

REFERÁTUMOK

Információkereső nyelvek kompatibilitása

1982. október 17-én Columbusban (Ohio, USA) nemzetközi konferenciát rendeztek az információkereső nyelvek kompatibilitásával kapcsolatos kérdések megvitatására. Az alábbiakban E. Svenonius bevezető előadását, illetve I. Dahlberg – a bevezetőben felvetett néhány problémáját körüljáró – hozzászólását ismertetjük.

A konferencia alapkérdése

A konferencia témája, némileg leegyszerűsítve, egyetlen kérdésben összegezhető: mennyire lehet két információkereső nyelv (IKNY) kompatibilis? Ez az alapkérdés azonban – különböző oldalról vizsgálva – más és más problémát vet fel. Az összetett adatbázisok szempontjából a teauruszok, az osztályozási rendszerek, illetve a keresőkérdés-típusok; a használók és a szöveges adatbázisok interakciójában a kérdés és a releváns választ alkotó szövegrészek megfeleltetését jelenti. Azaz, általánosítva, a konferenciának arra kell választ adnia: mennyire hangolható össze két nyelvhasználati modell?

A probléma megoldása csak nyelvészeti elmélet talán képzelhető el és rendkívüli gyakorlati jelentőséggel bír, hiszen ma már a felhasználói igények megközelítően teljes kielégítése is csak több adatbázis lekérdezésével oldható meg. Ez pedig – mivel az adatbázisok eltérő természetes, utasítás- és indexelő nyelvet használnak – elengedhetetlenné teszi a kérdés többszöri lefordítását. Az ideális megoldás, mind a gazdaságosságot, mind a jelentés „objektív” átvitelét tekintve, az lenne, ha a kérdező saját természetes nyelvén megfogalmazott kérdését az információkereső rendszer automatikusan ültetné át az egyes adatbázisok szókészletére, kiküszöbölve a közvetítőt, és algoritmikussá téve a keresést.

Történeti visszapillantás

A hatvanas évek elején az osztályozásméleti szakemberek még különbséget tettek az IKNY-ek konvertálhatósága (az egyik szótárról a másikra való áttérés lehetősége), illetve az információkereső rendszerek kompatibilitása (az egyik rendszer indexelő kifejezéseinek és referátumainak közvetlen átvételi lehetősége) [1] között. A kompatibilitást átkapcsoló nyelvek kialakításával [2], vagy szótárak lefordításával próbálták elérni [3]. Lassan azonban csökkent az ez irányú érdeklődés, és egészen addig ez volt a jellemző, amíg – a hetvenes évek végére –

a számítógéphasználat elterjedése és olcsóvá válása; az automatikus nyelvfeldolgozás szintaktikai és szemantikai problémáinak pontosabb megértése; Európában a nyelvi határok leküzdésének igénye; az USA-ban egy integrált, országos, hiteles tárgyszójegyzék könyvtári célú létrehozásának kísérlete, valamint a gépi fordítás és a kompatibilissá tétel megoldására irányuló erőfeszítések növekvő anyagi támogatása ismét fellendüléshez vezetett.

A jelenlegi kutató–fejlesztő munkálatok

A kompatibilitás iránti megerősödött érdeklődést 1977–1978-ban több konferencia jelezte, amelyek közül „Az információkereső nyelvek egységesített rendszere” című, latviai (1977. szeptember 6–8.) volt talán a legjelentősebb.

Európában a gyakorlati szakemberek többnyelvű teauruszok kialakítását tekintik elsődleges feladatuknak (olyan jelentős programok keretében, mint például az Euronet számítógépes terminológiai rendszere, az Eurodicautom; a brit PRECIS többnyelvűvé tétele vagy a francia TITUS II. dokumentációs IKNY-e), de foglalkoznak közvetítő és átkapcsoló nyelvek kidolgozásával is (ilyen például az UNISIST által kialakított Broad System of Ordering, vagy a V. HORSNELL irányításával készített *Intermediate Lexicon for Information Science*).

Az amerikai kontinensen folyó munkálatok iránya némileg eltérő: összetett adatbázisok automatikus átkapcsolási eljárásainak kidolgozására, meglévő szótárak integrálására stb. törekednek. A Massachusetts Institute of Technology pedig közös utasításnyelvet (CONIT) dolgozott ki a DIALOG, a MEDLINE, az Intrex és az ORBIT összekapcsolására, amely részlegesen ellenőrzött szótárral párosul. Az országosan egységesített hiteles tárgyszójegyzék (subject authority file) kialakítására tett kísérletek is ígéretesek, melyekben több meglévő fájl összevonásával próbálkoznak.

Nehézségek

Sajnálatos módon elhanyagolták a kompatibilitás kialakításának *hatékonysági és gazdaságossági vonatkozásait*, és ez kárbavesztett erőfeszítésekhez vezetett, mint például néhány korábbi gépi fordítással kapcsolatos munka esetében. Gondoskodni kell tehát a kompatibilitás kialakítására irányuló munkálatok értékeléséről.

Gyakran hiányos a kompatibilitás-programok *elméleti (nyelvészeti) háttere*. Meg kell vizsgálni például azt, hogy az átkapcsoló nyelveknek valóban specifikusabbaknak kell-e lenniük szókészletük tekintetében a lefordítandó nyelveknél, és hogy a szemantikai elemzés (tényezőkre bontás) valóban megfelelő eljárás-e az átkapcsoló nyelvek kialakítására.

További elméleti nehézséget jelent az, hogy a többnyelvű teauruszok és az automatikus átkapcsoló eljárások megkövetelik az *ekvivalencia operatív meghatározását*. A különböző mélységű hierarchikus szerkezetekkel dolgozó rendszerek közötti fordítás során pedig ún. generikus ekvivalencia megállapítására van szükség, ami bizonyos fokú információvesztéssel eredményez. A konverziós mechanizmus kialakításakor ezt is figyelembe kell venni.

Az indexelő nyelvek általában igen egyszerű *szintaktikai eszközökkel* operálnak, és ez újabb problémát okoz: nevezetesen azt, hogy a szintaxis jelentés-pontosító szerepe nem támogatja a „fordítót”. Az indexelő nyelvek szintaktikai jellemzői ismét mérhetők, és vizsgálandó ezek összefüggése a konverziós lehetőségekkel.

*

I. Dahlberg hozzászólása, amely rövidítve hangzott el a konferencián, a kompatibilitás fogalmi oldalával foglalkozik, így mind az elméleti alapot, mind az utolsó előtti probléma – az ekvivalencia definíció operatív tételének – megoldását elősegíti, méghozzá általánosítva, az összes rendező rendszerre vonatkoztatva.

Mit értünk a rendező rendszerek kompatibilitásán?

Rendező rendszer minden olyan eszköz, amely ismeretek szervezésére, leírására (indexelésére) és visszakeresésére szolgál; a fogalmakat és a közöttük fennálló relációkat szavakkal, illetve jelzetekkel fejezi ki; és ezeket az elemeket rendezetten tükrözi. Ilyenek az osztályozási rendszerek, a teauruszok, a tárgyszójegyzékek stb. A kompatibilitás a rendező rendszer ama tulajdonsága, amely *lehetővé teszi, hogy egy rendszer elemeit a másikéval együtt vagy azokkal felcserélhetően használják*. A kompatibilitásra számos esetben – egy rendszer kiegészítésekor, kooperatív osztályozó és információkereső vállalkozásokban stb. – szükség lehet. A rendszerek összevetése mindeddig a kifejezések és ezek „implicit” jelentésének összehasonlításával történt. Ez azonban nem vezet messzire, hiszen közsímet, hogy a kifejezések mennyire tökéletlen hordozói a fogalmi tartalomnak. A rendező rendszerek összevetése, korrelálása tehát nem korlátozható a nyelvi szintre. A feladat megoldásához rögzíteni kell a külvilág tárgyai és/vagy

ezek szellemi leképezései, és az ezeket jelölő nyelvi elemek kapcsolatát. E kapcsolatokkal a fogalomelmélet foglalkozik, így a kompatibilitás kérdéseiben is a fogalmak ismeretére kell támaszkodni.

Mi a fogalmi kompatibilitás?

A fogalmak formális szempontból az *1. ábra* szerint ábrázolhatók, ahol a háromszög csúcsai a fogalmak összetevőit, élei az ezek között fennálló kapcsolatokat mutatják. A BA reláció, a referenciára vonatkozó állítások, hozzák létre a jellemzőket. A szükséges jellemzők strukturált együttese alkotja a referencia-orientált definíció ama részét, amely a meghatározott fogalom intenzióját és extenzióját behatárolja. A predikációk nyomán keletkező ismeretelemek összessége egy ismeretegységnek tekinthető. A fogalmak meghatározásának tehát egy-egy értelmű megfelelést kell létesítenie a referencia (A), a jellemzők (B) és a nyelvi kifejezés (C) között.

A rendező rendszerek elemei között a fogalmi kompatibilitás három fokozata állhat fenn:

- a fogalmi egybeesés – ekvivalencia – ha két fogalom valamennyi jellemzőjében megegyezik;
- a fogalmi megfelelés – hasonlóság – ha két fogalom számos jellemzőjében megegyezik;*
- fogalmi korreláció – két fogalom viszonyítása a korreláció típusának jelzésével (pl. matematikai szimbólumok segítségével).



1. ábra A fogalmi háromszög

A fogalom-rekord és a fogalmi halmaz

A rendező rendszerek kialakításakor célszerű a fogalmakkal kapcsolatos ismereteket egy-egy *fogalom-rekordban* rögzíteni, melynek ilyen mezői lehetnek: a fogalom megnevezése, a fogalom osztálya, jelzete, definíciója, főlérendelt és egész-fogalma, szinonímája stb. A fogalomra vonatkozó eme információk összességét *fogalmi halmaznak* nevezik. A fogalmi halmazok és rekordok

* A hazai terminológiában „pontatlan ekvivalencia”-ként is ismert. – *A ref. megj.*

létrehozhatók utólag is egy olyan számítógépes rendszer birtokában, amely alkalmas az adatok tetszés szerinti válogatására, összerendezésére az elemzés és az összehasonlítás céljaira.

Rendező rendszerek fogalmi összehasonlítása

A fogalmi összehasonlítást nyelvi összehasonlításnak, eltérő nyelvű rendszerek esetén a fogalom-rekord fordításának kell megelőznie.

A fogalmi kompatibilitás vizsgálatában és megteremtésében számos eljárás követhető. Azonos nevű fogalmak esetén például a definíciók segítségével ellenőrizhető, hogy a fogalmak egybeesnek-e. Akár ekvivalens fogalmakról, akár homonimákról vagy poliszmákról van szó, a fogalom-rekordban jelölni kell a megnevezésbeli azonosságot.

A szinonimák kiszűrése a szinoníma, illetve a közvetlen fölérendelt mezők alapján történik, majd a fogalomnév és a definíció mezők vizsgálata után, a szinoníma mezőkbe kerülnek, vagy az azonos nevű fogalmaknak megfelelő eljárásokon esnek át.

A hierarchikus fogalmi összevethetőséget a fogalom-rekord szakcsoport, valamint a közvetlen fölérendelt és a legmagasabb fölérendelt mezők alapján határozzák meg. Ha e három mező összevetése alapján a rendszerek elemei megegyeznek, a fogalomnév és a definíció ellenőrzése után teljes vagy pontatlan ekvivalencia állapítható meg. E műveletek mutatják a rendező rendszerekben fellelhető egymásnak megfelelő hierarchiákat, az eltérő hierarchikus láncokat, illetve azt, hogy milyen hierarchikus szintek hiányzanak az egyik rendszerből. A fogalmi összevetés további támaszai lehetnek a referencia formai kategóriáira utaló mezők, például tárgy-entitás-absztraktum, folyamat-tevékenység-állapot. Eme összevetések alapján kialakítható a kompatibilitás mátrix.

Fogalmilag kompatibilis rendező rendszerek alkalmazása

Az UNISIST egyik jelentése, mintegy 12 évvel ezelőtt, javaslatot tett a különböző nyelvű információs rendszerek közötti átkapcsolást szolgáló közvetítő nyelvek létrehozására. A kompatibilitás mátrixok segítségével a fogalom-rekordok kiegészíthetők a fogalmat jelölő terminus idegen nyelvű megfelelőivel, s ez a mező – hosszú távon – lehetővé teszi a különböző rendező rendszerek közötti átkapcsolást. Ahogy a számítógéppel kezelt tezauruszok és osztályozási rendszerek száma nő, egyre könnyebb lesz ezeket a fogalom-rekordok megfelelő elemeivel kiegészíteni, és a rendező rendszereket – ezek segítségével – kompatibilissá tenni.

Szakmai kihívás

Az ilyen elemző és összehasonlító tevékenységek – a rendező rendszerek összehangolásán túl – elősegítik a fogalmak pontosabb megértését, hatékonyabb használatát. A fogalmak jelölésére szolgáló különböző nyelvű kifejezések – definíciókon alapuló – megfeleltetése pedig a terminológiai és fordítási munkálatok terén is előrelépést jelent.

Irodalom

1. HAMMOND, W.: Dimensions in compatibility. In.: Information systems compatibility. Szerk. Simon M. Newman. Washington, D.C. Soartan Books, 1965. 7 p.
2. COATES, F. J.: Switching languages for indexing = Journal of Documentation, 26. köt. 1970. 103 p.
3. HENDERSON, M. M. et al.: Cooperation convertibility and compatibility among information systems: A literature review. National Bureau of Standards Miscellaneous Publication 276. Washington, D.C., Government Printing Office. 1966. 78 p.

/SVENONIUS, E.: Compatibility of retrieval languages. Introduction to a forum = International Classification, 10. köt. 1. sz. 1983. p. 2-4.;

DAHLBERG, I.: Conceptual compatibility of ordering systems = International Classification, 10. köt. 1. sz. 1983. p. 5-8./

(Orbán Éva)

A kiadványban szereplő katalógus-adatok használata az Egyesült Királyság könyvtáraiban

A CIP és a kiadók

Könyvtáros körökben több, mint 100 éve vitatéma egy megfelelő program létrehozása a kiadványban történő katalogizálásra, de a könyvkereskedelem ezzel szemben kezdettől fogva fenntartással élt. A British Library 1977 februárjában kezdte CIP (Cataloguing-in-Publication = katalogizálás a kiadványban) programját, a könyvkereskedelmi előnyöket hangsúlyozva. Az előnyök – hirdette – abból származnak, hogy a kiadók sokkal jobban tudják ütemezni a nyomdai igényeket, ha a könyvtárak – a CIP gyorsaságának köszönhetően – előzetes rendeléseket (advance ordering) adnak fel. Ha pedig a könyv gyorsabban érkezik a könyvtárakba, akkor a pénz hamarabb áramlik vissza a kiadókhoz és a kereskedőkhöz. A feldolgozáson és a katalogizáláson