

A szókészlet mint a könyvtár- és információtudomány központi fogalma

Michael Buckland

Michael Buckland: Vocabulary as a central concept in library and information science. (In: *Digital libraries. Interdisciplinary concepts, challenges and opportunities. Proceedings of the third international conference on the Conceptions of the Library and Information Science. Dubrovnik, May 23-26. 1999.* Ed. Tatjana Aparac, Tefko Saracevic, Peter Ingwersen, Pertti Vakkari. Publ. Zavod za Informacijske studije Odsjeka za informacijske znanosti, Filozofski Fakultet, Zagreb, Lokve, pp. 3-12.) tanulmányát a szerző hozzájárulásával **Murányi Péter** fordította.

Bevezetés

A szókészletet általában nem tekintik a (digitális vagy másfajta) könyvtárak központi jellegzetességének. A könyvtári diskurzusok gyakrabban foglalkoznak a gyűjteményekkel, költségvetéssel, a személyzeti kérdésekkel, épületekkel, használókkal, menedzsmenttel és egyéb praktikus témákkal. Ebben az előadásban azonban közelebbről megvizsgáljuk, mit mondhatunk a „szókészletről” a könyvtárakban, különösen a digitális könyvtárakban.

A „szókészlet” gyakori, elfogadott kifejezés a mindennapi beszédben, de mostanáig nem kielégítő a szerepe a könyvtár- és információtudomány területén. Itt ez valami idegen, egzotikus dolognak tűnik,

amely szokásos lakóhelyéről elvándorolt egy másik környezetbe. Olyan bevándorlónak tűnik, amely hasznos, de valahogy gyanús. Amikor a könyvtár- és információtudományban használják, a „szókészletet” közönségesen és ügyetlenül minősítik. Beszélnek „természetes nyelvi szókészletről” és „kontrollált szókészletről”. Ezek a kifejezések magukban rejtik a helytelen „nem természetes nyelvi szókészlet” és feltehetőleg a veszélyes „nem kontrollált szókészlet” lehetőségét. Úgy tűnik, mintha a szókészlet fogalmát csak félig fogadták volna el szakterületünkön. Tételezzük fel, hogy nekünk kell ávennünk és meghonosítanunk. Mit találnánk elmondhatónak róla?

Három megállapítást teszek majd a szókészlettel kapcsolatban. Ezeket a következtetéseket lehet magától értetődőnek, triviálisnak vagy hibásnak ítélni. Lehet, hogy az olvasó egyetért következtetéseimmel, de azt gondolhatja, hogy túlságosan kiterjesztem a „szókészlet” jelentését. Ha így lenne, el kell különítenünk a mostani jelenség tárgyalását attól a nagyon különböző vitától, amely a „szókészlet” szó megfelelő használatát érinti.

1. *Gazdasági szempontok.* A szókészlet központi kérdés a digitális könyvtárak költséghatékonysága szempontjából, s ezért megtérül az, amit befektetnek. Az egész világon nagyon sokat költöttek arra, hogy a hálózatokon keresztül elérhetővé tegyék a dokumentumtárakat, s ugyanígy sokat fordítanak az indexelésre, kategorizálásra és egyéb metaadatok létrehozására. A helyzet az, hogy egyre több nehézség jelentkezik a dokumentumtárak használói számára, mivel növekszik azon dokumentumtárak száma és aránya, amelyekben számukra nem ismerősek a metaadat szókészletek. Ennek megjósolható eredménye a csökkenő eredményesség a kiválasztásban. (A „kiválasztás” szót általános fogalomként használjuk, amelybe beletartozik a keresés, szűrés és visszakeresés.) Ezért bármilyen technika, amely segíthet az ismeretlen metaadatok használatában, akár úgy, hogy ismertté teszi azokat, akár úgy, hogy enyhíti az ismeretlenség következményeit, hatalmas befolyással lehet arra, hogy a dokumentumtárak és metaadataik létrehozásába való hatalmas befektetések milyen mértékben térülnek meg. (Ez a gazdasági megfontolás áll „A keresés támogatása ismeretlen metaadat szókészletek esetén” című projektünk mögött. Címe:

[www.sims.berkeley.edu/research/metadata/.](http://www.sims.berkeley.edu/research/metadata/))

2. *Az identitás kérdései központiak az információtudományban.* Egy olyan világban, ahol az identitás politikája központi kérdés, az identitás kérdései az információtudományban is központiak, a szókészletek szerepéhez kapcsolódó okok miatt.

3. *A szókészlet mint a digitális könyvtárak központi összetevője.* Az állítás itt először az, hogy minden szűrő- és keresőrendszer modellezhető halmazok (vagy „gyűjtemények”) egyik állapotból a másikba való transzformációjának sorozataként, másodsor, hogy a „szókészlet” megfelelő fogalom értékek választékára egy adott halmazban (vagy gyűjteményben).

Néhány példa

Gondolkozzunk el a Kongresszusi Könyvtári következő tárgyszaváról: GOD – KNOWABLENESS-HISTORY OF DOCTRINES – EARLY CHURCH, CA. 30-600 (ISTEN – MEGISMERHETŐSÉG – A DOKTRINÁK TÖRTÉNETE – KORAI EGYHÁZ – KB. 30-600), amit az LCCN 8005064-hez kapcsolnak. Ez alig természetes angol. Amikor olvassuk, úgy hangzik, mint egy távirat. Szintaktikai szerkezete fordítottja a szokásos angolnak, ahol a minősítő mellénevek és kifejezések szokásos módon megelőzik azt, amit minősítenek. Úgy hangzik, mint egy normális angol szöveg, amikor a szavakat visszafelé olvassuk. A tárgyszót korrekten alakították ki, de mégis azt mondhatják, hogy elég természetellenes. Valaki, aki érdeklődik a Mindenható megértésének története iránt, véleményem szerint nagy valószínűséggel nem gondol a GOD – KNOWABLENESS szavakra.

A másik példa egy olyan információs rendszerből származik, amely specializált szókészletet használ osztályozási és keresési célokra. A Census Bureau (Népszámlálási Hivatal) CD-ROM-on megjelentette és a <http://govinfo.kerr.orst.edu/impexp.html> cím alatt is elérhetővé tette az USA importjának és exportjának numerikus adatait. Ezek az adatok nagyon fontosak a kormányzat és az ipar stratégiai politikai döntései számára. Ennek ellenére az, aki a gépkocsigyártás iránt érdeklődik, s az AUTOMOBILE szóra keres, semmit sem talál. A CARS szóra

való keresés RAILWAY OR TRAMWAY STOCK-hoz vezet. Vannak adatok, de azok a PASSENGER MOTOR VEHICLES, SPARK IGNITION ENGINE alatt található.

Rakétákat egyre nagyobb mértékben használnak harci cselekmények során. Mennyit exportál az USA évente? A ROCKETS szót használva az export adataira keresve olyan árukategóriákat kapunk, mint BEARINGS, TRANSMISSION, GASKETS, MISC. Ha a keresést az egyes számú alakra korlátozzuk (ROCKET), ez három további kategóriát eredményez.

- PHOTOGRAPHIC OR CINEMATOGRAPHIC GOODS
- ENGINES, PARTS ETC.
- ARMS AND AMMUNITION, PARTS AND ACCESORIES THEREOF

Ez a legutolsó kifejezetten harci fegyvereknek az USA-ból való exportjával foglalkozik. A ROCKET szó csak a MISSILE & ROCKET LAUNCHES AND SIMILAR PROJECTORS (9301009050) alatt található meg, a nagyobb export kategóriában (GUIDED MISSILES, 9306900020) azonban nem. Ennek a szekciónak az általános kategóriája: BOMBS, GRENADES ETC (9306). Világos, hogy azoknak a kutatóknak, akik használni akarják ezt az adatbázist, szükségük van egy eszközre, amely hidat ver a szakadék fölött, amely az általánosan használt terminológia és az ezen adatok kategorizálására kialakított, nagymértékben specializált osztályozási rendszer között található.

Néha a hatások árnyaltak s váratlanok, mint a MELVYL-ben, a Kaliforniai Egyetem online könyvtári katalógusában végzett két tárgyi keresés esetében.

FIND XSU VIETNAM WAR

Keresési eredmény: 0 rekord

FIND XSU VIETNAMESE CONFLICT

Keresési eredmény: 4190 rekord.

A szókészlettel kapcsolatos problémák fokozódnak, amikor idegen nyelvekkel foglalkozunk. A pél-

da egy olyan keresés, amely a német hadseregnek az I. világháború során folytatott légi fotózásával foglalkozik, nevezetesen a DRACHENPHOTOGRAPHIE-val. A DRACHEN német szó a sárkányra vonatkozott, a keresés pedig a korszak műszaki szakirodalmából a sárkányokat használó katonai légi fényképezésre. Az azonban, amit sikerült megtalálni, nem volt releváns, mivel a németek kifejlesztettek egy pányvázott megfigyelési léggömböt, amely aerodinamikai volt. Aerodinamikai volt olyan értelemben, mint egy sárkány és sárkányballonként volt ismeretes, németül DRACHENFESSELBALLON, vagy röviden DRACHEN. Így a DRACHENPHOTOGRAPHIE ebben a speciális kontextusban nem a sárkányból való fényképezésre utalt, hanem egy megfigyelési léggömbre. Még ha ezt tudta is a kereső, sokkal eredményesebb lehetett volna, ha tisztában lett volna azzal, hogy az angolban az AEROSTAT fogalmat használják a megfigyelési léggömbök szakszavaként.

Nyilvánvaló, hogy különböző nyelvek, mint az angol, a kínai és a német, különböző szavakat használnak. Egy adott nyelven belül is az a helyzet, hogy különböző területek eltérő szókészletet használnak. Ezek a különbségek sokszor nagyobbak, mint várhatnánk. Nézzük meg azokat a tárgyszavakat, amelyeket a Library of Congress Subject Headings és a Medical Subject Headings (MeSH) ad a partvidéki szennyezéssel (COASTAL POLLUTION) foglalkozó dokumentumokra. A COASTAL és a POLLUTION tárgyi kulcsszóként nem vezetett eredményre, de lehetett rekordokat találni olyan módon, hogy a két szót a címben kerestük. A kapott tárgyszavak rangsorolt sorrendben:

LCSH: MARINE POLLUTION; COASTAL ZONE MANAGEMENT; WATER - POLLUTION; PETROLEUM INDUSTRY AND TRADE; BEACH EROSION; COASTS; BARRIER ISLANDS; COASTAL CHANGES stb.

MeSH: SEAWATER; WATER POLLUTION; BACTERIA; WATER MICROBIOLOGY; AIR POLLUTION;

ENVIRONMENTAL MONITORING; BATHING BEACHES; ENVIRONMENTAL POLLUTION stb.

Figyeljük meg a változatosságot, s azt, hogy milyen kevés közös van a két listában. A tárgyszavak elfogadhatóak, de kitől várható, hogy néhánynál többet kitalál közülük Könnyebb felismerni a pertinens fogalmakat, mint megjósolni, hogy mik lesznek azok. Ebben az esetben legalább három különböző szókészlet játszott szimultán: az LCSH, a MeSH és az enyém.

Nemcsak a szókészletnek a tárgyi indexelésben való használata nyelvi tevékenység, hanem a nyelvhasználat is elkerülhetetlenül a kultúrán alapszik. A szókészlet használata jelez valamilyen társadalmi helyzetet.

Néhány lehetőség a helyzet javítására

A Dewey-féle Tizedes Osztályozás „330” jelzete ECONOMICS (közgazdaságtan) jelentésében valamifajta furcsa megjelenési formájú szó. Annak felismerése, hogy az indexelés és az osztályozás nyelvi tevékenység, nem új. A metaadat rendszerekre DOKUMENTÁCIÓS NYELVEKKÉNT vagy META-NYELVEKKÉNT utalnak. A Dewey számok jelentéssel bírnak, ha ön ismeri a számokat. De a jelentés többé vagy kevésbé homályos addig, amíg a használat során az ismeret ki nem fejlődik. A használónak arra van szüksége, hogy saját fogalmait lefordítsa a DDC (TO) fogalmaira, valami angolt a „Dewey” szótárára.

Melvil Dewey ezt a formát a Relatív Indexében adta, amit hajlamosak vagyunk az osztályozáshoz készített függeléknek tekinteni. Dewey maga a rendszere legfontosabb részének tekintette. (Olding, 1996, 82-83.)

„Ez a betűrendes index, a rendszer legfontosabb jellegzetessége olyan kulcsszavakból (headingekből) áll, amelyek sokféle forrásból lettek gyűjtve, s a rend-

szert használói kívánatosnak találták őket... Az Index hasonló vagy szinonim szavakat ad, ... így egy intelligens személy biztosan megtalálja a megfelelő számot... A Relatív Index azonos élőfejeivel ... biztosítja, hogy a hasonló tartalmú könyv bármely területről, amely az osztályozóhoz kerül, ugyanarra a helyre kerül, s bármely olvasó keresi ezeket a könyveket, azonnal a megfelelő helyre utalják.” (Olding, 1996, 89-91.)

„A rendszer tárgymutatója (Subject Indexe) csontváz szótárkatalógus, tartalmaz mindent, ami nincs benne a „névkatalógusban”. Ahelyett hogy könyvcímeket adnánk minden tárgyszó (head) alatt, a szám utal azokra a címekre egyszerűen és közvetlenül... Ezért egyesítjük a szótár- és a szakkatalógusok előnyeit, nem úgy, hogy összekeverjük őket s így elvész az egyik egyszerűségét és a másik kiválóságát, hanem igazán használva mindkettőt, mindegyiket a maga érdemével. A címekből csak egy halmazra van szükség, mivel a mi osztályozási számaink elérhetővé teszik őket mindkét katalógusunk számára.” (Olding, 1996, 104.)

A Dewey által használt megreformált helyesírás emlékeztet minket arra, hogy a „természetes” nyelv konvenciókon alapszik, s ellenállna a „teljes szövegű” keresésen alapuló visszakeresésnek, amely a szóhasználat konvencióitól függ.

Ebben az esetben mindenképp szükségesnek és érdekesnek látszik Dewey eredeti szövegének közlése is:

„This alphabetic index, the most important feature of the sistem, consists of headings gathered from a great variety of sources, as users of the sistem hav found them desirabl... The Index givs similar or sinonimus words,... so any intelijent person wil surely get the ryt number... The Relatív Index, with its cach-words... insures that books on the same faze of any subject cuming before the clasifyers shal be asynd to same place, and that any reader seeking these books shal be referd instantly to that place.” (Olding, 1996, 89-91.)

„The Subject Index of this sistem is a skeleton dictionary catalog, covering everything not covered by the 'name catalog', Insted of giving book titles under each hed, the number refers to all those titles simply and directly... We therefore unite advantajes of dictionary and clast catalogs, not by mingling them and so losing much of the simplicity of one and as much of excelence of the other, but by realy uzing both, each with its own merits. Only one set of titles is needed, for our clas numbers make this available for both catalogs.” (Olding, 1966, 104)

Több szókészlet

Minden keresőrendszer több szókészletet foglal magában. Még abban a legprimitívebb esetben is, amikor szerkesztetlen szövegeket keresünk szerkesztetlen kérdésekkel, legalább két szókészlet van:

1. A keresett dokumentumok szerzőjének(inek) szókészlete(i) és
2. A kereső szókészlete.
A működő kiválasztó rendszerek esetében a szókészletek száma valószínűleg sokkal nagyobb. Egy online könyvtári katalógus például általában további hármat tartalmaz. A két, már említett mellett az alábbiakat:
3. A katalogizáló szókészlete, amelyet a dokumentumok reprezentációinak létrehozására használ, módosítja, helyettesíti és/vagy kiegészíti a szerző szókészletét;
4. „See”, „See also” (Lásd, Lásd még) és más szindetikus (összekötő) szerkezetek módosítják, helyettesítik vagy kiegészítik a katalogizáló szókészletét; és
5. A kereső szókészlete, ahogy a keresőkérdést kialakította.

Pontosan azért, mert több szókészlet is van, mindig megvan a lehetősége a hibás párosításnak szókészletek közti átmenet során, vagyis egy disszonanciának a jelentésben. Ha a kereső A-ról kérdez

és a szerző B-t írt, lehet, hogy ugyanazt a jelentést különböző módokon fejezik ki (szinonimák), vagy lehet, hogy mindketten A-t írtak, de különböző jelentéssel (homográfok).

A közbeeső szókészleteket (a katalogizálóét, a szindetikus (összekötő) szerkezetet, a kialakított kérdést) úgy lehet tekinteni, mint ami normalizálni akarja a fogalom használatát s így helyesbítené az eltéréseket. A katalogizáló által adott subject heading (tárgyszó) helyesbíti a szerző által adott címet azáltal, hogy a dokumentum témáját szabványosított szókészlettel reprezentálja. Egy tapasztalt kereső tudja, hogyan kell módosítani a saját vagy mások által meghatározott igényt olyan fogalmakkal, hogy a rendszer használhatóan válaszoljon.

A szókészletnek annyi újrareprezentálása van, ahány átmenetre sor kerül a szókészletek között. Minden újrareprezentálás lehetőséget ad a kiválasztó és a dokumentum közti disszonanciák helyesbítésére, de ugyanakkor disszonanciákhoz is vezethet. A vietnami konfliktus példája esetében a katalogizáló szókészlete különbözik mind a szerzőkétől, mind a keresőkétől, ezzel disszonanciát és problémát okozva, amely egy újabb átmeneti szókészlet megalkotását teszi szükségessé a javítás érdekében. Ebben az esetben egy kereszthivatkozás VIETNAM WAR USE VIETNAMESE CONFLICT (VIETNAMI HÁBORÚ LÁSD VIETNAMI KONFLIKTUS) szüntetné meg az eltérést. Ennek alternatívája az, hogy egy közvetítő kereső (ember vagy gép) tudhat eleget ahhoz, hogy terminológiai változást hozzon létre a rendszer szókészletéhez igazodva.

Mivel a szókészletek közti csaknem minden átmenet lehetőséget kínál arra, hogy disszonanciát szüntessen meg vagy keltsen, a többértelműség, amely zavarossá teszi a jelentést azzal, hogy egy jelentést kettőként vagy többként mutat be, szintén valószínű eredmény. A többértelműség bevezetésének lehetősége megsokszorozódhat, mivel minden további átmenet egy újabb szókészletbe, növeli a többértelműségek és hibák lehetőségét.

A több szókészletről úgy beszéltünk, mintha mindegyik egyértelmű és homogén lenne, de bármelyik szókészlet belső okokból is következetlen lehet, akár az egyén használatának variációin keresztül, vagy mert több szereplőt vontak be, mindegyiket a saját személyes használatával (szinkron változat). Azt sem szabad feltételeznünk, hogy bármely egyén vagy csoport következetes szókészletének használatában az idők során egy változó világban (diakron változat), vagy hogy az idők folyamán bekövetkező változások következetesek minden résztvevő szempontjából. A szókészletet arra használják, hogy jelezzék a jelentést, de a nyelv instabil, dinamikus. A szavak terjeszkedőek abban értelemben, hogy új jelentéseket vehetnek magukra. A jelentések különbségei, kétértelműségei és bizonytalanságai, ha egyszer felismerték őket, feloldhatóak mások jelentéseinek vizsgálatával. A párbeszéd kihozza a különbségeket. A különbségek változtatást generálnak. Ebben az értelemben a szavaknak saját életük van, új jelentéseket vesznek fel függetlenül attól, hogy a szótárszerkesztők mit írtak. Továbbá a jelentések összetettek: Létezik mind a denotáció (jelentés, értelem), amire a szó utal, és a konnotáció (mellékjelentés), indirekt nüanszok és asszociációk, amelyek színezik a felfogott jelentést.

Forma és jelentés

Az előzőkben utaltunk a szókészletek közti meghatározatlan „disszonanciára”. A digitális technika használata oda vezet minket, hogy a szókészletre mint karaktersorozatból álló fogalmakra gondoljunk, amelyeket lehet kezelni, de a szavak jelentése szubjektíven és szituációban formálódik, a szókészlet használata társadalmi.

Van egyfajta dualitás a szavak formája és jelentése között: a szó ugyanazon formájának különböző lehet a jelentése különböző szókészletekben, és ugyanaz a jelentés kifejezhető különböző formájú

szavakkal. A forma fontos az információs rendszerekben, mert a technika a szavak fizikai jellemzőivel dolgozik, nem jelentésükkel. A szoftver karaktersorozattal dolgozik, lexikai egységekkel, nem fogalmakkal. Amikor a „fogalom” szót használják a számítógépes visszakereséssel kapcsolatban, ennek használata metaforikus, általában néhány algoritmikus kapcsolatra utal lexikai egységek között. Már mindössze két szókészletből származó bármely fogalompár közötti kapcsolat is minimálisan négy lehetőséget tartalmaz:

Ugyanaz a forma, ugyanaz a jelentés: Ugyanaz a szó.

Ugyanaz a forma, különböző jelentés: Homográfia

Különböző forma, ugyanaz a jelentés: Szinonima

Különböző forma, különböző jelentés: Különböző szavak

Az „ugyanaz” itt gyakorlatban azt jelenti, hogy azonos módon elfogadható valamilyen célra, nem pedig azonosságot szigorú értelemben. A tökéletes szinonimák ritkák az angol nyelvben. Általában van némi különbség a jelentésben.

A jelentés cseréje és az identitás kérdései

Az egyik szókészletből a másikba való jelentésvesztés nélküli átmenet szigorú megszorításokat igényel. A jelentésvesztés nélküli transzformáció formális feltétele az, hogy egyértelmű és visszafordítható ekvivalenciának kell lennie minden transzformációban. Ez valószínűtlen, kivéve azt az esetet, ha teljesen zárt rendszerről van szó, amelyben minden kifejezésformát egyértelműen előre meghatározták. Ez a feltétel nem érhető el, amikor emberek is bekapcsolódnak, így a több szókészlet jelenléte a kiválasztó rendszerekben a lehetséges jelentések többségét tükrözi, ez létezik minden jelentés cseréje esetén. A kiválasztó rendszerek párbeszédes interakcióként funkcionálnak a szókészletek között,

amelyek legalább két irányban működnek. A kereső szókészletét próbára teszik a kiválasztó rendszer szókészletének korlátai, megkötései; a dokumentumok szókészlete is módosított, „kontrollált”. Mindkettőt a konvergenciát megcélözva alakítják át annak érdekében, hogy elérjék a jelentés megegyezését.

Létezik azonban a szókészletek bővülése is, ahogy a kereső, a katalogizáló, a kereszthivatkozások készítője folyamatosan alkalmazkodik az új kihívásokhoz, változó körülményekhez és az új szókészlet-hez. *Berman: Prejudices and Antipathies (Előítéletek és antipátiák)* című tanulmányában bőségesen szolgáltat a Kongresszusi Könyvtár Subject Headingjeinek (tárgyszórendszerének) megváltoztatására vonatkozó javaslatokat (Berman, 1971).

Az identitás kérdései központiak a társadalmi életben. *Edward Tylor* klasszikus meghatározása: „*A kultúra és a civilizáció, széles etnográfiai értelemben az a komplex egész, amelynek része a tudás, hit, művészet, erkölcs, jog, szokás és egyéb képességek és viselkedésmódok, amelyeket az ember a társadalom tagjaként sajátít el.*” (Tylor, 1871).

Az identitás a különbséggel magyarázható meg. Semmi sem határozható meg kielégítően önmaga fogalmaival. „Én énmagam vagyok,” azt jelenti, hogy nem vagyok te. Jobban hasonlítok rá, mint másra, inkább egyetértek ezzel a személlyel, mint azzal. Jobban szeretem az egyik csoport hitét, művészetét, erkölcsét, jogszokásait és szokásait, mint egy másikéit.

A kategorizálás és a szókészlet a kategóriák reprezentációjához alapvető az ilyenfajta megkülönböztetéshez. A megkülönböztetés és a kategorizálás az alapja a társadalmi világnak és alapja a kiválasztó rendszereknek. Egy szerep eljátszására a könyvtárosok, tanárok és mások bekapcsolódnak a kultúra átadásába, az indexikális kapcsolatok formálásába, beleértve az *ez és az, köztünk és köztük* közötti megkülönböztetést és jellemzést, a könyvtár- és információ-tudományt olyan műveletek és kérdések

részévé téve, amelyek központiak a társadalom számára. Ez egy további ok arra, hogy a szókészletek közötti átmenet („mapping, leképezés”) a könyvtár- és információtudomány központi kérdése, vagy legalábbis annak kell lennie.

A szókészlet meghatározása a könyvtár- és információtudományban

Ha azt gondoljuk, hogy a szókészlet bármilyen fontossággal bír a könyvtár- és információtudományban, ahogy azt állítottam, akkor talán többként kell kezelnünk a „szókészlet” fogalmát, mint egy nyelvészeti tanulmányokból kölcsönzött szót. Ha a szókészlet fogalmát be akarjuk vezetni a könyvtár- és információtudományba, akkor definíciót kell keresnünk arra, vagy legalább meg kell értenünk, mit értünk „szókészlet” alatt, ami intellektuálisan kielégítő módon megfelel a könyvtár- és információtudomány területének. Ez a konferencia a digitális könyvtárakkal foglalkozik. Mi lehetne hatékony definíció a „szókészletre” a digitális könyvtárakkal való összefüggésben? Hogy erre a kérdésre válaszolhassunk, először a szó szokásos jelentéseit kell megfontolnunk, utána a digitális könyvtárak természetét.

Az *Oxford English Dictionary* (1989, 19. köt., 721.) négy definíciót ad a „vocabulary” szóra.

1. Szavak gyűjteménye vagy listája jelentésük rövid magyarázatával.
2. Egy adott személy, osztály, foglalkozás vagy hasonló nyelvének értéktartománya.
3. *Az egy nyelvet képező szavak összessége.*
4. *Képletes értelemben.* Artisztikus vagy stilisztikus formák, technikák, mozdulatok stb. halmaza, amelyek elérhetőek egy adott személy stb. számára.

Az a mögöttes jelentés, hogy a „szókészlet” a jelentés különböző kifejezéseinek felsorolását, a repre-



zentációs formák repertoárját jelöli. A nyelvészetben a „type” (típus) és a „token” (jelölés) szavakat használják, ahol egy adott szó minden esete jelölés, s minden különböző szó típus. A különböző típusok kiterjedését vagy repertoárját nevezhetjük szókészletnek. Így az is megfelelő lehet, ha indexfogalmak vagy tárgyszavak értéktartományát vagy repertoárját nevezzük „szókészletnek”.

De a könyvtár- és információ-tudományban a jelentés kifejezésére használt fogalmak gyakran vagy a természetes nyelv eléggé természetellenes adaptációi (pl. GOD – KNOWABLENESS...), vagy mesterséges jelölést alkalmaznak, mint például azt, hogy 330 jelenti a közgazdaságot. Valójában a jelentés reprezentációjának ilyen rendszerei a szakterület specialitásai. Az ilyen leíró rendszerek alkalmazása valamifajta nyelvi tevékenység, s régóta használják rá a DOKUMENTÁCIÓS NYELVEK vagy META-NYELVEK fogalmát a metaadatok nyelvének jelentésében. Ezért nagyon megfelelő a könyvtár- és információ-tudományban kiterjeszteni a „szókészlet” használatát úgy, hogy utaljon tezauszban elfogadott fogalmak, osztályozási rendszerben használt számok és bármely kategorizálás kódja értéktartományára vagy repertoárjára. A digitális könyvtárakra speciálizált szövegtérnyezetben sem látszik ok arra, hogy ne használjuk a „szókészletet” olyan technikai fogalomként, amely bármely MARC mező vagy más metaadat mező kiterjedését vagy repertoárját jelzi.

A „szókészlet” technikai fogalomként való kezelése a könyvtár- és információ-tudományban bármely metaadat terjedelmének jelentésében újabb lehetőségeket nyit, mivel a digitális könyvtárak rendszerének teljes szerkezete reprezentálható fogalmak (term-ök) halmazaként (vagy gyűjteményeként) és egyik halmazból a másikba, egy abból következőbe való átmenetként. Csak kétféle átmenet tűnik lehetségesnek:

1. Transzformáció, amikor katalóguscédulák halmaza származik könyvek halmazából, vagy ami-

kor vektorok halmaza származik digitális szövegekből;

2. Szétosztás (vagy újra)rendezés), mint például amikor a cédulákat betűrendbe rendezik, vagy amikor rekordok részhalmazát kiválasztják egy visszakeresett halmazból.

Ahogy *Christian Plaunt*tal mostanáig sikerült meghatároznunk, minden digitális könyvtári szerkezet, valójában minden szűrő- és keresőrendszer modellezhető ilyen módon halmazok sorozataként és (származtatott) halmazokba való átmenetként. A digitális könyvtárak szerkezetei halmazok hierarchiái. Léteznek dokumentumtárak hálózatai, mindegyik dokumentumok gyűjteményét tartalmazza, tipikusan betűkből álló szavakból képzett bekezdésekből épül fel. A metaadatok mezőkből állnak, amelyek gyakran tartalmaznak almezőket és így tovább. Ennek a modellnek a részleteit és a digitális könyvtárakra vonatkozó következményeit máshol dolgoztuk fel (Buckland és Plaunt 1994; Plaunt 1997, Buckland és Plaunt 1997). Itt az érvényes, hogy ha a digitális könyvtárakat fogalmak halmazának tekinthetjük, s ha fogalmainkat meghatározuk a könyvtár- és információ-tudományban, a „szókészlet” jól meghatározható bármely halmaz értéktartományaként. Ha így lenne, a „szókészlet” azonnal a szakterület központi technikai fogalmává válna.

Összegzés és következtetések

Az információkeresésben a „szókészlet” kifejezés gyakran utal a természetes nyelv stilizált alkalmazására indexelő fogalmak kialakítása érdekében. A közelebbi vizsgálat a szókincset hatékony és átható elképzelésnek mutatja, mert a digitális könyvtárak több nyelvet s ebből következően több szókészletet alkalmaznak. A jól ismert könyvtári katalógussal kapcsolatos minden tranzakció legalább öt különböző szókészletet foglal magában: a szerzőkét, az

indexelőkét, a szindetikus (összekötő) szerkezetet, a keresőét és a megszerkesztett kérdését. Az indexelés, akár „természetes”, akár mesterséges jelöléssel, leíró ezért nyelvi tevékenység. Hagyományos és megfelelő, ha a metaadat rendszerekre DOKUMENTÁCIÓS NYELVEKKÉNT utalunk.

Ha a „szókészlet” szót szokásos értelmében vesszük, vagyis ami a használt, különböző szavak terjedelmét vagy repertoárját jelöli, akkor ésszerűnek tűnhet az is, hogy bármifajta metaadat, például bármely MARC mező terjedelmeként is használjuk. Ha egy szót technikai fogalomként akarunk használni bármely területen, például a könyvtár- és információtudományban, legjobb, ha ennek a területnek jelentéséről megegyezés születik. Ésszerű és hasznos lenne a „szókészlet” használata bármely halmazban (vagy gyűjteményben) található szavak terjedelmének jelölésére használni, beleértve minden metaadatot, ez nagy általánosítást biztosít, amikor az általunk javasolt funkcionális modellel kapcsolatban használják.

A szótárak közti leképezés, egy adott szókészletben megtalálható fogalomtól egy másik szókészletben annak megfelelő fogalomhoz vagy fogalmakhoz egyre inkább szükséges, ahogy a kényelmes hozzáférés egyre több és több dokumentumtárra, és további, kevésbé ismert metaadatokra terjed ki.

A szókészlet alapvető a digitális könyvtárak gazdagságossága szempontjából, mivel az ismeretlen terminológia gátolja a hatékony keresést. A szókészlet azért is fontos, mert központi jelentőségű az identitás kérdéseiben, amelyek ugyanakkor központi jelentőségűek a társadalom számára. A szókészlet, ha technikai definíciót kap a könyvtár- és információtudományban, értékek választéka vagy terjedelme egy halmazban, a digitális könyvtárak szerkezetének és használatának központi jellemzője.

Van másik megfontolás is. Ez leegyszerűsítés, de mégis úgy gondolom, hogy a könyvtár- és információtudomány eszméjének történelmi fejlődését jobban megérthetjük, ha két különböző hagyomány fo-

galmaiban gondolkodunk, amelyeket „dokumentum hagyomány”-nak és „formális hagyomány”-nak nevezek.

A „formális hagyomány”-ba azokat a technikákat és technológiákat értem bele, amelyek a logikán és algoritmuson alapulnak: lyukkártyák, digitális számítógépek, adatfeldolgozás, számítástechnika, mesterséges intelligencia és az információkeresés történeti hagyományai, ahogyan az ACM SIGIR találkozóin tükröződnek. Ez a formális hagyomány az, amely olyan sokat tett azért, hogy konferenciánk témája – a digitális könyvtárak – lehetségessé váljanak. De ez a hagyomány definícióktól és megbízható műveletektől függ, és szembenáll az emberi nyelv és az emberi viselkedés változékonyságával.

A „dokumentum hagyomány”-ba helyezném a dokumentumszolgáltatás olyan történeti gyakorlatát, mint a bibliográfia, könyvtárosság, levéltárosság és rekordok kezelése. Ebben a hagyományban a dokumentumokkal jelentéssel bíró tárgyként és több célt szolgáló használatuk jelentésében foglalkoznak: gyakorlati használhatóság, oktatás, rekreáció, műveltség és különböző társadalmi szolgáltatások. Ennek a hagyománynak van meghatározott logikája: az következik belőle, hogy a szakma gyakorlata kiterjed bármely jelentéssel bíró tárgyra, bármilyen formában, amely (potenciálisan) mindent tartalmaz, ami segíti a tudást, továbbá annak megértése is (következik belőle), hogy a dokumentumoknak köze van a tudáshoz, jelentéshez, tanuláshoz, leíráshoz, nyelvhez és kétértelműséghez (Buckland, 1997). Ebből az következik, hogy a könyvtár- és információtudomány eszméje nem lehet teljes, ha nem foglalja magában a kulturális tanulmányokat is, s végül is a könyvtár- és információtudomány érett, jól fejlett eszméjének szükségszerűen elő gyökerekkel kell kapcsolódnia a humán tudományok és a kvalitatív társadalomtudományok kérdéseire.

Úgy tűnik, hogy a két hagyomány alapvetően inkompatibilis, mert alapvetően különböző bázisról

indultak. Egyiket sem választhatjuk azonban kizárólagosan, ha hatékonyak és gyakorlatiasak is akarunk lenni. A szókészlet azonban mindkét hagyomány központi része. Különböző módokon mindkettőnek foglalkoznia kell a szókészlet kérdéseivel, s ez szolgálhat valamilyen találkozási helyként. A szókészlet kérdése végső következtetésem szerint fontos ennek a konferenciának a számára, mert a szókészlet természete és szerepe központi a könyvtár- és információtudomány bármely hitelt érdemlő koncepciója számára.

Köszönetnyilvánítás

Ennek az előadásnak az eszméi nagy mértékben a szerző Ron Dayvel és Christian Plaunttal, valamint „A keresés támogatása a kevésbé ismert metaaadat szókészletek esetén” című projekt (DARPA Contract N66001-97-C-8541; AO = F477; <http://www.sims.berkeley.edu/research/metadata/>) közreműködőivel való együttműködéséből származnak.

Irodalom

- BERMAN, S.: Prejudices and Antipathies. A Tract on the LC Subject Headings Concerning People. Jefferson, NC: McFarland. First published 1971. Scarecrow Press, 1993.
- BUCKLAND, M. K.: What is a „document”? In: Journal of the American Society for Information Science, 48. 1997. 804-809. Reprinted in: Historical Studies in Information Science, 1998. (T. B. Hahn and M. K. Buckland, eds.). Medford, NJ: Information Today, 1998. pp. 215-220.
- BUCKLAND, M. K.– PLAUNT, C.: Selecting Libraries, Selecting Documents, Selecting Data. In: Proceedings of the International Symposium on Research, Development and Practice in Digital Libraries 1997, ISDL 97, Nov. 18-21, 1997, Tsukuba, Japan. Tsukuba, Japan : University of Library and Information Science. pp. 85-91.
<http://bliss.berkeley.edu/papers/isdl97/isdl97.html>
- BUCKLAND, M. K.; C. Plaunt. (1994). On the construction of selection systems. In: Library Hi Tech, 48. 15-28.
<http://bliss.sims.berkeley.edu/papers/analysis/analysis.html>
- NORGARD, B. A. – BERGER, M. B.– BUCKLAND, M. K. – PLAUNT, C.: The online catalog: From technical services to access service. In: Advances in Librarianship, 17. 1993. 11-148.
- OLDING, R. K.(ed.): Readings in Library Cataloguing. Hamden, CT : Archon Press, 1997.
- OXFORD English Dictionary. 2nd ed. Oxford : Clarendon Press, 1989.
- PLAUNT, C.: A Functional Model of Information Retrieval Systems. Doctoral dissertation. University of California, Berkeley, 1997.

ISBN SZÁM MAGYAR E-KÖNYVEKNEK

Először adott ki a Magyar ISBN Iroda ISBN számokat magyar elektronikus könyvekre.
A 963-86133-0-0 ISBN számot Chi Wicca: NemS negyedek című szépirodalmi műve,
a 963-86133-1-9 ISBN számot pedig Perjés László: Intermarketing magyar szemmel
című szakkönyve kapta.

(www.prim.hu/online/cikk.prm?id.14014)