

A dokumentumban számos hivatkozást találunk, amelyek a részleteket tárgyalják, és az előző verziókra mutató csatolók lehetőségét adnak a történeti áttekintésre is.

/OWL Web Ontology Language, 2005. április 25.
<http://www.w3c.hu/forditasok/OWL/REC-owl-features-20040210.html/>

(Varga Anett)

Tezauruszok, taxonómiák és ontológiák – etimológiai megjegyzés

Gilchrist rövid dolgozatát a terminológiai zűrzavar enyhítésére írta, egyben bizonyos prognosztizálást lehetővé téve. Elsőként az *Oxford English Dictionary*-ből idézi a címben szereplő terminus technicusok meghatározását.

Tezaurusz

A tudás „kincses háza” vagy „raktára”, miként egy szótár vagy enciklopédia. Értelem szerint rendezett szavak vagy fogalmak gyűjteménye, antinómiák és szinonimák szótára.

Taxonómia

Osztályozás, főként általános törvényeivel és alapelveivel összefüggésben, amely a tudománynak, tudományágnak vagy témának része; lényege az osztályozás, vagy kapcsolatos az osztályozással, az élő szervezetek módszeres osztályozásával.

Ontológia

A lét tudománya, a metafizika része; a léttel, a dolgok lényegével vagy a léttel absztrakt voltában kapcsolatos.

Mindhárom definíció sokak számára ismerős, de az információtudomány dolgozói kissé zavarossá tették. A három között átfedések, néha ellentétek jöttek létre. Nincs értelme például e háromnak „korrekt” jelentésével nagyképszerűsüdni, de hármójuk különbségei és hasonlóságai gazdagítják a használókat.

A szerző a LISA-ban megnézte e szavak előfordulását:

tezauruszra	2313,
taxonómiára	285,
ontológiára	163 találat volt.

Korábban (1997-ben) pl. az ontológia szó az új használatában nem is fordult elő. Ám azóta a taxonómia és ontológia szó előfordulása növekszik mind a publikált, mind a szürke irodalomban, leggyakrabban az információkeresés, tudásmedzsment, navigálás szavak társaságában.

Tezauruszok

A tezauruszról legtöbbünknek *P. M. Roget* Thesaurusa jut eszébe, aki munkáját 1852-ben jelentette meg először. Az angol nyelv szinonimaszótárát készítette el, amelyben „műfajt” teremtett. Szerkezete – Roget „verbal classification”-nek hívja – a szavak és frázisok jelentésének hasonlóságán alapszik. (Magyarul talán a szóbeliségen nyugvó osztályozásnak lehetne nevezni.) Roget a természettörténet akkor még uralkodó iskolájához igazodik, amely az élőlények rendszerezésében szintén a hasonlóságot tekintette mérvadónak.

Roget méltatói és kutatói egymástól függetlenül közvetlen kapcsolatot vélnek felfedezni olyan filozófusokkal, mint *Descartes*, *Bacon*. Közvetlen elődjének pedig a XVII. század angol nyelvészét, *J. Wilkinst* tekintik, aki a nyelvi kétértelműséget kívánta leküzdeni egy, a latin helyére kimunkált „filozófiai” nyelvvel, célja az egyetemes kommunikáció elősegítése volt.

Vickery 1960-ban írta, hogy a mechanizált tezauruszok problémái a rokon jelentések hálózatán nyugszanak. Az „osztályozás” szó használata óvatosságot igényel, mert az első tezauruszok nem alkalmazták az osztályozási technikát. A *tezauro-facet* volt e tekintetben az első lépés 1969 körül, illetve a fazettás osztályozás és a tezauruszok megjelenése ugyanabban a kötetben.

Most már előttünk állhat a klasszikus információs (kereső) tezaurusz: ellenőrzött szótár, amely indexelésre és visszakeresésre használatos. Kezdetben lyukkártyákhoz alkalmazták, a számítógépes munkákban pedig batch üzemmódban. A tezaurusz tovább él az online világban. Használatának stabilitásához hozzájárult „nemzetköziesítése” az ISO révén, *J. Aitchison* és társainak kézikönyve, és más fontos eszközök. Az online korszak azt a problémát hívta elő, hogy létre kell hozni olyan eszközt, amely különböző tezauruszok között létesít kapcsolatot a *term*-ek közötti hierarchikus, asz-

szociatív és ekvivalenciarelációk hálózatával. Egy ilyen köztes vagy közvetítő nyelvre alkalmas példa a nálunk jól ismert BSO. Mindenesetre a probléma akuttá vált. Jelenleg is széles körű kutatások ismeretese az adatbázisok közötti keresésekre.

A fentiekkel párhuzamosan a mesterséges intelligencia és a szakértői rendszerek területén ún. „szemantikai hálózatokat” dolgoztak ki. Ezeket J. Milstead ismertette 2000-ben: „Fogalmilag eléggé közel állnak a teauruszokhoz, a hálózatok mutatják a term-ek szemantikai kapcsolatait a szövegekben.” Jó példa erre N. Ford könyvtárosoknak szánt tankönyve, illetve annak egy fejezete, szemantikai térképe a kardiovaszkuláris betegségekről. A fenti eredmények irányították a figyelmet a taxonómia felé.

Taxonómiák

Mi az oka a taxonómia kifejezés megjelenésének az új, kibővült jelentéssel?

- Az *információs túlterheltség* (hagyományos be rendezések nem hatékonyak nagy adatbázisokhoz).
- *Információs műveltség* (főként a végfelhasználók küszködnek tudásbeli hiányokkal).
- A *szervezés terminológiája* (a közzétett osztályozási eszközök és teauruszok nem igazodnak a szervezetek nyelveihez, amelyek 80%-a belsőleg kapható meg).
- A *szervezés „rombolása”* (a bővítések, adatbázisok egyesítése, a létrehozás különböző szakaszaiban kulturális kérdéseket vetnek fel). Az eltérő használói csoportok megosztják az információt és tudást a szervezetben.

Ezek kezelésére a taxonómia és eszközei valók. Egy 2000-ben közzétett tanulmány szerint a taxonómiának öt eltérő, számos érintkezést mutató jelentése volt:

- *Webcímjegyzékek*, amelyeknek hálózatközi használata alakult ki. Valójában az osztályozás egy formája, csúcskifejezések többszintű hierarchiájával. Közülük lehet kiválasztani azt a néhány term-et, amely egy-két géphez tartozik. Jó példa erre az Open Directory Project, amely mintegy 350 ezer (!) term-et kezel, létrehozásában 35 ezer önkéntes vett részt.
- A taxonómiák *támogatják az automatikus indexelést*.
- A taxonómiák *automatikus kategorizálást hoznak létre*.

- Az *előkészületek befejezése*. Kérdésformula létesítése: homonimák egyértelműsítése, szinonimák csoportosítása (klaszterálása), a keresés feladása hálózatokra.
- *Testületi taxonómiák*. Eszközök, pl. útmutatók, szavak térképei, vagy a létező teauruszok összeosztása megatezauruszá.

Ontológiák

Vickery 1997-ben tanulmányt tett közzé az akkori ontológiákról és vezető gondolkodóiról. Innen az idézet: „Az ontológiát úgy kell meghatározni, mint a megosztott konceptualizálás formális, explicit leírását.” Ehhez kell egy kis magyarázat.

Konceptualizálás: a világ jelenségeinek (fenomenon) absztrakt modellje azáltal, hogy e fenomenonnak releváns fogalma (concept) határozható meg.

Explicit: a használt fogalmak típusát jelenti, a használat megkötöttségeit, és ezeket határozottan definiálják.

Formális: géppel olvasható (szabályok szerint leírt, előírással).

Megosztott: (shared) arra vonatkozik, hogy egy ontológia magáévá tette a megegyezéssel tudást, azaz nem személyi vagy individuális, hanem egy közösség fogadta el.

Példái közül a WordNet említhető, amelyet a Stanford University Kognitív Tudományok Laboratóriuma fejlesztett. Kb. 100 ezer szó jelentését tartalmazza 5 kategóriába sorolva: főnevek, igék, melléknevek, határozószók és „szerep” szavak. A jelentések kapcsolatban állnak szinonimákkal, antonimiákkal (ellentétes jelentések), hiponimákkal (létezik valamiként, angolul „is-a...”) meronimákkal (része kapcsolat fogalmak között), és a szóalakok morfológiai relációival. A WordNet tehát szótári eszköz, és az interneten szabadon érhető el. Olyan szakterületeken építenek ontológiákat, mint a gépi fordítás, ismeretek újrahaznosítása, információkeresés stb. Legambiciózusabb felhasználása a tudásmenedzsmentben jött létre.

Az ontológiai eszközöknek három fajtája rajzolódik ki:

- A *szervezési* ontológia leírja az információ metamodelljét, például az információs források szerkezetét, formátumát.
- A *szakterületi* ontológiák az információs források tartalmát írják le.

- A *vállalkozási* ontológia az üzleti folyamatokat modellálja.

Befejezésül hozott példája *T. Berners-Lee* és társainak 2001-ben javasolt szemantikus webje, amelyet ma a legfejlettebbnek tart, jól definiált jelentésekkel az ember és a gép jobb együttműködésével. Gilchrist szinte átszellemülten ír róla, illetve idéz, a nézeteit is általánosítja. A szemantikus web egyetlen és decentralizált rendszer; nem ismeri a diszkriminációt pl. egy hevenyészett vázlat és kiglancolt előállítás között, a hétköznapi és akadémiai információ, kultúrák, nyelvek, médiák között. Berners-Lee nyomán az ontológiákat úgy határozza meg, hogy az „egy dokumentum vagy fájl, amelyet formálisan a term-ek közötti relációk határoznak meg”.

Konklúzió

Wittgenstein szerint úgy lehet megismerni egy szó jelentését, hogy megvizsgáljuk használatát. Vizsgálva a teauruszok, taxonómiák és ontológiák használatát, kitűnik, hogy köztük sok az átfedés. A zavar lehetőségét növeli az is, hogy a részt vevő szereplők mások és mások. Egy poszt-roget-i teaurusz az információtudomány képviselőire tartozik, a taxonómia az információtechnológusok és rendszerfejlesztők ügye, az ontológiákat a filozófusoktól veszik, akik egyebek mellett mesterséges intelligenciával foglalkoznak. Néhány tényező azonban kiemelhető:

- a szervezésben növekvő trend a külső és belső információ összeegyeztetése,
- jelenleg az információk óriási tömege áll rendelkezésre,
- elérhető a viszonylag olcsó számítógépes erőforrás.

A teauruszok főként a belső információra figyeltek. A taxonómia mind a külső, mind a belső információ hatalmas mennyiségű inputjaival számol. Az információ minősége is a taxonómiát kívánja a tartalmak automatikus elemzésében. Az ontológiák világában a számítógépi erőforrások kiemelkedő fontosságúak.

Az is közös vonása a teauruszoknak, taxonómiáknak és ontológiáknak, hogy a természetes nyelvvel foglalkoznak. Ha adva van számos tárgy, csoportokba rendezésük lehetséges, ahogy a könyvtárak hagyományosan csinálták. A deskriptorok címkézhetők, csoportokba sorolhatók, erre születnek a csoportokat jelentő fogalmak és tezaurofacetták. A taxonómia él az osztályozás és tezausztechnika lehetőségeivel az automatikus kategorizálásban is. Alkalmazhatók a tárgyak nagy száma esetén is. Az osztályozás lehetőséget teremt a fogalmak közötti kapcsolatok elemzésére, az asszociációkra. A szemantikai hálózatok pedig többet is jelentenek, mint amit a relációkról a teauruszok nyújtanak. A vizsgálatok kiterjednek a szavak széles körére, és mivel szövegfüggőek, többet adnak egy definíciónál vagy használatra vonatkozó megjegyzésnél is. Feltűntetik a szóalakokat, a morfológiai változatokat, a szintaktikai helyzetüket. Így jönnek létre a „gazdagított” teauruszok. Ezek az ontológiák alapvető összetevői. Az is látható, hogy tudományközi csoportok szükségesek az olyan álmokhoz, amilyenek a szemantikus web bizonyult. (Felhívom a figyelmet arra, hogy Gilchrist tanulmánya végén e téma szakirodalmának legjavát nyújtja. – A ref.)

/GILCHRIST, Alan: Thesauri, taxonomies and ontologies – an etymological note. = Managing Information, 2004. május, p. 44–48./

(Horváth Tibor)

Tárgyszavazás Svédországban

Hogyan készül nemzetközi szabványok alapján nemzeti tárgyszójegyzék egy olyan országban, amely mindig is a saját útját akarta járni?

Svédországban hosszú története van a tárgyszavazásnak. A *Királyi Könyvtár (Royal Library = a svéd nemzeti könyvtár)* már az 1920-as években létrehozott egy korszerű, a mai elveknek is megfelelő tárgyszókatalógust. Ugyanebben az időben az ország úgy döntött, hogy saját osztályozási rend-

szert alkot, és ezzel nem követi a többi skandináv ország útját, amelyek vagy a Dewey-t vagy az ETO-t alkalmazták. Az osztályozási rendszer mintegy fél évszázadra elterelte az érdeklődést a tárgyszavazásról.

Az 1970-es években jött létre a svéd szakkönyvtárak közös katalógusa (*LIBRIS*). A részt vevő könyvtárak egy része erőteljes propagandát fejtett ki a tárgyszavazás népszerűsítése érdekében.