

FALUDI ANDREA

LEG Zrt. és KRE TERMIK  
faludiandi@gmail.com

Faludi Andrea: A dokumentáció mint információtudomány és a terminológia kapcsolata  
Alkalmazott Nyelvtudomány, XVII. évfolyam, 2017/2. szám  
doi:<http://dx.doi.org/10.18460/ANY.2017.2.003>

## **A dokumentáció mint információtudomány és a terminológia kapcsolata**

Relationship between Documentation (as Information Science) and Terminology. Traditionally, information acquisition, storage and retrieval were exclusive to Information Science. Today, however, Terminology is increasingly participating in these areas and activities. Information science tools (thesauri, industrial applications and ontologies) and documentation and terminology are equally working with and categorizing the texts based on terminological units. The aim of the study is to map out and present the connections between documentation as information science and terminology based on the relevant literature of these fields.

### **1. Bevezetés**

A tanulmány célja a dokumentáció mint információtudomány és a terminológia kapcsolatának feltérképezése és bemutatása, elsősorban a releváns (információtudományi, dokumentációs, terminológiai) szakirodalomra támaszkodva. Ha a dokumentáció szakirodalma után érdeklődünk, legelőször a könyvtártudomány, újabb nevén az információtudomány területéről kapunk találatokat. Az információk kinyerésével, gyűjtésével, tárolásával és újrafelhasználásával hagyományosan az információtudomány foglalkozik. A szakirodalmat tanulmányozva rögtön szembetűnik, hogy a terminológia kulcsszerepet játszik a dokumentációban. A terminológia és a dokumentáció egymással szoros összefüggésben álló, több szinten együttműködő területek. A dokumentáció a dokumentumokból kinyert információkat terminológiai egységek segítségével rendezi, az információtudománnyal foglalkozó szakemberek pedig mind az információkinyerés, mind a visszakeresés folyamatában terminológiai egységekkel dolgoznak.

A szakterület alapfogalmait az UNESCO által 1976-ban kiadott négy nyelvű terminológiai szótár a *Terminology of documentation* (a továbbiakban: Tod) gyűjti össze. A szótár **a dokumentáció** fogalmát az alábbiak szerint összegzi: Dokumentumok és adatok folyamatos és rendszerezett kezelése, beleértve az elhelyezést, azonosítást, feltárást, elemzést, tárolást, visszakeresést, újrafelhasználást és megőrzést a szakemberek számára.<sup>1</sup>

Tehát – ellentétben a műszaki dokumentáció készítésével – elsősorban **a tárolást és a visszakereshetőséget**, nem pedig a dokumentáció írását, létrehozását jelöli meg saját céljaként.

Az információtudomány szakirodalmával némileg ellentétben Cabré a műszaki dokumentáció és a fordítás irányából közelít a terminológia és dokumentáció kapcsolatához, és a terminológia dokumentációs célú felhasználására vonatkozóan három területet különböztet meg: (1) a szakírást, (2) szakszövegek fordítását és tolmácsolását és (3) a műszaki dokumentációt (1999: 50). (Erről, a terminológiához szintén szorosan kapcsolódó műszaki dokumentációs és dokumentum-menedzsment irányzatról részletesebben l. Fóris–Faludi, 2017.)

Az információ sokféle megközelítésével kapcsolatos nézőpontokat, tudományágakat fogja össze az International Society for Knowledge Organization (ISKO, <http://www.isko.org/>) 1989-ben alapított nemzetközi szervezete, mely rendszeresen szervez konferenciákat, tudományos folyóirata *Knowledge Organization* címmel 1973 óta jelenik meg.

## 2. A dokumentáció rövid története

A **dokumentáció tudományának** létrejöttét legalább két dátumhoz kötik. Az első dátum a *Pharmazeutisches Zentralblatt* című referáló folyóirat (EN abstracting journal) első megjelenése (1830), amely a gyógyszerkémia szakirodalmát fogta össze, és ugyan más néven, de 1945-ig folyamatosan megjelent. A dokumentáció második születésnapjának 1895-öt tekintik, amikor megszületett az úgynevezett IIB (Institut International de Bibliographie), amely céljával nem kisebb feladatot választott, mint az egyetemes bibliográfia elkészítését. A szervezet 1931-ben az International Institute for Documentation, 1937-ben pedig az International Federation for Documentation nevet vette fel. 1988-ban keresztelték a maival megegyező International Federation for Information and Documentation (FID) névre. A dokumentumok leírásának és elérhetővé tételének új szempontjai – mint a tartalomorientáltság, aktualitás, gyorsaság stb. – a dokumentáció tevékenységét már szükségszerűen a könyvtáron kívülre helyezték. (A könyvtárak folyamataikban lassúnak, lomhának bizonyultak erre az új típusú feladatra.) A dokumentáció fogalma is meglehetősen dinamikusan változott a 19. század első felében. A fogalmi változást a FID maga is figyelemmel követte, és az alábbi definíciókat tette közzé: „A dokumentáció bármely fajta **dokumentum** gyűjtése, osztályozása és terjesztése az emberi tevékenység bármely területéről.”<sup>2</sup>, illetve később, 1953-ban: „A dokumentáció bármely fajta **információ** gyűjtése és megőrzése, osztályozása és válogatása, terjesztése és hasznosítása.”<sup>3</sup> (Polzovics, 1959: 327). A FID számtalan kiadványt jelentetett meg, legutolsó szakfolyóirata a *FID Review* volt. A FID 2001-ben súlyos pénzügyi problémák miatt megszűnt (web 2).

Maga Paul Otlet (az IIB, és egyben a dokumentációtudomány atyja) még ennél is szélesebb értelmű definíciót alkotott, szerinte a dokumentáció

„tudomást szerez mindarról, ami valamely kérdés szempontjából eredeti vagy fontos” (Otlet, 1934: 317)<sup>4</sup>. A későbbiek során előtérbe került a dokumentációs nyelv fontossága is. Polzovics Iván (1962) fontosnak tartotta, hogy a dokumentáció egységes nyelven valósuljon meg, hiszen maga a dokumentáció az, ami képes elhárítani a nyelvi akadályokat (Verók, 2011).

A tényleges dokumentációs munka kezdete Magyarországon 1947-re tehető. [E]kkor megalakul egy műszaki dokumentációs központ, amely Műszaki Dokumentációs Intézet (Hungarian Technical Documentation Institute) néven nemzeti tagként belép a FID-be (Balázs, 1960: 59).

Két év múlva, 1949-ben hivatalos, állami feladattá vált a dokumentáció ügye. Ez az eredmény a Magyar Tudományos Tanácsnak köszönhető, amely felismerte, hogy szükség van az egymástól függetlenül létrejött és működő dokumentációs műhelyek összefogására, központi irányítására. Így jött létre az Országos Dokumentációs Központ, mely 16 hónapnyi működés után meg is szűnt, ám a szakmai dokumentációs központok tovább működtek. Egy 1949-ben hozott kormányrendelet lehetővé tette, hogy az előírtakon túl más szakterületi dokumentációs központok is létesüljenek, így az 1950-es évek elején több szakmai dokumentációs szervezet is létrejött, többek között például a földrajz, az irodalomtörténet és a jog területén (Kégli, 2009).

Körülbelül az 1970-es évektől az úgynevezett információs (vagy tudásalapú) társadalom kialakulásával, illetve az információrobbanás után, az információ előállítás, elosztása, terjesztése, használata és kezelése jelentős gazdasági, politikai és kulturális tevékenységgé vált. Egyre több intézmény, szervezet kezdett dokumentációval foglalkozni, sőt az információt érintő tevékenységek is egyre több szakterületre diverzifikálódtak.

Tehát a szakterület dinamikus fejlődése egyben a szétfejlődését is jelentette. Mára a dokumentáció átszövi a mindennapjainkat, nemcsak szervezeti szinteken, de gyakorlatilag magánéletünk nagy részét is dokumentációval töltjük. Szervezeti formában mindenfajta könyvtár, információs ügynökség, dokumentációs szervezet foglalkozik dokumentációs feladatokkal. Dokumentációs rendszerek kiépítésével, karbantartásával az informatikus, könyvtáros szakembereken túl nagyon sokan foglalkoznak, ugyanis minden szervezetben szükséges ezeknek a munkáknak az elvégzése.

Ez a fajta robbanásszerű fejlődés a dokumentációs rendszerek sokszínűségében is megmutatkozik, melyeket gyakran tudásmenedzsment megoldásoknak is neveznek. Ide sorolhatók például a *vezetői információs rendszerek (VIR)*, az *adattárház*, az *adatbányászati rendszerek*, a *dokumentumkezelő megoldások*, a *tartalomkezelő megoldások*, a *csoportmunkát támogató megoldások* vagy a *komplex tudásmenedzsment rendszerek*.

### 3. A dokumentáció mint információtudomány

A **dokumentáció** és az **információ tudománya** (avagy könyvtártudomány) szoros kapcsolatban állnak egymással. Verók szerint a könyvtár és a dokumentáció tulajdonképpen ugyanazt a tevékenységet jelöli: „Nem eltérő intézményekről van szó, hanem eltérő nézőpontokról. Az egyik megelégszik a hordozók fizikai rendjével, a másik nézőpont a tartalmakra orientál” (2011: 55). Verók értelmezésével némiképp szembemegy, hogy Polzovics már 1959-ben is úgy látta, hogy „az információ fogalmának teljes elkülönítése a dokumentációtól kevésbé megokolt és (...) gyakorlati célra nem is szükséges, sőt kontraindikált” (1959: 330). Polzovics 1961-ben terminológiai vitába keveredett Josef Koblitzcal, a berlini Institut für Dokumentation helyettes igazgatójával is. Vitájuk „A dokumentáció, az információ és a tájékoztatás fogalma. Terminológiai vita” című kiadványban követhető nyomon magyar nyelven (Polzovics, 1961). A vita elindítója Koblitz „Dokumentation und Information. Eine Terminologische Untersuchung” című tanulmánya volt (Koblitz, 1959), mely szerint az addig a dokumentáció egyik funkciójának tekintett tevékenységi kör **információ** elnevezéssel már önálló, a dokumentációtól elkülönült ismeretággá nőtte ki magát. Polzovics szerint viszont a megjelölt tevékenységi kör jelentős fejlődése ugyan indokolttá teheti külön fogalmi megjelölését, de úgy tartja, hogy ez idő szerint sem elvi, sem gyakorlati szempontok nem szólnak amellett, hogy az és a rávonatkozó ismeretek összessége mint ismeretág a dokumentáció fogalmi köréből kirekesztődjék (Polzovics, 1961: 5).

#### 3.1. A dokumentáció és az információ fogalmának viszonya

A fogalmak körül kialakult vita oka a könyvtártudomány fejlődésében, elméleti változásában keresendő. Koltay és Prókai (2010) tisztázzák, hogy a könyvtártudományi terminusok változása nagyjából 50 éves ciklusokban nyomon követhető. Ahogyan a 19. század végén és a századfordulón a könyvtári tevékenység kulcsfogalma a **bibliográfia** volt, a 20. század közepére a **dokumentáció** lépett a helyére, majd napjainkra az **információ** fogalma került a középpontba (Koltay–Prókai, 2010: 270). Az egyes fogalmak közkeletűvé válása az alábbi dátumokhoz köthető: 1885: bibliográfia, 1934: dokumentáció, 1988: információ (Koltay–Prókai, 2010: 271). Érdekesség, hogy 1988-ban – amikor az információ fogalma uralkodóvá válik a szakterületen – az UNESCO az **alkalmazott nyelv tudomány** alá sorolta a dokumentációs tevékenységeket (Horváth, 1999: 19). A dokumentációval kapcsolatos tevékenységeket az UNESCO tudományfelosztási nomenklatúrájában ma is az alkalmazott nyelvészet részdiszciplínáiként találjuk meg. Ezek a következők: *referálás (abstracting)*, *automatizált dokumentáció (automated documentation)*, *információkereső nyelvek (documentary languages)*, és maga a *dokumentáció (documentation)* (Koltay–Prókai, 2010: 277).

A dokumentáció mint tudományterület az alkalmazott nyelvtudomány körébe való sorolása nem új keletű: Szépe György 1967-ben a következőképpen fogalmazott:

Az alkalmazott nyelvészet a társadalom által feladott feladatokat kíván a nyelvészet segítségével – vagy azzal is! – megoldani. Ezek a feladatok a társadalmi gyakorlat különféle területéről származhatnak: ilyen a közvetlen termelés, a dokumentálás, a távközlés, az oktatás, a kultúra stb. Ennek megfelelően beszélhetünk pl. dokumentációs vagy oktatási stb. alkalmazott nyelvészetről (Szépe, 1967: 3).

A dokumentáció alkalmazott nyelvészetbe sorolását az is indokolja, hogy a dokumentumok többsége nyelvi közlés terméke, a használt nyelv pedig valamely természetes nyelv. Ezért nem célszerű ezt a nyelvészettel párhuzamos diszciplínát külön felépíteni, hanem fel kell használni a nyelvészet idevágó eszközeit, módszereit és eredményeit (Szépe, 1967).

### **3.2. A dokumentáció mint információtudományi szolgáltatás**

Az információtudományi értelemben vett dokumentáció szolgáltatásai közül a legfontosabb az úgynevezett *referáló szolgáltatás*. A referáló szolgáltatás az eredeti tartalom kb. tizedére tömörített tartalmi összefoglalókat (referátumokat) kínál. A referátum leküzdí a nyelvi korlátokat (mindig egy nyelven történik, napjainkban természetesen az angol nyelvű referátum a legelterjedtebb), és az eredeti szöveg lehető legobjektívebb leírását adja. Minden referátumnak öt fontos kritériumnak kell megfelelnie, ezek a teljesség, pontosság, objektivitás, rövidség és érthetőség. A referáló szolgáltatás során a referátum mellett rendszerint megadják még az eredeti tanulmány bibliográfiai leírását indexelő kifejezéseket, tárgyszavakat, deskriptorokat (Verók, 2011).

Mára elérhetővé vált az automatizált referáló szolgáltatás is. Angol nyelvterületen már régóta automatizálták a referátumok készítését, ám Magyarországon ez még mindig nem jellemző gyakorlat. Lengyelne Molnár Tünde (2008) nevéhez köthető a magyar nyelvű offline kivonatoló program elkészítése. Az automatikus kivonatolás alapja, hogy szövegstatistikai eljárásokkal fontosságuk szerint súlyozzák az eredeti szöveg mondatait, majd a „legfontosabb” mondatokat az eredeti szövegben való előfordulásuk sorrendjébe helyezik. A mondatok előfordulásának gyakoriságán túl szerepet kap a releváns szakszavak előfordulásának gyakorisága is (Lengyelne, 2008).

A referátumok magyar nyelven történő automatizálásának elsődleges problémája, hogy magyar nyelven hiányoznak az úgynevezett szakterületekre jellemző gyakorisági szótárak, amelyek egy-egy szakterület szókincsét gyakorisági mutatókkal ellátva közölnék. Ezekre azért lenne szükség, hogy egy adott témájú szakszöveg automatizált kivonatolását valódi, az adott szaknyelvre valóban jellemző gyakorisági adatok mentén végezhesék el.

## 4. A dokumentáció mint információtudomány és a terminológia kapcsolata

### 4.1. Documentary language, azaz információkereső nyelvek

Az olyan nyelveket, amelyeket a „dokumentációs rendszerek indexelés, tárolás és visszakeresés céljára használnak”<sup>5</sup>, „documentary language”-nek nevezzük. (A *Tod* valójában több angol nyelvű terminust is kínál ennek a fogalomnak a megnevezésére: *documentation language, information storage and retrieval language, information language*; franciául: *langage documentaire*; németül: *Dokumentationssprache*). Ennek a fogalomnak a magyar szakirodalomban elterjedt megnevezése: *információkereső nyelvek*.

Az információkereső nyelvek olyan mesterséges nyelvek, amelyek az információk megtalálásában a természetes nyelvből a szakmai nyelvekbe átöröklődő spontán használat okozta nehézségeket hivatottak kiküszöbölni (Ungváry–Vajda, 2002: 8).

Többek között jelentéstani problémákat (szinonímia, homonímia, poliszémia stb.). Ezért a dokumentumok leírásához, indexálásához szabványosított nyelveket alkalmaznak, kulcsszavakat vagy deskriptorokat használnak, amelyek szakszövegek esetében a terminusok.

A dokumentumok indexálásához használt szabványosított terminológia lehetővé teszi, hogy a dokumentáció szisztematikus és egyértelmű legyen. A terminológia a teauruszok készítéséhez és felhasználásához is hozzájárul, olyan lépésekben, mint a fogalmi osztályozás tervezése, a fogalmak kiválasztása és a közöttük lévő kapcsolatok meghatározása, illetve a szinonimák, homonimák kezelése (Cabré, 1999: 51). **Összességében elmondhatjuk, hogy az információkereső nyelvek (szerencsés esetben) szabványosított terminológiai rendszerek.**

Koltay és Prókai (2010) is figyelemreméltónak tartja a szemiotika, a könyvtár- (vagy információ-) tudomány és a terminológia kapcsolatát, mivel mindegyik tudományterület érdeklődése kiterjed a tartalom és reprezentációi, a jelölt és a jelölő, a hivatkozott tárgy és a hivatkozás közötti kapcsolatok vizsgálatára.

#### 4.1.1. Könyvtári osztályozás – az információkereső nyelvek egy alfaja?

A magyar nyelvű szakirodalom alapján az **osztályozás és információkereső nyelvek** kifejezéssel jelölt fogalmak értelmezése körül is bonyodalmakba kerülünk.

Utóbbi kialakulási körülményei folytán és nyelviségéből eredően is általánosabbnak, s eltérő célúnak is felfogható lenne, de a továbbiakban egymás szinonimáinak tekintjük őket (B. Hajdu–Babiczy, 1998: 38).

Így ebben az értelmezésben az osztályozási rendszer és az információkereső nyelvek feladata egyaránt „az információk tartalmának leírása, tárolása és a

lehetőleg több szempontú visszakeresés biztosítása” (B. Hajdu–Babiczky, 1998: 38).

A Tod alapján azonban az *osztályozás az információkereső nyelvek (documentary language)* egy alfaja, amely a fogalmakat osztályokba rendezi, a közöttük lévő kapcsolatokat pedig leggyakrabban hierarchikus viszonyokkal fejezi ki<sup>6</sup>. (Tehát nem minden esetben hierarchikus kapcsolatokkal dolgoznak, hiszen *a kifejezetten csak hierarchikus kapcsolatokat alkalmazó osztályozást* nevezzük **hierarchikus osztályozás**-nak (EN *hierarchical classification*).

Az információkereső nyelvek a felhasználás célját tekintve is többfélék lehetnek: például tudományokat osztályozók, ismeretközvetítők és ismeretszervezők (B. Hajdu–Babiczky, 1998: 43).

Az **osztályozás** a könyvtártudományban a tartalmi feltárás egy módszere, illetve szintje. (Az öt feltárási szint: *bibliográfiai leírás, osztályozás, referálás, tömörítvény és szemle*). Ez az eljárás a dokumentumról megállapított tartalmi ismereteket fogalmi láncsal fejezi ki, ezáltal alkalmas az információk átfogó tartalmi csoportosítására. Az ismérvek több formában jelenhetnek meg, pl. összefoglaló táblázatokban, tárgyszójegyzékekben (B. Hajdu–Babiczky, 1998: 10). Az osztályozás során „a tárgyakat hasonlóságuk mértéke alapján egybegyűjtjük, és különbségeik mértéke alapján elkülönítjük egymástól” (B. Hajdu–Babiczky, 1998: 21). Ez az eljárás alapvetően megegyezik a tudományos osztályozással, azonban a gyakorlatban rendszerint eltér tőle, amelynek oka például, hogy a könyvtári osztályozásnak a dokumentumok fizikai oldalát, megjelenési formáját is ki kell fejeznie. Így a könyvtári osztályozás számára „a tudományok osztályának tükrözése csak másodlagos, eszköz szerepet játszik” (B. Hajdu–Babiczky, 1998: 24).

A könyvtári osztályozással kapcsolatos kutatások az 1950-es években oda jutottak, hogy az addig kialakult és domináns hierarchikus – nem-faj, egész-rész összefüggések – túl merevek, elégtelenek és képtelenek a sokrétű valóság-, illetve tudattartalmak összefüggéseinek leképezésére (Pálvölgyi, 2011b).

Hatékonyabb információkereső rendszerek kidolgozásához szükségessé vált az összefüggések világában való több szempontú megközelítés és eligazodás biztosítása, ahogyan az elsinore-i konferencián (1964) elhangzott: „az osztályozás a fogalmi egységek közötti relációk kialakításának módszere” (Pálvölgyi, 2011b). Jason E. L. Farradane kidolgozta a *relációs indexelés*-t, majd elméleti modelljére alapozva számos fejlesztés kezdődött a sokrétű összefüggések ábrázolására. Ennek köszönhetően az 1960-as évektől világszerte és Magyarországon is teauruszok építésébe kezdtek, melyekben az összefüggések ábrázolása már elemi értékű volt. Az összefüggések (relációk) ábrázolásának „diadalmenete” azóta is folytonos (Pálvölgyi, 2011b).

#### **4.1.2. Osztályozás: a fogalmak leírása és elhelyezésük a fogalmi rendszerben**

Mára elmondható, hogy a könyvtári osztályozás mindig a fogalom szintjén zajlik (B. Hajdu–Babiczky, 1998: 24), azonban a fogalmak „leírása” – Ranganathan alapján – legalább három szinten történik. Először kinyerik/meghatározzák a fogalmat, majd ezt egy nyelvi jelölővel látják el, majd ezen jelölő alapján meghatározzák annak helyét a már kialakított osztályozási rendszerben, és hozzárendelnek egy „újabb” jelölőt, az adott osztályozási rendszer jelkészletéből (*jelzetelés*) (B. Hajdu–Babiczky, 1998: 24). Az első lépés a szemantika egyik alapvetéséből adódik, miszerint a fogalmak absztrahált ismeretek, ezeket pedig címkékkel látjuk el, hogy a nyelvi kommunikációban használhassuk őket.

A fogalom és jelölő közötti kapcsolatok leírásával a *terminológia* foglalkozik. A terminológiai vizsgálatok abból indulnak ki, hogy a nyelv feladata a világi környezetből származó ismeretek pontos megfogalmazása, rendszerezett tárolása és továbbítása, vagyis a fogalmakban absztrahált ismereteket kell a nyelvnek mint kommunikációs rendszernek közvetítenie. Ezért a fogalmat, valamint a fogalom és a jelölő viszonyát, a jelentést állítják a kérdéskör középpontjába. A fogalmak nyelvi megjelenítése a terminusokkal történik (Fóris, 2014: 29).

A terminusok rendszerbeli helyének meghatározása a jelölőtől függetlenül, a fogalom alapján történik, az úgynevezett „fogalomazonosító” segítségével. Utóbbi lehetővé teszi a fogalom nyelvi jelölő nélküli kezelését, így nincs szükség „jelzetelésre”, továbbá kikerülhetjük a nyelvi jelölők „csapdáját”, amely magával hozná a szinonímia és homonímia jelenségének problémáit is.

Fóris szerint a nyelvi hálózat, és azon belül a domének terminológiai hálózata skálafüggetlen hálózatokként modellezhetők. Fóris terminológiai hálózatok modellje Cabré terminológiai modelljén és a hálózatelméleten nyugszik. A modell alapja, hogy a terminológiai megközelítés három összetevője – a kognitív, a nyelvi és a kommunikációs – maguk is skálafüggetlen hálózatokat alkotnak, és a kommunikációs folyamat ezek összekapcsolásával írható le. Ezért elengedhetetlen, hogy mindhárom tényezőt együttesen vizsgáljuk (Fóris, 2012).

#### **4.2. Többnyelvű információkeresés**

A több nyelven történő dokumentáció szükségsszerűvé teszi a *többnyelvű információkeresés*-t (EN Cross-Language Information Retrieval – CLIR), amely annyiban különbözik a hagyományos, saját nyelven történő kereséstől, hogy az úgynevezett nyelvi és kulturális akadályokat is át kell hidalnunk valamilyen eszközzel. Alapvetően két megoldás lehetséges: az egyik, hogy a lekérdezéseket lefordítjuk a célnyelvre, a másik, hogy magukat a dokumentumokat fordítjuk le

a forrásnyelvre. Ez három úton valósulhat meg: gépi fordítási technikák útján, tudásalapú technikákkal, illetve korpuszalapú technikákkal (Peters, 1997).

A többnyelvű információkeresés problémájával könyvtártudományi oldalról magyar nyelven Vig Dániel foglalkozott részletesebben előadásában. Ahogy elmondta, többnyelvű információkeresés során az egyik fő feladat, hogy olyan eszközöket fejlesszünk ki, melyek összepárosítják a különböző nyelvű, megegyező kifejezéseket, illetve a legközelebbi hasonló jelentésű szavakkal kapcsolják őket össze (Vig, 2015).

Tehát a könyvtár- illetve információtudomány is felismerte, hogy ez nem egyszerű, gépiesíthető, a nyelvi „felszint” érintő feladat, hanem annál sokkalta bonyolultabb, ha úgy tetszik fordítási, ha úgy tetszik terminológiai, szemantikai, szemiotikai feladat.

A mai kutatások a dokumentumfordítás iránya helyett inkább a lekérdezőfordításra koncentrálnak, annak nagyobb flexibilitása miatt (Nie, 2010: 20). A lekérdezőfordítás legnagyobb bizonytalanságát az jelenti, hogy a lekérdezők általában nem állnak két vagy három szónál többől. Ezért nehéz megtalálni a megfelelő célnyelvi kifejezéseket, terminusokat. Továbbá az információkeresés velejárója, hogy a lekérdező „teljességéhez” kívánatos, hogy a lefordított lekérdező az összes lehetséges jelentést magában hordozza. Ez már az egyszerű egynyelvű információkeresés során is lényegi momentum, de a többnyelvű információkeresésnél még hangsúlyosabb. Tehát a többnyelvű információkeresés nemcsak egy egyszerű fordítási feladat, hanem egy olyan művelet, mely mélyen beépült a keresésbe (Nie, 2010: 22).

#### **4.2.1. Szótáralapú technológiák**

A szótáralapú lekérdező-fordítás, amelyekben minden kifejezést és fogalmat a másik nyelvben lévő megfelelőjével összepárosítanak – vagy ha ez nem lehetséges, akkor annak egy értelmében legközelebb álló, releváns szóval kapcsolják össze –, megfelelő első kísérlet a többnyelvű információkeresésre (Vig, 2015).

Ezek a megoldások azonban mindig gyengébben teljesítenek, mint az egynyelvű keresések. A kutatások szerint a gép által olvasható automata lekérdezőfordítás az egynyelvű keresés pontosságának csupán negyven–hatvan százalékát éri el (Hull–Grefenstette, 1996). Hull és Grefenstette azt is modellezte, mennyivel jobb eredmények érhetők el, ha a lekérdezőfordítás nem kizárólag szóalapú, hanem a többszavas kifejezések nem gépi fordítását is tartalmazza. Így közel 90 százalékos eredmény is elérhető.

Az automatikus lekérdezőfordítás sikertelenségének első és legfontosabb oka, hogy az általános szótárakban nem találjuk meg a szaknyelvi kifejezéseket, esetleg téves, nem naprakész vagy nem kifejezetten célnyelvi megfelelőket találunk, illetve, hogy a többszavas kifejezések nem lefordíthatók kizárólag

általános szótárak használatával. Továbbá némely kifejezést azért sem lehet csak szótári alapon gépesítve lefordítani, mivel a szótár nem fedheti le az összes lehetséges előfordulást. Tehát a kétnyelvű szótárak használatának fő problémái, hogy a fordítás pontossága bizonytalan, nehezen tud megbirkózni a ragozással, a vegyületek neveivel, a kifejezésekkel, a tulajdonnevekkel, a különböző helyesírási változatokhoz tartozó és egyéb különleges szavakkal. Összefoglalva: az általános bizonytalanság és a hiányosság azok a problémák, amelyek a szótáralapú lekérdezésfordítás esetén szinte elkerülhetetlenek. A fordítási bizonytalanságot többféle eszközzel próbálták csökkenteni, többek között többnyelvű ontológiák használatával (l. pl. Abusalah et al., 2009).

#### 4.2.2. Tudásalapú technológiák

A tudásalapú technológiák a többnyelvű információkeresés céljából többnyelvű szótárakat, teauruszokat és ontológiákat alkalmaznak. Ezek a technológiák segíthetik áthidalni a nyelvi jel és jelentés között lévő szakadékot. Abusalah et al. (2009) például egy kétnyelvű ontológiát hoztak létre, mivel az ontológia mint általános ismeretgyűjtemény alkalmas arra, hogy kontextust adjon a lekérdezéseknek, így téve lehetővé a szavak fordításának pontosítását. A rendszer átlagos pontosságát is mérték, eredményük szerint a kétnyelvű ontológiák rendszere jelentősen meghaladja a szótári rendszer precizitását: utóbbi negyvenkét százalékos találati pontosságot ért el, míg az ontológiai rendszer hatvanhárom százalékot (Abusalah et al., 2009: 16). Mégis – egyelőre – **a teauruszalapú rendszerek rendelkeznek az egyik legjobb pontossági aránnyal a többnyelvű információkeresés területén.** A teaurusz egy olyan ontológia, amely terminológiákra specializálódik, és fogalmi kapcsolatokat biztosít. Az ISO 25964-1-es szabványa írja le, hogy hogyan építhető be egy tárgykör tartalma egy többnyelvű teauruszba. A teauruszoknak azon fajtáját, amelynek feladata, hogy egy-egy tudományterületen segítse az információfeldolgozást és az információkeresést, *információs* vagy *információkereső teaurusz*-nak nevezzük (Pálvölgyi, 2011a: 122). Az információkereső teaurusz olyan összetett szótár, mely egy-egy szakterület fogalmait írja le természetes nyelven, továbbá feltárja és feltünteti a legfontosabb logikai kapcsolatokat (Pálvölgyi, 2011a: 123). Kiemelendő, hogy „az osztályozás, indexelés és visszakereshetőség mellett a teauruszok egyben egy szakterület terminológiáját is egyöntetűsítik” (Horváth–Papp, 2001: 97). Ugyanakkor a szakmai fogalmak jelölői, épp úgy, mint a köznyelvi lexikai egységek, folyamatosan változnak, ezért a teauruszcikkek folyamatos ellenőrzésére, nyomon követésére van szükség. Hasznos lenne más, új nézőpontok bevonása is a fejlesztésükbe, ugyanis hosszú távon a teauruszok egyik gyengeségének tekinthető, hogy az egyes fogalmakat lexikai egységekkel

jelölik (szemben például a terminológiai adatbázisokkal, ahol fogalmi azonosítókat használnak).

Az ontológiák lényegében speciális taxonómiák, amelyek amellett, hogy megmutatják a fogalmak egymáshoz képesti fogalmi hierarchiáját, a formális logikai szabályok megjelenítésével még többre is képesek az emberi gondolkodás reprezentálásában (Berners-Lee et al., 2001, idézi Tóth, 2010: 192–193).

Az ontológiák építésének társadalmi motivációi között első helyen áll a tartalmak kereshetőségének, megjelenítési és kreatív újrahasznosítási lehetőségeinek növelése. Ez olyan szótárak fejlesztését teszi szükségessé, amelyek a világ létezőinek lehető legteljesebb szabványos leírására vállalkoznak. Minden ontológia a világ tudásának egy-egy fontos szegmensét reprezentálja, amelyek együttesen egy globális hálózati metaadat-infrastruktúra kiépítésének a lehetőségét hordozza magában (Tóth–Kovács, 2013: 5).

A modern számítógépes ontológiák fejlesztése és különböző területeken való sikeres alkalmazásuk azon a felismerésen alapul, hogy a tudás megszerzése nem elegendő: „ismereteinket csak akkor alkalmazhatjuk a gyakorlatban, ha ismerjük a különböző elemek közötti összefüggéseket és képesek vagyunk leírni azokat”<sup>7</sup> (Fóris, 2013: 422).

A teaurusz és az ontológia közötti különbség – kivált felhasználásukban (információkeresés, tartalomkinyerés, feldolgozás) – nem igazán szembetűnő. Azonban amíg „a teauruszokat és osztályozási rendszereket hagyományosan a dokumentumok (források) tartalmi feltárásához és kereséséhez használják” (Ungváry, 2004), addig „[A]z ontológiákat a szakértői rendszerekben és ismeretbázisokban használják, az információkeresés szemantikailag automatizáltabb támogatására” (Ungváry, 2004). Emellett mindenképpen megemlítendő, hogy míg az ontológiák az általános, alapfogalmak közötti relációkkal foglalkoznak, a teauruszok szűk szakterületek lexikai egységei közötti kapcsolatok megjelenítésére specializálódtak. A lényegi különbség továbbá abban ragadható meg, hogy a teauruszok egyes szakterületre specializált különféle kapcsolatokat (eredmény, ok, hatás-ellenhatás stb.) is definiálnak expliciten.

Kissé leegyszerűsítve, **az ontológia és a teaurusz két fontos dologban tér el: a relációk típusaiban, és a későbbi felhasználásukban**, vagyis hogy szeparáltan, egy szűk terület terminológiájának építését tűzzük ki célul, vagy a szemantikus webbe tervezzük implementálni, és összefűzzük egy általános, csúcsontológiával, esetleg információkinyerésre vagy dokumentumklaszterezésre használjuk.

A teauruszok és az ontológiák közös hátránya, hogy kialakításuk, fenntartásuk és fejlesztésük rendkívül költséges, valamint megfelelő alkalmazásuk alapos szaktudást igényel.

Napjaink fejlesztései a hibrid technológiák irányába mutatnak, ennek a tendenciának köszönhető az úgynevezett **ontoterminológia** megjelenése is. Az ontoterminológia azért jöhetett létre, mert az ontológia és a terminológia célja egyaránt az információ és tartalom közvetítése, kommunikáció az ember és ember vagy ember és gép között (Roche, 2012). Az ontoterminológia elméletének kidolgozásával Christopher Roche a terminológiai adatbázisok készítésének módszertani alapelveit is új alapokra helyezte.

Roche élesen elhatárolja a terminológiai és a fogalmi rendszert (Roche, 2006, 2009, 2012). A fogalom és terminus éles elválasztása együtt jár a fogalom- és terminusdefiníció megkülönböztetésével, így Roche, a maga által fejlesztett ontoterminológiai adatbázisában, egyszerre kétféle definíciót is használ, úgymond egyet a géppel, és egyet az emberrel való kommunikációhoz. Ennek szemléltetésére a klasszikus szemantikai háromszög helyett bevezet egy úgynevezett dupla szemantikai háromszöget is (lásd részletesen Roche, 2006, 2009, 2012).

A jelentés és a kapcsolatszerkezet azonosításához adott a *jelölt* (*signified*), amely nem feleltethető meg a *fogalomnak* (*concept*), azonban kapcsolódnak egymáshoz, ahogy a *jelölő* (*signifier*) és az *azonosító* (*identifier*) is. A fogalmat magát a fogalom logikai specifikációja identifikálja (Roche et al., 2009 [2011]), azaz a hierarchikus kapcsolatok; tehát egy fogalom egy rendszer egészében létezik, önmagában nem. Utóbbi megegyezik a terminológia nézőpontjával, miszerint a terminusoknak különböző környezetekben (kontextusban, doménben) különböző szemémái, azaz fogalmi jegyei lehetnek. „Egy terminus összes jegye a terminológiai háló különböző helyein való előfordulása alapján adható meg” (Fóris, 2005: 39).

Látszólagos bonyolultságuk ellenére már az 1990-es évek legelején elkezdődött a vegyes (és tudásalapú) technológiák alkalmazása a többnyelvű információkeresésben, amelynek egyik leghíresebb példája az EMIR (European Multilingual Information Retrieval, azaz Európai Többnyelvű Információkeresés) projekt. Az EMIR céljairól a projekt weboldalán kaphatunk részletes információkat. Az 1990 novemberétől 1993 novemberéig tartó projektben egy olyan indexálási rendszer kidolgozását tűzték ki célul, amely lehetőséget ad a felhasználóknak arra, hogy különböző nyelvű szövegeken saját nyelvű lekérdezéseket futassanak, illetve, hogy egy bizonyos nyelvű szöveget több különböző nyelven is lekérdezhessenek.

Az EMIR bizonyítja a többnyelvű adatbázisokban tett többnyelvű lekérdezések megvalósíthatóságát, beleértve a többnyelvű dokumentumokat is (Peters 1997). A kísérlet rangsorolós Boole-i műveleteket használ különböző típusú (pl. idióma) szótárakkal együtt. Fontos kiemelni, hogy ehhez természetesen szükségesek a tárgykörhöz tartozó fogalomtárak, és az alapos manuális szerkesztés is.

## 5. Összegzés

A releváns (információtudományi, dokumentációs, terminológiai) szakirodalomra támaszkodva megállapítottuk, hogy az információtudomány feltárási eszközei (tezauruszok, ipari alkalmazások, ontológiák) és a terminológia tudománya is az információk feltárásával, tárolásával és újrafelhasználásával foglalkoznak. A dokumentumok indexálásához használt szabványosított terminológia lehetővé teszi, hogy a dokumentáció szisztematikus és egyértelmű legyen. Az információkereső nyelvek (szerencsés esetben) szabványosított terminológiai rendszerek.

A tudásalapú technológiák segíthetnek áthidalni a nyelvi jel és jelentés között lévő szakadékot. Az információkinyerés automatizálásának továbbra is elsődleges problémája, hogy magyar nyelven hiányoznak az úgynevezett szakterületekre jellemző gyakorisági szótárak, amelyek egy-egy szakterület szókincsét gyakorisági mutatókkal ellátva közölnék.

Az egyes tudományágak közötti szorosabb együttműködés elősegíthetné a minél sikeresebb „hibrid” technológiák kidolgozását. A továbbiakban célszerű lenne annak vizsgálata is, hogy az egyes területek hogyan tudják szolgálni egymást akár a többnyelvű dokumentációkezelés (illetve fordítás) vagy akár a mesterséges intelligencia kutatásának területén.

A többnyelvű dokumentáció a mindennapokban is szükségszerűvé teszi az ezen rendszereken belüli folyamatos terminológiai munkát, terminológiai harmonizációt is (erről részletesen I. Fóris–Sermann, 2010).

## Irodalom

- Abusalah, M., Oakes, M., Tait, J.** (2009) Cross language information retrieval using multilingual ontology translation and query expansion base. *Polibits* 2/40. pp. 13–16. Web: <http://www.scielo.org.mx/pdf/poli/n40/n40a3.pdf>
- B. Hajdu Á., Babiczky B.** (1998) *Bevezetés az információkereső nyelvek elméletébe és gyakorlatába.* Budapest: Universitas Kiadó.
- Balázs S.** (1960) Magyarország képviselője és szerepe a FID-ben. *Magyar könyvszemle* 1960/1. 58–64. Web: [http://epa.oszk.hu/00000/00021/00239/pdf/MKSZ\\_EPA00021\\_1960\\_76\\_01\\_054-064.pdf#page=5](http://epa.oszk.hu/00000/00021/00239/pdf/MKSZ_EPA00021_1960_76_01_054-064.pdf#page=5) (Megtekintés: 2017. 06. 29.)
- Cabré, M. T.** (1999) *Terminology: Theory, Methods and Applications.* Amsterdam: John Benjamins.
- Fóris Á., Faludi A.** (2017) A dokumentáció és a dokumentum-menedzsment kapcsolatai, összefüggései a terminológiával és a fordítói gyakorlattal. *Fordítástudomány*, megjelenés alatt.
- Fóris, Á., Sermann, E.** (2010) A terminológiai szabványosítás és a terminológiai harmonizáció. *Magyar Terminológia* 3/1. 41–54.
- Fóris Á.** (2005) *Hat terminológia lecke.* Pécs: Lexikográfia Kiadó.
- Fóris Á.** (2013) Network Theory and Terminology. *Knowledge Organization* (International Journal) 40/6. pp. 422–429.
- Fóris Á.** (2014) A szaknyelvek terminológiai szempontú megközelítése. In: Veszelszky Á. és Lengyel K. (szerk.) *Tudomány, technolektus, terminológia – a tudományok, szakmák nyelve.* Budapest: Éghajlat Könyvkiadó. 27–36.
- Horváth T., Papp I.** (szerk.) (2001) *Könyvtárosok kézikönyve* 2. kötet. Feltárás és visszakeresés. Budapest: Osiris.

- Horváth T.** (1999) A könyvtártudomány és információtudomány alapjai. In: Horváth T. és Papp I. (szerk.) *Könyvtárosok kézikönyve*. 1. kötet. Alapvetés. Budapest: Osiris. 17–75.
- Hull, D., Grefenstette, G.** (1996) Query across languages: A dictionary-based approach to multilingual information retrieval. In: *Proceedings of the 19th ACM SIGIR Conference*. pp. 49–57. Web: [https://www.academia.edu/25198572/Querying\\_across\\_languages](https://www.academia.edu/25198572/Querying_across_languages) (Megtekintés: 2017. 07. 23.)
- Kéglí F.** (2009) Hatvan éve alakultak meg a szakmai dokumentációs központok. *Könyvtári Figyelő* 55/3. 445–457. Web: <http://ki.oszk.hu/kf/2010/10/hatvan-eve-alakultak-meg-a-szakmai-dokumentacios-kozpontok> (Megtekintés: 2017. 07. 23.)
- Koblitz, J.** (1959) Dokumentation und Information. Eine Terminologische Untersuchung. *Dokumentation* 6/1. pp. 3–10.
- Koltay T., Prókai M.** (2010) Terminológiai változások a XX-XXI. századi könyvtártudományban. *Magyar Terminológia* 3/2. 269–284.
- Lengyelne Molnár T.** (2008) *Automatikus referátumkészítés*. Doktori disszertáció. ELTE BTK IKT.
- Nie, J.-Y.** (2010) *Cross-language information retrieval*. San Rafael, Calif: Morgan & Claypool. Web: <http://disi.unitn.it/~bernardi/Courses/DL/clir.pdf> (Megtekintés: 2017. 07. 23.)
- Otlet, P.** (1934) *Traité de documentation, le livre sur le livre. Théorie et Pratique, Vol. I*. Bruxelles: Editiones Mundaneum, Palais Mondial.
- Pálvölgyi M.** (2011a) *Információkereső nyelvek I*. Eger: Eszterházy Károly Főiskola. Web: [http://www.tankonyvtar.hu/hu/tartalom/tamop425/0005\\_11\\_infkereso\\_nyelvek\\_i\\_scorm\\_03/adatok.html](http://www.tankonyvtar.hu/hu/tartalom/tamop425/0005_11_infkereso_nyelvek_i_scorm_03/adatok.html). (Megtekintés: 2017. 07. 23.)
- Pálvölgyi M.** (2011b) *Információkereső nyelvek II*. Eger: Eszterházy Károly Főiskola.
- Peters, C.** (1997) Across languages, across cultures – Issues in multilinguality and digital libraries. *D-Lib magazine* 3/5. Web: <http://www.dlib.org/dlib/may97/peters/05peters.html> (Megtekintés: 2017. 07. 23.)
- Polzovics I.** (1959) Dokumentáció és információ. *Magyar Könyvszemle* 75/4. 325–332.
- Polzovics I.** (1961) *A dokumentáció, az információ és a tájékoztatás fogalma. Terminológiai vita*. Budapest: OSZK Könyvtártudományi és Módszertani Központ.
- Polzovics I.** (1962) *Bevezetés a szakirodalmi dokumentációba a műszaki és természettudományok terén*. Budapest: OMKDK.
- Roche, Ch.** (2006) Lexical and Conceptual Structures in Ontology. In: *Proceedings of Advances in Applied Artificial Intelligence, 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2006*. Annecy, France, June 27–30, 1034–1041.
- Roche, Ch.** (2012) Ontoterminology: How to unify terminology and ontology into a single paradigm. In: Calzolari, N. & Choukri, K. & Declerck T. & Uğur Doğan, M. & Maegaard, B. & Mariani, J. & Odijk, J. & Piperidis, S. (eds.). *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) 2012*. May 23–25, Istanbul, Turkey. European Language Resources Association (ELRA). 2626–2630. Web: [http://www.lrec-conf.org/proceedings/lrec2012/pdf/567\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/567_Paper.pdf). (Megtekintés: 2017. 07. 23.)
- Roche, Ch., Calberg-Challot, M., Damas, L., Rouard, Ph.** 2009 (2011) Ontoterminology – A New Paradigm for Terminology. In: *KEOD 2009 – Proceedings of the International Conference on Knowledge Engineering and Ontology Development*. Funchal–Madeira, Portugal, 2009. October 6–8, 321–326. Web: [http://hal.archives-ouvertes.fr/docs/00/62/21/32/PDF/Proceedings\\_KEOD\\_2009\\_128\\_CR.pdf](http://hal.archives-ouvertes.fr/docs/00/62/21/32/PDF/Proceedings_KEOD_2009_128_CR.pdf). (Megtekintés: 2017. 07. 23.)
- Szépe Gy.** (szerk.) (1967) *A Tudományos Tájékoztatás Elmélete és Gyakorlata II. Nyelvfeldolgozás és dokumentáció*. Budapest: Országos Műszaki Könyvtár és Dokumentációs Központ.
- Tod = Wersig, G., Neveling, U.** (1967) *Terminology of documentation*. Paris: The Unesco Press.
- Tóth M.** 2010. A szintaktikai és szemantikai paradigma. Egy globális hálózati metaadat-infrastruktúra felé. *Magyar Terminológia*. 3/2. 183–213.

- Tóth M., Kovács Z. Cs.** (2013) *Az ontológiaépítés módszertana*. Egyetemi tankönyv. Debrecen. Web: <http://docplayer.hu/1361325-Az-ontologiaepites-modszer-tana-toth-mate-es-kovacs-zoltan-csaba.html>. (Megtekintés: 2017. 07. 23.)
- Ungváry R.** (2004) Tezaurusz és ontológia, avagy a fogalmi ismertetőjegyek generikus öröklődésének formalizálása. *Tudományos és Műszaki Tájékoztatás Könyvtár- és Információtudományi szakfolyóirat*. 51/4. Web: [http://tmt.omikk.bme.hu/show\\_news.html?id=3615&issue\\_id=450](http://tmt.omikk.bme.hu/show_news.html?id=3615&issue_id=450). (Megtekintés: 2017. 07. 23.)
- Ungváry R., Vajda E.** (2002) *Könyvtári információkeresés*. Budapest: Typotex.
- Verók A.** (2011) A dokumentáció. In: Ottovay L. (átdolgozó: Verók A.): *Az általános tájékoztatás eszközei*. Eger: Eszterházy Károly Főiskola.
- Vig D.** (2015) *Nyelveken és kultúrákon átívelő többnyelvű információkeresés*. Elhangzott: XXXII. Országos Tudományos Diákköri Konferencia. Pedagógiai, Pszichológiai, Andragógiai és Könyvtártudományi Szekció. Debreceni Egyetem. Debrecen, 2015. 04. 15.

### Internetes források

web1: [searchsoftwarequality.techtarget.com/definition/documentation](http://searchsoftwarequality.techtarget.com/definition/documentation)

web2: [http://konyvtar.hu/wiki/Fédération\\_Internationale\\_de\\_Documentation#cite\\_note-0](http://konyvtar.hu/wiki/Fédération_Internationale_de_Documentation#cite_note-0)

EMIR (European Multilingual Information Retrieval).

[http://cordis.europa.eu/project/rcn/8780\\_en.html](http://cordis.europa.eu/project/rcn/8780_en.html)

### Jegyzetek

---

<sup>1</sup> Eredeti szöveg: „The continuous and systematic processing of documents or data, including e.g. location, identification, 'acquisition', analysis, storage, retrieval, circulation and preservation for the specialized information of users.” (Ford.: F. A.)

<sup>2</sup> Eredeti szöveg: „Documentation c'est réunir, classer et distribuer des documents de tout genre dans tous les domaines de l'activité humaine.” (Ford.: Verók 2011)

<sup>3</sup> Eredeti szöveg: „Collection et conservation, classification et sélection, dissémination et utilisation de toute information.” (Ford.: F. A.)

<sup>4</sup> Eredeti szöveg: „La documentation consiste à prendre connaissance de ce qui a été dit d'original ou d'important, sur une question.” (Ford.: Verók 2011.)

<sup>5</sup> Eredeti szöveg: „An 'artificial language' used by documentation systems for purposes of indexing, storage and retrieval.” (Ford.: F. A.)

<sup>6</sup> Eredeti szöveg: „A documentary language which groups concepts into classes, uses mostly hierarchical relations for arrangement of classes, and reflects the structure of arrangement by notation.” (Ford.: F. A.)

<sup>7</sup> Eredeti szöveg: „The development of modern computerized ontology and its successful practical application in various fields is based on the recognition that obtaining knowledge is not enough: we can apply our knowledge in practice only if we know the connections among its various elements and are able to describe them.” (Ford.: F.A.)