásáról, szerkesztéséről, közös menedzseléséről van szó. Ilyen pl. az Alliance nevű alkalmazási program (Decouchant, D.–Quint, V.–Salcedo, M. R.: *Structured Cooperative Authoring on the World Wide Web*), amely a <http://www.w3.org/pub/Conferences/WWW4/Papers/91/> címen található.

2. A hiperdokumentum készítésének egy lehetséges módszere

A hiperdokumentum struktúráját és a benne való mozgás funkcióit végeredményben az szabja meg, hogy mire akarjuk használni.

Agosti említett cikkében a következő hét keresési, illetve munkafunkciót említi és elemzi: szemantikus asszociáció, navigáció, szekvenciális olvasás, asszociatív olvasás, visszatérés, történet (bejárt út), a keresési stratégia fejlesztésének támogatása. Nem térve ki a részletekre és eltekintve a fenti felsorolás utolsó három, technikai jellegű munkafunkciójától, azt mondhatjuk, hogy a dokumentumban lehetőséget kell adni

- a szabad bolyongásra (asszociatív olvasás, navigáció),
- a racionális ismeretszerző vagy oktató olvasási pályára,
- és a keresés különböző fajtáira.

E három funkciót elégítené ki egy háromszintű struktúra, egy tartalomjegyzékszerű, egy fogalmi háló (a glosszáríum), és maga a dokumentumgyűjtemény. Az első vezérelten készül, a második és a harmadik szabadon építhető.

A hiperdokumentumban nem túl sok, a vázolt céloknak megfelelő számú kapcsolóponttípust kell megjelölni. A tartalomjegyzék szervező típusú kapcsolatokat igényel, de mivel a tanuló olvasási pálya elsősorban ezen a szinten vezérelt, a retorikai és pragmatikus kapcsolatok fontosabbak. A fogalmi hálóban, a glosszáríumban tartalmi, szemantikai kapcsolatokat kell létrehozni. Megfontolandó kérdés, hogy a dokumentumok között léte-süljön-e szemantikai kapcsolat.

A dokumentumarchívum – tekintettel arra, hogy különböző formátumban készült múltbéli dokumentumok is belekerülnek a hiperdokumentumba – vegyes formátumú lehet. Ezek kezelését a szolgáltatógépnek meg kell oldania. A dokumentumon belüli kapcsolatok feltárása végett célszerűnek látszik a kapcsolatok külön adatbázisban való tárolása.

3. Javaslat egy hiperdokumentum készítésére a digitális (elektronikus) könyvtár témakörében

A HUNGARNET Egyesület Könyvtáros Szakosztálya elnökségének néhány hónapja felvettem egy, a digitális könyvtár témakörét felölelő hiperdokumentum elkészítésének lehetőségét. A projekt célja kettős. Egyrészt a Hálón világosan látható, hogy másutt nagy erőket és fejlesztési kapacitásokat mozgatnak meg a digitális könyvtárak, archívumok létrehozása érdekében, és részletesen vizsgálják, kutatják annak részleteit. Magyarországon a kiváló MEK kezdeményezésen túl csak elszórt egyéni akciókkal találkozhatunk, amelyek gyűjtemények építésére irányulnak (ami jó), de a szisztematikus építkezésre kevés ismeretünk van. Ezért hasznos lenne közösen megismernünk a témára vonatkozó kutatási és fejlesztési eredményeket.

Másrészt a projekt lehetőséget adna a hálózati közösségnek arra, hogy gyakorlatot szerezzen egy olyan technikában, amely láthatóan programmá válik a következő években. A könyvtáros-informatikus közösségnek a projekt lehetőséget adna bekapcsolódni ebbe a folyamatba.

A hiperdokumentum készítésének legfontosabb alapelvei:

- Készül egy hierarchikusan szervezett tartalomjegyzék, egy hálózatosan összekapcsolható fogalomgyűjtemény, hiperglosszáríum formájában, és egy szöveges archívum.

- A háromszintű hiperdokumentum szabadon ellátható kapcsolópontokkal (link). A kapcsolópontok nem annyira a dokumentumokra, mint inkább azok belsejébe mutatnak. Ez az első fázisban feltételezi a dokumentumrészek kézi – esetleg szerző általi – strukturálását, a tartalomjegyzéknek és a fogalmi hálónak megfelelő kapcsolópontok elhelyezését. (Megjegyzendő, hogy a hipertext automatikus létrehozásának módszerein és eszközein is dolgoznak a szakma kutatói, lásd az *Information Processing & Management* ez évi 2., tematikus számát.) A lehetséges kapcsolattípusok listáját a program kezdetén össze kellene állítani, és dönteni tárolási stratégiájukról.
- Mindhárom szinthez bárki hozzájárulhat új tételekkel, akár meglévő dokumentumokkal, akár saját maga által készített tételekkel. Ha meg tudjuk oldani, akkor a tételek bevitelét a Hálón keresztül automatizálni kell.
- Meglévő dokumentumok esetén meg kell szerezni a szerző engedélyét a dokumentum itthoni tükrözésére, valamint belső kapcsolódási és kapcsolópontok elhelyezésére.
- Párhuzamosan kell magyar és legalább angol nyelvű dokumentumokat gyűjteni és összekapcsolni. A tartalomjegyzék és a fogalomgyűjtemény mindenképpen kétnyelvű kell legyen.
- A hiperdokumentum felügyeletét egy (legfeljebb háromtagú) kuratórium lássa el, csupán kevés

felügyeleti joggal, amely inkább időszerűvé váló módosításokra, esetleges törlésekre (ha egy közleményből frissített változat jelenik meg) és definícióütközések feloldására terjed ki.

A nem tisztázott szakmai és szervezési kérdéseket a program indulásakor a kuratórium és önkéntes résztvevők vegyék sorra. Elengedhetetlennek látszik programozó szakemberek bevonása is egy életképes modell elkészítésére.

Nyomtatott irodalom

- [1] WEST, L.–MURRAY-RUST, P.: Steps towards the global linking of knowledge. = *Managing Information*, 4. köt. 4. sz. 1997. p. 36–39.
- [2] AGOSTI, M.–GRADENIGO, G.–MARCHETTI, P. G.: A hypertext environment for interacting with large textual databases. = *Information Processing & Management*, 28. köt. 3. sz. 1992. p. 371–387.
- [3] THISTLEWAITE, P.: Automatic construction and management of large open Webs. = *Information Processing & Management*, 33. köt. 2. sz. 1997. p. 161–173.
- [4] BARON, L.–TAGUE-SUTCLIFFE, J.–KINNUCAN, M. T.–CAREY, T.: Labelled, typed lincs as cues when reading hypertext documents. = *Journal of the American Society for Information Science*, 47. köt. 12. sz. 1996. p. 896–908.

Beérkezett: 1997. IX. 22-én.

Rendezvénynaplár

Online Information '97

London, 1997. december 9–11.

Szervező: Learned Information Europe Ltd.
 Woodside, Hinksey Hill,
 Oxford OX1 5BE
 Tel.: +44 1865 388000
 Fax: +44 1865 736354
 URL: www.learned.co.uk

6. BOBCATSSS szimpózium

Budapest, 1998. január 26–28.

Szervező: Magyar Könyvtárosok Egyesülete
 1054 Budapest, Hold u.6.
 Tel./fax: 111-8634
 E-mail: mke@mail.c3.hu

Európai telematika: Az információs társadalom fejlesztése

Barcelona, 1998. február 4–7.

Szervező: Margot Donkersloot
 CONCORD Project
 CSC Datacentralen
 14, Avenue d'Hyppocrate
 B-1932 Woluwe St. Etienne
 Belgium
 Tel.: +32 2 714 7595
 Fax: +32 2 721 3357
 E-mail: conference@cse.de.be

CRIS 98 Európai Kutatási Információs

Rendszerek: Út az innovációhoz – konferencia

Luxembourg, 1998. március 11–13.

Szervező: Lieve Van Woensel
 European Commission, DG XIII-D.2
 EUFO 2191
 rue Alcide de Gasperi
 L-2920 Luxembourg
 Fax: +352-4301.34989
 E-mail: „Lieve.Van Woensel”
 @lux.dg13.cec.be