

Szükség van-e az osztályozásra a Google utáni korban?



HJØRLAND, Birger: „Is classification necessary after Google?” című tanulmányát (Journal of Documentation, vol. 68. 2012. no. 3. p. 299–317.) Koltay Tibor tömörítette.

Bevezetés

Az osztályozás (az indexeléssel, a dokumentumleírással és a metaadatok dokumentumokhoz való rendelkezésével együtt) a tudásszervezés (tartalmi feltárás) része.¹ Gyakorlata több mint száz éves múltra tekint vissza. A könyvtárosképzésnek is része, amióta 1876-ban *Melvil Dewey* létrehozta az első „könyvtárgazdasági” iskolát. Jövőbeli kilátásait azonban a digitális technológiák mind gyakorlati, mind elméleti tekintetben jelentősen megváltoztatták. Utólag az is felmerül, volt-e valaha szilárd elméleti alapja.

A tartalmi feltárás fejlesztési stratégiájának

kimunkálásakor abból kell kiindulnunk, hogy a használók hozzáférnek az internethez, ahol egymással versengő rendszerek és szolgáltatók állnak rendelkezésre. A használók számára mindegy, hogy az őket érdeklő dokumentumokat a helyi könyvtár osztályozta-e vagy más. Az információ megtalálásához használható legjobb tartalmi feltáró eszköz egy kattintásnyira van, és nincs szüksége sem a második legjobbra, sem a könyvtárak, a kiadók, vagy a bibliográfiai adatbázisok által párhuzamosan kínált több száz feltáró rendszer valamelyikére. Mindebből következik, hogy a tartalmi feltárás elméletének és gyakorlatának arra kell törekednie, hogy magas színvonalú szolgáltatásokat nyújtson, még-

hozzá globális léptékben, például úgy, ahogyan ez a szakterületi bibliográfiai adatbázisokban vagy az OCLC WorldCat-ban megvalósul. Ha a központosított osztályozások minősége nem elég jó, a használók gyorsan használhatatlannak minősítik őket, tekintettel arra, hogy vannak alternatívák.

A bibliográfiai osztályozások válságának másik jele, hogy a Thomson Reuters hivatkozási (idézettségi) adatbázisaiban² indexelt folyóiratokat egy elemi és intuitív, tehát nem tudományos alapokon nyugvó rendszer alapján osztályozzák.

Kitérő: néhány megjegyzés az ETO-ról

Az ETO mindenek előtt jelentős szakmai tudást tükrözött és számos tudományterületen összhangban volt az ott folyó diskurzussal. Eredetileg nemcsak monográfiák, hanem cikkek és más dokumentumok feltárására készült, ezért jóval részletesebb volt, mint például a Dewey Tizedes Osztályozás (DDC). Az ETO most is létezik és jelenleg is fejlesztés alatt áll. Nem lehetünk azonban büszkéek rá, mivel szakmai közösségünk nem volt képes megfelelően karbantartani és korszerűsíteni, és közel botrányos, hogy az új kiadásban olyan sok az elavult rész.

Az ETO-t valaha a Nemzetközi Információs és Dokumentációs Szövetség (FID) kezelte. A FID azonban 2002-ben feloszlott. Ezt megelőzően számos (nemzetközi és nemzeti) bizottság dolgozott a rendszer tökéletesítésén. Ma nem találkozunk semmi hasonlóval, ami a digitális környezetben problémát okoz, mivel egy minőségi tartalmi feltáró rendszer fejlesztése túl nagy feladat egy-egy könyvtár vagy olyan, kisebb szakértői csoport számára, mint amely a DDC-t a Kongresszusi Könyvtárban gondozza. Ha láthatóbbá akarjuk tenni a könyvtár- és információtudományt (könyvtártudományt) és a Google utáni korszakban hozzáférhetőbbé akarjuk tenni az információkat, valószínűleg nagyléptékű, a

gyakorló könyvtárosokat, a könyvtártudományi és más területekről érkező kutatókat a munkába bevonó, nemzetközi együttműködésekre lesz szükség.

A hagyományos könyvtári osztályozási rendszerekhez hasonlóan az ETO úgy készült, hogy alkalmas legyen a könyvek polcon való elrendezésére is. Ez olyan korlátozásokat von maga után, amelyek az online keresés esetében szükségtelenek. Ez valószínűleg részét képezi annak, hogy miért nem tölti be régi szerepét. Ennek okán a raktári és az információ-visszakeresési funkcióval külön-külön kell foglalkoznunk.

Azt is el kell mondanunk, hogy az ETO és a hozzá hasonló rendszerek feltehetőleg arra az elképzelésre alapozódtak, hogy az osztályozás semleges, objektív és tartalomtól függő döntésekre épülnek, tehát például „A” fogalom „B” fogalommal meghatározott összefüggésben áll, függetlenül hatókörétől vagy perspektívájától. Ezek a feltételezések hibásak. Helyesebb, ha azt gondoljuk, hogy specifikus, különböző témákat és paradigmákat megcélzó tartalmi feltáró eszközre van szükség. Az a széles körben elterjedt megközelítés, hogy az osztályozás szabványosítható és ilyen módon különböző kontextusokban újrafelhasználható, kérdéses, mivel a különböző szakmai közösségek (diskurzusközösségek) saját terminológiát, jelentéseket és relevancia-kritériumokat alakítanak ki. Nyomós érvek szólnak tehát a mellett, hogy az osztályozásnak a különböző szakterületek, ismeretelméleti közösségek és felhasználói csoportok igényeinek kielégítése érdekében testre szabottnak kell lennie. A feladat tehát nemcsak az, hogy aktualizáljuk az olyan rendszereket, mint az ETO vagy, hogy minden diszciplínára kiterjedő teauruszokat és más tartalomfeltáró rendszereket hozzunk létre. Ehelyett átfogó keretet kell adnunk a különböző alternatívák kialakítására és megvitatására. Ez lehetőséget kínál a felhasználóknak arra, hogy informált döntéseket hozzanak az információkeresés során. Azt is lehetővé teszi, hogy széles alapon nyugvó specifikus eszközöket fejlesszünk ki. Mindazonáltal továbbra is fontosak, sőt a tartalmi feltárás fennmaradásához feltehetően

szükségesek lesznek az olyan nagy, nemzetközi projektek, amelyek az *irodalmi igazolás*, azaz a szakirodalomban található fogalmak vizsgálata alapján osztályozzák a tudást. Az ETO-nak pedig a jövőben az előíró jelleg helyett leírónak és kontextuális jellegűnek kell lennie.

Mi az osztályozás?

Azt mondhatjuk, hogy az osztályozás a következő, egymással összefüggő folyamatok együttese:

- ⊙ osztályok meghatározása;
- ⊙ az osztályok közötti viszonyok, kapcsolatok meghatározása (az osztályozási rendszer megalkotása);
- ⊙ elemek (dokumentumok) hozzárendelése az adott osztályozási rendszerhez.

Mindaz azonos a következő, egymással összefüggő folyamatokkal:

1. fogalmak meghatározása;
2. a fogalmak közötti szemantikai relációk meghatározása;
3. annak meghatározása, hogy mely elemek tartoznak az adott fogalom alá (egy adott „dolgnak” egy fogalomhoz való rendelése).

Fontos, hogy megértsük, hogy az osztályozási rendszerek és szöveges indexelő nyelvek (a szabad szöveges rendszerek kivételével) az „osztályozási rendszerek” különböző típusai közé tartoznak.

A szabad szöveges rendszerekben nem a bibliográfiai rendszerhez kötődő személyzettől ered az osztályozás, mivel csak a dokumentum szerzője által adott „osztályozások” lehetségesek. Ezeket az osztályozásokat a szabad szöveges rendszerekben az adott szöveg jelentései közötti olyan relációk reprezentálják, amelyeket azok a szakmai közösségek alakítanak ki, amelyekhez az adott szöveg tartozik. Minden más rendszerben a könyvtárosok (vagy a bibliográfiai rendszerek „metaadatainak” kialakításáért, gondozásáért és használatáért felelős személyek) által kialakított osztályozásokkal találkozunk. Minden ilyen

rendszer ellenőrzött szókincsnek nevezhető, amelynek jellemzője, hogy normatív és előíró jellegű. Ezeket a rendszereket bibliográfiai osztályozásoknak nevezhetjük.

A bibliográfiai osztályozások nem függetlenek a tudományos osztályozásoktól, ezért a könyvtártudomány nem lesz képes az osztályozás megfelelő elméletét kimunkálni, ha figyelmen kívül hagyja a tudományok osztályozásának problémáit.

A tartalmi feltáró rendszerek a következő három általános kategóriát fedik le: *terminológiai listák*, amelyek a terminusok listáira helyezik a hangsúlyt (gyakran definíciókkal), *osztályozások és kategóriák*, amelyek a tematikus csoportok létrehozását hangsúlyozzák, *relációs listák*, amelyek a terminusok és a fogalmak közötti összefüggésekre koncentrálnak.

A tartalmi feltáró rendszerek (nem teljes) taxonómiája ennek megfelelően a következő.

Terminológiai listák:

- ⊙ egységesített névalakok listái,
- ⊙ szójegyzékek,
- ⊙ szótárak,
- ⊙ földrajzi névtárak.

Osztályozások és kategóriák:

- ⊙ tárgyszójegyzékek,
- ⊙ osztályozási rendszerek,
- ⊙ taxonómiák,
- ⊙ kategorizálási rendszerek.

Relációs listák:

- ⊙ tezauszok,
- ⊙ szemantikai hálók,
- ⊙ ontológiák.³

A legfontosabb különbség a tartalmi feltáró rendszerek különböző típusai között leginkább abban van, hogy különböző típusú szemantikai relációkat jelenítenek meg. A hagyományos osztályozási rendszerekben a hierarchikus viszonyok, valamint a szinonimák és a homonimák közötti kapcsolatok a legfontosabbak. Az ontológiákban a szemantikai kapcsolatok széles skálája lehetséges. A kérdés az, hogy más rendszerek szükségesek-e a tartalmi feltárás elméletének kidolgozásához, továbbá, hogy a többi

tartalmi feltáró rendszer az ontológiák olyan speciális típusának tekinthető-e, amelyekben a szemantikai viszonyok korlátozottabb skálája áll rendelkezésre.

Az osztályozások tezaszauruszokká alakíthatók, ami azt mutatja, hogy a tartalmi feltáró rendszerek formája kevésbé fontos. Az osztályozás középponti kérdése – ahogy arról már szóltunk – az, hogy milyen módon tudjuk meghatározni a fogalmak közötti szemantikai viszonyokat. Meg kell jegyeznünk, hogy egy adott tartalomfeltáró rendszer sokkal több, mint rendszerjellemzőinek összessége. A használat során a könyvtárak számos megalapozott döntést hoztak egy-egy fogalom osztályba sorolásáról, ami fontos ismeretek felhalmozódását jelentheti, amit viszont ritkán tükröz a tartalmi feltárás irodalma, tehát inkább tacit tudás.

A kutatásban az információkeresés hagyománya nagyobb befolyásra tett szert, mint az osztályozásé, így a két terület kutatói között a megosztottság a jellemző abban a tekintetben, hogy fontosnak tartják-e az ellenőrzött szókinccsel kapcsolatos vizsgálatokat. A hagyományos értelemben vett osztályozás elleni egyik érv pedig az, hogy az ellenőrzött rendszerek mindegyike a maga önkényes döntéseire épül, ezért a több rendszerben (adatbázisban) történő egyidejű keresés hatékonyságát rontja.

Az osztályozás ismérvei

Az alapvető elméleti kérdések közül az első az, hogy miként határozzuk meg annak kritériumait, hogy „A” dokumentumot az „X” osztályba soroljuk-e? A másik pedig az, hogy milyen ismérvek alapján hozható meg a döntés arról, hogy „A” és „B” fogalom között „X” szemantikai viszony határozható meg.

A könyvtári hagyomány főárama azt a szabályt követi, hogy egy könyv legalább 20%-ának „X”-ről kell szólnia, hogy az adott osztályhoz soroljuk. Az automatikus indexelésben viszont „A” szó előfordulása határozza meg, hogy „X”-hez soroljuk-e vagy nem.

A felhasználói igényekhez igazadó indexelés esetében az anticipált felhasználói kérdések határozzák meg, hogy mikor soroljuk „A”-t „X”-hez. Ilyenkor az indexelő azt kérdezi, hogy melyik deskriptor alá kellene az adott entitást sorolni.⁴ Ezzel összhangban mondható, hogy az indexelés lényege, hogy megbecsüljük, miben járul hozzá az adott cikk a tudáshoz.⁵

Annak érdekében, hogy az indexelés jó minőségű és konzisztens legyen, az indexelőnek alaposan ismernie kell a téma szerkezetét és azt, hogy mivel járul hozzá a dokumentum a tudás fejlődéséhez.

Amíg az osztályozással kapcsolatos kutatások fő iránya arra az objektivista értelmezésre épül, hogy a dokumentumnak van valamilyen tárgya, egyre inkább teret nyer az a kisebbségi értelmezés, mely szerint a dokumentumhoz annak érdekében rendeljük „X” tárgyat, hogy valamely specifikus tevékenységet támogassunk. Az osztályozás jövőbeni elmélete és gyakorlata szempontjából ez az utóbbi nézet meghatározó.

Amikor azt a kérdést vizsgáljuk, hogy miként jelölhetjük a fogalmak közötti szemantikai kapcsolatokat, az ellenőrzött szókinccsre épülő eszközöket használó könyvtárosok gyakran feltételezik, hogy ezzel kapcsolatos döntéseik a logikán vagy az adott fogalmak közötti rögzített szemantikai kapcsolatokon alapulnak. Az ellenőrzött szókinccset azonban interpretációnak kell tekintenünk: egy olyan szemantikai rétegnek, amelyet a felhasználó által bevitt kifejezés és az alapul szolgáló adatbázis közé illesztünk abból a célból, hogy jobban képviselje a felhasználók eredeti szándékát jelző kifejezést.⁶ Egy ilyen interpretáció gyakran a tapasztalaton alapul, továbbá különböző perspektívák és érdekek a szemantikai kapcsolatok értékelésének eltérő ismérveit eredményezi. A héliumot például a kémiában nemesgáznak tekintjük. *Stowe* (kvantummechanikán alapuló) periódusos rendszerében viszont az alkáliföldfémek között szerepel.⁷ A modern tudomány nem oldotta meg a hélium helyét, és nyitott kérdés, hogy van-e erre a kérdésre helyes válasz. Az ilyen megfontolások teszik fontossá, hogy az osztályozásról úgy gondolkozunk, mint ami nemcsak logikus döntésekről szól, hanem

szövegek értelmezésén, továbbá különböző érdekek egyeztetésén alapul.

A deskriptorok, osztályozási jelzetek stb. elvben az adott, osztályozandó szakterület történetét feldolgozó könyvek fejezetcímeinek felelnek meg. Minden ilyen mű tükrözi szerzőjének és a korszellemnek a szubjektivitását, amit nem lehet és valószínűleg nem is helyes kiküszöbölni.

A könyvtártudománynak tanulmányoznia kellene a műfaj-fogalmak és más, a dokumentumok megfogalmazására használt módszerek használatát, továbbá szükséges volna, hogy a kutatók a saját gyakorlatukba átültessék ezeket. Ez a tudás nemcsak az indexeléshez, hanem minden, az információhoz kötődő szakmai tevékenységhez szükséges. Ezen kívül, ennek a tudásnak a meglétét tételezzük fel az olyan magas színvonalú könyvtárakban és bibliográfiai adatbázisokban, mint az Egyesült Államok Nemzeti Orvostudományi Könyvtára és a Medline. Ha az ilyen adatbázisok előállításakor azzal csökkentenék a költségeket, hogy nem alkalmaznak szakértő munkatársakat, végeredményben eltűnne a tartalmi feltárás és az osztályozás, mivel a jövőben csak a legjobb osztályozási gyakorlatokra lesz szükség.

Hogyan kapcsolódik a könyvtári osztályozás más osztályozásokhoz?

Amikor a könyvtári szakemberek osztályoznak egy adott könyvet, azok a fogalmak, amelyeket használnak, a szakirodalomból származnak és nem elsősorban ők maguk hozzák ezeket létre. Ahogy azt *Hulme* írja, az irodalom valódi osztályozását a könyv írója végzi el, a könyvtáros csak rögzíti azt.⁸ Ez a szemlélet azonban korán eltűnt a tartalmi feltárásból.

Ezzel a hagyományos nézettel szemben az osztályozásnak azokat a célokat kellene tükröznie, amelyek elérése érdekében létrehozták őket. Emellett tudnunk kell, hogy a különböző tudományok, elméletek és az emberi tevékenység többé-kevésbé különböző módon osztályozzák a világot.

A könyvtártudomány és a tartalmi feltárás számára ebből az a tanulság, hogy az osztályozás független az ezeken a területeken kívül felhalmozott tudástól. Az osztályozáskor figyelembe kell venni annak célját és azoknak a tevékenységeknek a természetét, amelyeket az támogatni hivatott.

Miért van szükség az osztályozásra: a tényeken alapuló gyakorlat esete

Ha azt akarjuk, hogy az osztályozások relevánsak legyenek, eltérő nézetek és érdekek figyelembe vételével kell megalkotnunk őket. Ennek érdekében az osztályozásoknak lehetővé kell tenniük, hogy az információkeresés az ezekhez fűződő relevancia-kritériumok alapján történjék. A Google és más hasonló információ-visszaereső rendszerek lenyűgöznek bennünket, viszont tudnunk kell, hogy miként osztályozzák és rangsorolják a releváns információt. Azt gondolhatnánk, hogy ezek az információkereső rendszerek semlegesek és objektívak, de ez lehetetlen, hiszen minden rendszer torzít.

A keresőgépek esetében lehetséges, hogy különböző találatok vagy rangsorolások elérésére kalibrálják őket. Ennek eléréséhez vagy arra, hogy értékeljük ezeket a rendszereket, szükségünk van osztályozásra, amely megmutatja, hogy mit kellene megtalálnunk. A könyvtártudomány területén eddig a „felhasználói relevancia” alapján ítéltük meg a relevanciát. Amennyiben azonban bízunk kellene például orvosi dokumentumokban, helyesebb lenne, hogy a relevanciáról alkotott ítéletünk tudományos ismérveken (például a kutatási módszereken) alapulna, ahelyett, hogy a felhasználók véleményére alapoznánk. Egyértelműen ezt teszi a tényeken alapuló gyakorlat néven ismertté vált mozgalom, amelyik a tényeken alapuló orvosláson alapul.

A tényeken alapuló gyakorlat elvei szerint a dokumentumokat világosan meghatározott ismérvek alapján osztályozzuk (a kutatási módszerek hierarchiájára, vagyis arra építve, hogy

mi számít bizonyítéknak). A tényeken alapuló gyakorlatot azonban túlzott formalizmus jellemzi, ráadásul túlságosan mechanikus ahhoz, hogy erre alapozzuk a tudományos osztályozást. Függetlenül attól, hogy elfogadjuk-e a tényeken alapuló gyakorlat filozófiáját, az osztályozás fontos marad. Ha ugyanis a kutatási módszerek egy-egy adott rangsorát a legjobbnak fogadjuk el, a cikkeket a rangsorban felsoroltak alapján kell osztályoznunk. Ha kialakítunk ilyen rangsorokat, lehetségessé válik, hogy algoritmusokat dolgozzunk ki, amelyek – bizonyos mértékig megbízható módon – alkalmasak az automatikus osztályozásra, bár magát az osztályozást előre létre kell hozni, így az eredmény „emberi felügyelettel működő gépi tanulás” vagy „emberi felügyelettel működő gépi osztályozás” lesz. A tényeken alapuló gyakorlat nem a végső válasz az osztályozás problémáira, azonban sokkal egészségesebb megközelítést kínál az információkereső rendszerek és az osztályozás megalapozására, mint azok a paradigmák, amelyek a könyvtártudomány és a tartalmi feltárás területén domináltak az elmúlt 30 évben. A tényeken alapuló gyakorlat léte annak a fontos következtetésnek a levonására is alapot ad, hogy nem lehetünk meg osztályozás nélkül.

Miért van szükség az osztályozásra a Google utáni korszakban?

A tartalomfeltáró rendszereket olyan ontológiáknak tekinthetjük, amelyek sajátos fogalomalkotásra alapozva adnak válogatást fogalmakból és a hozzájuk tartozó szemantikai relációkból. Az ontológiák többet jelentenek az objektív valóság semleges visszatükrözésénél. Olyan világnézetekből építjük fel őket, amelyek – más célok rovására – bizonyos célok és értékek elérésére jók.

Figyelembe véve az osztályozásnak és a tartalmi feltárásnak ezt az új elméletét megérthetjük, hogy miért van szükség osztályozásra bármely, könyvtári, dokumentációs és információs munkaterületen. Az osztályozás kategóriái egyszerűen a releváns információ szolgáltatásának ismérveivel azonosak.

Jegyzetek

1. Az eredetiben a *knowledge organization* (tudásszervezés) terminus szerepel. Idehaza azonban a *tartalmi feltárás* (tárgyi feltárás) kifejezés nagyobb népszerűségnek örvend.
2. A Web of Science (a Science Citation Index, a Social Sciences Citation Index, az Arts & Humanities Citation Index és a Conference Proceedings Citation Index)
3. HODGE, G.: Systems of knowledge organization for digital libraries. Beyond traditional authority files. Washington, DC: Council on Library and Information Resources, 2000. <http://www.clir.org/pubs/reports/pub91/contents.html> (2012. június 6.).
4. SOERGEL, D.: Organizing Information: Principles of Data Base and Retrieval Systems. Orlando, Academic Press, FL: 1985.
5. ROWLEY, J.E. – FARROW, J.: Organizing Knowledge: An Introduction to Managing Access to Information, 3rd ed., Aldershot: Gower, 2000.
6. LEISE, F. – FAST, K. – STECKEL, M.: What Is A Controlled Vocabulary? Boxes and Arrows, 2002 [http://www.boxesand-arrows.com/view/what_is_a_controlled_vocabulary_\(2012.junius.27.\)](http://www.boxesand-arrows.com/view/what_is_a_controlled_vocabulary_(2012.junius.27.))
7. CHANNON, M.: The Stowe table as the definitive periodic system” = Knowledge Organization, vol. 38. 2011. no. 4. p. 321–327.
SCERRI, E.: The Periodic Table, Its Story and Its Significance, Oxford, Oxford University Press, 2007.
8. HULME, E.W.: Principles of book classification” = Library Association Record, vol. 13, 1911, p. 354–358, p. 389–394. p. 444–449.