

Szöveg és tartalom az információs társadalomban

1. Amikor Varga Katalin doktori disszertációjának módszertani kézikönyv-változatával először találkoztam a kötelezpéldányunk átpörgetésekor, azzal az örömmel csodálkoztam rá, hogy milyen sok, a jelenlegi információkutatató, szöveg(eket) feldolgozó és reprezentáló munkánkra rímelő gondolatot lelek fel benne. A tartalmi osztályozás eszközei és gyakorlata iránti fogékonyság, valamint a tájékoztatásban szerzett tapasztalatok is arra ösztönöztek, hogy majd alaposan elolvaszam.

Amikor azonban a recenzió írására felkértek, más megközelítéssel kellett a mű átolvasásának nekilássak. Nemcsak a saját célú befogadás okán forgattam immár, hiszen a többi lehetséges olvasó számára kell bemutatnom – ugyanakkor az első benyomások is élnek, s a felhasználói célcsoport egyik tagjaként is értékelni fogom. Mindezek okán recenziómban hosszas tartalmi ismertetést rövidre fogott módszertani észrevételek követnek.

Varga Katalin egyik meghatározó tagja a ma még csupán „tizenkét apostolnyi” hazai könyvtár- és információtudományi PhD fokozatot elnyert könyvtári szakemberek csoportjának. Témaválasztása telitalálat: a tudássá nemesülő (feldolgozással tudássá nemesített) információözön tudatos válogatás eredménye, melynek közvetítése napjainkban és a közeljövőben is a könyvtárosok, információs szakemberek elsőrendű társadalmi feladata. A tudatos válogatás és a közvetített tudás felhasználhatóságának egyik alapfeltételét a jelenkori információs társadalomban kell megteremtenuünk a tudástársadalom számára. Mi ez az alapfeltétel? Az információkat tartalmazó dokumentumok, *kognitívumok*, *szövegek* – ezekkel a szakkifejezésekkel most barátkozzunk meg! – tartalmának elemzése, majd a szöveget reprezentáló szöveg, az eredeti tartalmat az információs szolgáltatásban helyettesítő *szurrogátum* előállítása.

2. A szöveg – információ – tartalom összefüggéseit a könyv első fejezete taglalja. Az információfeldolgozás és annak fejlődési irányai iránt érdeklődők figyelmét cím szerint „Az információtudomány és a nyelvtudomány kapcsolata”, továbbá „A tartalomelemzés” című alfejezetek ragadják meg. S valóban vita felett áll, hogy az információ reprezentálása és kereshetővé tételének nyelvészeti problémái napjaink fontos kutatási témái közé tartoznak. Az interdiszciplinaritás általános kortűnet, a két említett tudománynak is létezik közös területe: a szövegelemzés.

A szerző Varga Csabától idéz, mikor azt a tételt vezeti be, hogy az információs társadalom potenciálisan ismeret-társadalom. Miben különbözik ez a tudástársadalomtól? A tudástársadalom az értelmezett ismeretrendszer társadalma – potenciálisan pedig a tudásjel társadalma.

A tartalomról szóló alfejezet kulcsmondata: „...az információközvetítés a tudástársadalomban értelmezett ismeretek közvetítését kell, hogy jelentse.” Ebben a részben olvashatjuk a tudás 21. századi követelmények szerinti definícióját Horváth Tibor megfogalmazásában. Állításuk lényege az, hogy a dokumentum (az abban tárgyiasult tudás) a könyvtárba (az egyik lehetséges információs rendszerbe) bekerülő input, a tudás a könyvtárban zajló kommunikációs folyamat tartalmává, tárgyává válik.

Itt olvashatunk először a szöveg(ek) üzenetéről, amelyeknek információs értéke van az információs rendszer(ek) használói számára.

Mi a tartalom? – A szöveg jelentése, állítja a szerző, majd ábrával is erősíti tételét, amely arról szól, hogy a szöveg a tudás, a téma és a tartalom háromszögének középpontjában áll, és ezek az elemek elválaszthatatlanok egymástól.

Bár a szerző vizsgálatának tárgya szűkebben a szaktudományos szöveg, nem mulasztja el, hogy felsorakoztassa a szövegre vonatkozó összegyűjtött ismereteit. A szövegről szóló alfejezet a szövegnyelvészeti irányzataiban nem jártas olvasó intellektusának szóló kihívás: ebben a részben húsz idézett tételt olvashatunk a szöveg különböző megközelítésű definícióiként. A szerző törekszik arra, hogy tipográfiai is elkülönüljenek a tételei bizonyítása szempontjából legfontosabb elemek. Ezt a logikát követve recenzorként azzal az eszközzel élek, hogy az idézett tételek közül kiemelem a – könyvtári felhasználói szempontból – legfontosabbnak ítélt és továbbgondolandó megállapításokat. Ezek a következők:

- szöveg minden, a kommunikáció célját szolgáló emberi produktum, legyen az verbális vagy írott szöveg, vagy nem-verbális szöveg (pl. rajz, festmény, fénykép, film, zene, tánc stb.) – bármi, ami szándékolt üzenetet hordoz;
- az a szöveg, amely megfelel a szövegszerűség (ennek ismérveit később taglalja a szerző) hét ismérvének, amelyek sorban: kohézió, koherencia, közlési szándék, elfogadhatóság, hírérték, helyzetszerűség, intertextualitás;
- A szöveg egyrészt jelentéskomplexum, egységes egészet alkot, másrészt szerkesztett, globális struktúrájú, tehát részekre tagolható;
- A szöveg nemcsak részekből és viszonyokból álló összetett jelként, hanem értelmes cselekvésként is értelmezhető.

A szövegkutatás egyik fő feladata annak vizsgálata, hogy a szöveg miként fejti ki a hatását az adott kommunikációs szituációban. Miért fontos ez? Mert az információfeldolgozás célja, hogy kapcsolatot teremtsen az információ előállítója (az üzenet küldője) és az információ (üzenet) címzettje között. Az információfeldolgozás során arra kell koncentrálni, hogyan bontható ki a szövegből az üzenet, illetve (a felhasználói oldalról) mely információs igényekre mely létező szövegek tartalmazzák a választ.

A szöveg tematikus hálóját a témamegjelölő szavak, kifejezés láncolata biztosítja. A szövegszerűség kritériumai egy megvilágításban a kommunikációs környezet, a szöveg-világ, a szövegértékűség, míg más megközelítésben: a szövegvilág, szövegértékűség, szövegsemantika – szövegértelem, szemantikai progresszió, lineáris és globális kohézió, szövegpragmatika és szövegfókusz, szövegszerke-

zet, mikro- és makroszerkezet, a cím, fókuszmondat – tételmondat, kulcsszó. Ezek szerepét és jelentőségét nyelvészeti és információfeldolgozási szempontból is részletesen kifejti Varga Katalin.

A szerző az izotópia, az izotópiásík(ok) és izotopikus lánc, valamint a kvantálás, a (szöveg)réma fogalmát is az első fejezetben vezeti be. A hipertext avagy hiperszöveg szintén azért említődik ebben a fejezetben, mert nemcsak az információkeresés egyik eszköze, hanem a szöveg megjelenítésének egyik formája is – valójában a szövegszervezés egyik új módszere. (A recenzens itt megjegyzi, hogy a mű e részében az olvasó maga is megtapasztalhatja az intertextualitás jelenségét és jelentőségét: a fejezetben taglaltak megértéséhez ugyanis ismerni kell a kognitívum jelentését is: egy feltárási ismeretmennyiség ismérveivel együtt, ill. teljes vagy rész-szöveg, amelynek legalább egy önállóan meghatározható témája van stb.)

Ezután új alfejezethez érünk, amelyben az információtudomány szöveg-modelljéről esik szó, igen röviden. A szöveg-modell kialakítása azért szükséges, mert a tartalmi elemzés és feltárási erre építhető. A modell érvényes a szövegelemzésre és a szurrogátumkészítésre. A szöveg-modell szerkezete a kognitívumból (a feldolgozási egységből), a szövegből és a szövegvilágból áll.

A tartalomelemzésről szóló alfejezetben nyelvtudományi, információtudományi, kommunikáció-elméleti stb. alapdefiníciókat ismerhetünk meg, amelyek az információ-feldolgozó könyvtárost a tartalmi elemzés felelősségének felismeréséhez vezetik. E fejezet áttekintést nyújt a tartalomelemzés „hagyományos” (értsd: hetvenes-nyolcvanas évek-béli) alkalmazási területeiről. Ezt követi egy filozófiai alapként szerepeltetett idézet, és hozzá kapcsolódó igen gyakorlatias tanácsok az osztályozási jelzetek következetes alkalmazásáról. A tartalomelemzés elemeit (így a célját is?), valamint a típusait felsorolásként tekinthetjük át.

Ez lehetne a kötet egyik legizgalmasabb fejezete. Átolvasva viszont az a benyomás alakul ki, mintha egy logikailag ugyan helyes vázra egy sokforrású szakirodalmi szemlét feszítenének rá, amelynek kiértékelését rábízák az olvasóra.

A tartalomelemzési módszerekkel, továbbá a tartalomelemzés és az információtudomány kapcsolatával foglalkozó záró alfejezetek mégis közelebb visznek a módszertani kézikönyv céljához. A szerző rámutat a feltárási tárgyat képező médiumok rohamos fejlődésére, és ennek folytán a feltárási, az üzenetközvetítés módszereiben is elvárt megújulásra. A tartalomelemzés klasszikus módszerei az információs társadalom munkásai számára is számos tanulsággal szolgálhatnak. 3. „A szöveget reprezentáló szöveg” címet viselő második fejezetben elérkezünk a szurrogátum értelmének mélyebb kifejtéséhez. Magyarázatot kapunk arra, hogy miként viszonyul a szurrogátum az eredeti szöveghez, és hogyan helyezhető el a dokumentum a tartalmi feltárási szintjeihez képest. A kézikönyvnek ez a fejezete eligazít a tekintetben is, hogy melyek a szurrogátum szövegösszetevői (elemei), hogy milyen a viszonya a szövegszerűséghez kommunikációs, szövegszerűségi, szövegkörnyezeti, szövegpragmatikai és szövegszerkezeti szempontból. A fejezet (szerintem) legfontosabb mondatát szó szerint idézem: „...ahogy a szövegértést tanítani kell az iskolában, tanítani kell a szurrogátumok értelmezését is, hogy a polgárok valójában hasznosan tudják az információs szolgáltatásokat használni-

ni". Eljutunk tehát ahhoz a felismeréshez, hogy az információs társadalom tagjait fel kell készíteni, oktatni kell az információs rendszerek adta lehetőségek kiaknázására.

A szurrogátum minőségének mérhetőségét, a reprezentálandó szövegek adott információs rendszerben való helyettesítését illetően a szerző egy francia és egy német szerző szempontrendszerét ismerteti. Itt érdemes megemlíteni, hogy a kézikönyv doktori disszertáció-változatának vitáján többen is hitet tettek a tartalomelemzés és szurrogátumkészítés minőségbiztosítás körébe vonásának jelentősége mellett.

A tartalmi feltárást vezérlő paradigmák témáját számos szempont szerint tárja fel Varga Katalin. A hierarchikus osztályozás, a mellérendelő osztályozás (akár párhuzamos használatának) előnyei megjelennek a szurrogátum minőségében csakúgy, mint a szintaktikai indexelési eljárásokéi. A kontextusmegőrző indexelés c. alfejezetben kerít sort a szerző az általa igen alaposan ismert és alkalmazott PRECIS, illetve más prekoordinált indexelő rendszereknek, mint a Library of Congress Subject Headings-je és a Deep Structure Indexing System összehasonlító elemzésére. Megállapítja, hogy a szöveg kiemelt tartalmi elemeinek (személy- és földrajzi nevek, speciális jellemzők, felhasználói célcsoportok stb.) jelentős szerepe van a szurrogátum szöveggé szervezésében. (Értelmezésemben egy új kognitívum létrehozásában, amelyről készülhet egy új szurrogátum... és így tovább).

A szöveg tartalmi feltárássának sem nem osztályozó, sem nem szétszedő-összerakó módszere, hanem a kihagyás elvét alkalmazó eszköze a tömörítés. A módszer taglaló alfejezet tankönyvszerű pontossággal, lépésről lépésre tekinti át a szurrogátumkészítésben is szükségszerűen alkalmazott tömörítés folyamatát, ismérveit. A tartalmi feltárást a tartalmi keresést segítő eszközök, jelesül a számítógép adta lehetőségek alkalmazása, kiaknázása szempontjából is vizsgálja a szerző.

Végül a komplex tartalmi reprezentációt felkínáló nemzetközi szakirodalmi adatbázisok példáján keresztül kaphatunk képet a tartalmi feltáró eszközök kombinációjáról, amelyek eredménye akár komplex szurrogátumnak is tekinthető.

4. „Az automatikus szövegelemzés, automatikus osztályozás” című fejezet olvasásakor mélységes tisztelet fogott el a számítógépes nyelvészettel foglalkozó szakemberek iránt. Mennyi, az elődöktől felhalmozott tudás felhasználását és saját innovációt igényeltek az eddigi eredményeik is! Közben az ördög is dolgozott bennem, mikor megfogalmaztam a kérdést: szeretném tudni (utána is fogok járni), hogy a nemzetközileg legismertebb, legelfogadottabb szakirodalmi adatbázisok előállításánál milyen mértékben alkalmazzák a számítógépes tartalomelemzést, az automatikus kulcsszó-meghatározást, a szintaktikai és szemantikai elemzést, az automatikus szövegtömörítést (ideértve a mély és a teljes, avagy a részleges vagy felületes szövegelemzést, a transzformációt és a szöveggenerálást), valamint az automatikus osztályozást? Mert projektek vannak, voltak és indulnak, az ezekből leszűrt tapasztalatok viszik előre a szakmai világot, segítenek döntéseket hozni egyes fejlesztési területek mélyebb kiaknázásáról – vagy más fejlesztési irányoktól való elfordulásról.

A fejezet kifejtése közelebb vitt ahhoz, hogy megértsem a szimbolikus tudás összetevőinek kapcsolatát, felfedezzem a tudásbázis és a teauruszok rokonságát. Ez volt a pozitív hatás. A negatív hatást úgy érzékeltethetem, hogy jelzem az elbizonytalanodásomat: ha igaz az, hogy az elektronikus dokumentumok terjedésével növekszenek a dokumentumtermésben rejlő információk közötti eligazodás (legyőzendő) nehézségei, és az is igaz, hogy „a tartalmi feltárás az egyik legidőigényesebb és legdrágább munkafolyamat”, miért következik ebből, hogy az automatikus megoldások kutatása az igazán jó befektetés a jövőbe?

Idézek: „Arra is vannak próbálkozások, hogy a reprezentatív bekezdések kiválogatásával hozzunk létre tömörített szöveget. Így megőrizhető a kohézió... Mégis ezek a módszerek elég gyenge eredményeket mutatnak, az... automatikusan előállított referátumok kb. 46%-ban egyeznek meg a manuálisan készítettekkel.”

Mi olcsó és mi drága? Drága-e a képzett, a felhasználói igényeket jól ismerő szellemi munkás, aki az eredeti szöveget úgy osztályozza (mintha egy „címkézett irányított gráf” működne benne – lásd Ungváry R. előadása), referálja, tömöríti, készít belőle szurrogátumot akár, hogy az a leghívebben reprezentálja a szerző(k) eredeti mondanivalóját (a szöveg üzenetét)? Olcsóbb-e annak a nemzetközi ismertségű számítógépes szövegelemzést alkalmazó adatbázisnak az előállítás, amelyik bárki, illetőleg a felhasználók virtuális csoportjai számára tartalmaz relevánsnak mondott információkat? Drága-e a kutatás? Jó irányokban halad-e? Míg a hiperszövegek esetében szinte biztosak lehetünk abban, hogy az emberi agy működéséhez hasonló asszociatív ismeretszerzésre alkalmasak, az automatikus tömörítéssel kétségkívül együtt járó kihagyásról nem derülhet-e ki később (egy többedik generációs kutatás eredményeként), hogy éppen az tartalmazta a rejtett tudásjelet?

Egy példa: a műben hivatkozott kutatások jelentős része az USA-ban folyik. A Library of Congress élen jár a számítástechnika-alkalmazásban. Mégis, a Kongresszusi Könyvtár Kutatószolgálatánál tett látogatáskor azt jelentette ki az egyik, in-house adatbázist építő dokumentátor: „Hogyne, számos kereskedelmi adatbázist használunk. A felületes keresésekhez igen alkalmasak. A mi speciális igényű felhasználói körünk (szénátorok, képviselők és a szakértők) számára azonban mi magunk dolgozzuk fel a dokumentumok bizonyos körét, mert így érvényesíteni tudjuk az információk profiljaik és a dokumentumok közötti megfeleltetést.” Lehet, hogy drága a módszer, de valószínűleg nagyobb a pertinencia-értéke, mint 46%.

Ugyanakkor teljes rokonszenvvel és nagy várakozásokkal fordulok magam is a szövegelemzést és információkeresést támogató kutatások felé. Az elektronikus dokumentumokból/-ről való tájékoztatás növekvő igénye okán nagy jelentőségűnek tartom a természetes nyelvű szövegek tartalmi kategorizálását teauruszokkal és más tudásbázisokkal (pl. topic-okkal). A hírügynökségek szöveges és nem-szöveges (képi, hangzó- stb.) anyagainak, a televízió- és rádióműsoroknak az információk értéke a hagyományos sajtóéhoz fogható, hírértéke gyorsaságban néha meghaladja ezt. A tartalmi alapú keresésükre alkalmas számítógépes megoldásokra már láthatunk példákat.

Igéretes kutatást ír le a kézikönyv, amikor a GE egyik kutatócsoportjának „az információkeresés új generációja” elnevezésű projektjét mutatja be. A módszer

neve kiterjesztett tematikus keresés, lényege az, hogy a természetes nyelven megfogalmazott felhasználói kérdéseket az információs rendszer lefordítja a keresőnyelvre, és lefuttatja az adatbázisban. A maximum 30 referátumnyi listát a felhasználó összeveti az eredeti kérdésével. A relevánsnak jelzett referátumokkal a rendszer kiegészíti az eredeti keresőkérdést, és újra lefuttatja. Kísértetiesen emlékeztet a folyamat az emberi aggyal végzett keresés lépcsőire.

A 3. fejezetről szóló ismertetésben – úgy tűnhet – túltengenek a szubjektív elemek. Nem így van, megemlítek minden jelentősnek ítélt, e fejezetben taglalt szempontot, és egyben hozzáfűzök néhány kritikai észrevételt is.

5. Míg az első három fejezetben az elméleti alapvetést találhatja meg a mű olvasója, addig az „Új módszertani megoldások a könyvtári tartalomelemzésben” c. negyedik fejezet „igazi”, szerzői véleményt és tapasztalatot összegző kézikönyv az információ minőségi szelektálásához. A szerző alaptétele a minőségi kritériumok alkalmazása. Ennek érdekében teszi fel a következő kérdéseket, adott könyvtári szolgáltatások (információs rendszerek) mércéjére vonatkozóan:

- a) hitelesek-e a szövegelemzések és a szövegreprezentációk?
- b) más szóval: hiteles-e a tartalmi feldolgozás?

Ehhez a kérdésfelvetéshez és a válasz megkereséséhez szakmai és vezetési bátorság kell, szemléletváltás és megújult módszertan alkalmazása. A szerző az automatizálás egyértelmű igénye mellett manuális kvantálást javasol, valamint az információs rendszert építők befogadási igényességének és felelősségének, az emberi intellektusnak a jelentőségét hangsúlyozza. Varga Katalin továbbkérdez: mit ér az információs társadalom a tartalom feldolgozása nélkül? És kimond egy sokak által felismert igazságot: a hatékony információszolgáltatás feltétele, hogy valamennyi tudományterület rendelkezék megfelelő színvonalú információs rendszerrel.

A szövegszerűség hét kritériumával már találkoztunk egy korábbi fejezetben. Itt azt az aspektust hangsúlyozza a szerző, hogy a szövegértékűség hogyan válhat minőségi mércévé. A kritériumok vizsgálata során egy-egy szöveg értékelése is meg kell, hogy történjék. Ennek során fel lehet tární azt is, hogy a szöveg kisebb egységei önállóan is hordoznak-e értékes üzenetet. A tartalmi feltárás fő válogatási elvének és szempontjának kell lennie, hogy megfelel-e a szöveg tematikusan az információs rendszer elvárásainak. Itt észlelhetjük a „borítótól borítóig” feltárási gyakorlattal együtt járó „tévedhetetlenség” és a válogatási felelősség ütközését. A válogatott feltárást kiegészítheti a repertorizálás és a szöveg internetes kereshetősége. Az információs rendszer igényességéhez tartozik, hogy nem vezetheti félre a felhasználót, ehhez azonban a szurrogátum fel kell, hogy tüntesse a tárgyszavak aktuális kapcsolatrendszerét.

„Az izotópia, mint a kvantálás metronja” c. alfejezetben a pedagógia tárgyköréből válogatott szövegeken mutatja be a szerző a kvantálás gyakorlati problémáit. Leszögezi, hogy a témák (izotópiák) meghatározása, majd a kognitívumok meghatározására vonatkozó döntés a tartalmi feltárás leglényegesebb pontja. Ekkor dől el, hogy a szöveg a későbbiekben mennyire specifikusan lesz kereshető.

„A tartalom elemzése” c. alfejezetben a téma elemzésével: a kulcsszavak mennyiségi elemzésével, a tétel- és fókuszmondatok, a motívumok elemzésével,

a szöveg tematikus szerkezetének feltérképezésével foglalkozik Varga Katalin. Rámutat a tárgyszóképzés jelentőségére és felelősségére, a tárgyszavak aktuális kapcsolatrendszere kimutatásának fontosságára, a felhasználói célcsoport meghatározásának gyakorlati hasznára.

A szerző egyik korábbi állítása nyomán nem meglepő, hogy a magyar szakirodalom tartalmi reprezentációjának jelenlegi állapotát külön alfejezetben mutatja be. Tíz, eltérő szakterület irodalmát gyűjtő és közvetítő hazai köz- és szakkönyvtár, ill. felsőoktatási könyvtár információs rendszerét hasonlítja össze, és ehhez további három adatbázis elemzését is társítja. Ennek keretében tizenhat szakirodalmi adatbázist elemez és illet kritikával

- a) a tartalmi feltáró eszközök száma és félesége
- b) a feltáró eszközök szabályozottsága
- c) a tartalmi elemek jelölése
- d) a generikusság – specifikusság, továbbá
- e) a relációk – kontextus kifejezése szempontjából.

Az „Összegzés” c. zárórészt olvasva teljesen egyetértek a szerzővel abban, hogy „az automatizálás segítség, de nem megoldás”, és „a módszerek megújítása minőségi fejlődést eredményez”. Helyes a szövegnyelvészet és a tartalomelemzés módszereinek alkalmazása is.

6. Végül következzen néhány módszertani észrevétel. A kézikönyv szerkezeti felépítése logikus. A témakifejtés azonban több helyütt egyenetlen: tudományos mélységű feltárás vegyül tankönyvszerű didaktikával, szerkezetileg túlméretezett felsorolásokkal vagy alig kidolgozott alfejezetekre tagolással. A kézikönyvből hiányzik a hivatkozott függelék. A szerző előbb alkalmaz olyan fogalmakat, amelyekről azt is hihetné az olvasó, hogy elvárt a készség szintű ismeretük. Ugyanakkor később nagy alapossággal tárja fel e fogalmak jelentését. Ugyancsak megkönnyítené a kézikönyv gyakorlati alkalmazását, ha a példák és felsorolások következetes tördelésére (egybertartás, hierarchiaszintek) nagyobb gondot fordítottak volna – ezt nem a szerző hibájául rovom fel, de figyelmébe ajánlom következő műve vagy az itt ismertetett kötet újrakiadásakor. A 2. kiadáshoz érdemes lenne tárgymutatót szerkeszteni.

7. A kézikönyv olvasásához a főként magyar, angol és német nyelvű műveket felsorakoztató irodalomjegyzék alapos áttekintése után fogtam hozzá. Ekkor még nem tudtam, hogy valójában a mű (a szöveg) szövegvilágának egyik lehetséges elemével ismerkedem... Mire a kötet végére értem, nagyfokú tudatosság alakult ki bennem. Már látom a biztató jövőt: a szerzőktől (nemsokára) autoreferátum helyett autoszurrogátumot kérnek a kiadók. A tartalomelemzésre és a szövegprezentációra ez is egy lehetséges megoldás.

Haraszi Pálné

(Varga Katalin: Szöveg és tartalom az információs társadalomban. Módszerek és lehetőségek az információ minőségi szelektálására. Pécs: Pécsi Tudományegyetem Felnőttképzési és Emberi Erőforrás Fejlesztési Kara, 2005. /Humán szervező/munkaügyi/menedzser sorozat/)