

hetjük például a jelzáloghitelek és a takarékbetétek (típus) havi kamatlábát az elmúlt évre (időbázis) az Egyesült Államokban (a részletezés szintje). Ehhez előre kell tudnunk (a dokumentációból), hogy ezek a változók szerepelnek-e az adatbázisban. Vagyis előre tudunk mindent az adatról, kivéve magát az értéket. Ez az érték az, amit le hívunk vagy visszakeresünk az adatbázisból.

Egyes keresőrendszerek (DIALOG) lehetővé teszik a változókra való szöveges keresést is. Ilyenkor azonban sok nehézséget és pontatlanságot okoz a lehetséges többféle írásmód, a különböző rövidítések, a többféle lehetséges megnevezés stb. Célszerű ezért inkább az adatbázis *kódrendszerével* végezni az adatlekvást. Ilyen kódrendszere a legtöbb numerikus adatbázisnak van a különböző változókra. Használatuk növeli a pontosságot és időt takarít meg.

A végső lépés a visszakeresett adatok alakítása és megjelenítése. Egyes rendszereknek (I. P. SHARP) az adatok online manipulálására is vannak eszközeik. Más rendszerekben csak arra van lehetőségünk, hogy a saját mikroszámítógépünkbe letöltve az adatokat, azokat ott utólag, offline módon alakítsuk az igényeinknek megfelelő formára.

A numerikus adatbázisok használata során alapvetően fontos a dokumentáció igénybevétele, mert az adatok nem magyarázzák önmagukat. Ne várjunk azonban segítséget a dokumentációtól a visszakeresett adatok *interpretálásában*. Azt például nem találjuk meg ott, hogy mi az adatokban látható infláció oka.

A szöveges szolgáltatórendszerekben megszoktuk, hogy a költségek a kapcsolati időtől és a kinyomtatott eredmény terjedelmétől függenek. A numerikus szolgáltatórendszerekben azonban gyakran ettől eltérő árképzéssel találkozunk: a költség a számítógép központi egységének a használati idejétől függ. Ilyen rendszerben tehát lehet, hogy egy

terjedelmes eredményt adó egyszerű adatlekvás viszonylag olcsó, míg egy kevés eredményt adó, de bonyolult feldolgozással járó munka meglepően drága. Amíg kevés a tapasztalatunk, nagyon nehezen becsüljük meg a várható költséget.

Ha más számára végezzük az információkeresést, fontos az alapos előzetes interjú. Ha az ügyfél számunkra idegen terminológiát használ, viágosan tisztáznunk kell, hogy mire gondol.

Miért van szükségünk numerikus adatbázisokra?

A numerikus adatbázisok lényegében nem tartalmaznak több adatot, mint a nyomtatott kézikönyvek, használatuk megtanulása pedig kétségtelenül többlet-erőfeszítést igényel. Miért van mégis szükségünk rájuk?

A számítógépek az adatbázisok gyakori és gyors aktualizálását teszik lehetővé. Az adatbázisok ezért mindig sokkal frissebb adatokat tartalmazhatnak, mint nyomtatott megfelelőik. Kiküszöbölhető velük az adatok lemásolása során fenyegető elírások veszélye. Nagy időmegtakarítást nyújt a számítógép az adatok manipulálásában, átalakításában. Végül, a számítógépek óriási tárolókapacitása hatalmas adatmennyiségek gyors elérésére ad módot.

Az adatok mennyisége nemcsak óriási, de nagyon gyorsan növekszik is. Már nemcsak az adatok hagyományos keresése válik egyre lehetetlenebbé, hanem az is, hogy az egyre több szükséges adatot egyáltalán kinyomtassuk. A kiutat a számítógépes adatbázisok jelentik.

/SUOZZI, P.: *By the numbers: an introduction to numeric databases.* = *Database*, 11. köt. 1. sz. 1987. p. 15– 22./

(Válas György)

Információs hálózatok nyelvi eszközei és kompatibilitási problémái

Az információs hálózatok típusai és a nyelvi kompatibilitás

A hálózatok *deduktív* létrehozásakor már az adatbázisok tervezésénél figyelembe veszik a nyelvi kompatibilitás követelményét: a hálózatban összekötött adatbázisok egységes, vagy legalább összehangolt elvek szerint felépülő nyelvi eszközöket használnak. Gyakoribb azonban, hogy az információs hálózatok *induktív* módon jönnek létre, vagyis már működő információs rendszereket kapcsolnak össze. E rendszerek feldolgozási módszere és technológiája eltérő. Az ilyen típusú hálózatokban a nyelvi kompatibilitást radikális átalakítás nélkül kell elérni olyan mechanizmus segítségével, amely elvégzi a fordítást az adatbázis kibocsátó nyelvéről a hálózat szolgáltatási

nyelvére. Hasonló problémák merülnek fel a hálózat szempontjából külső (pl. külföldi) adatbázisok átvételkor.

A külföldi adatbázisok információkereső nyelveinek struktúrája igen különböző. Mivel a nyelvi eszközök egységesítése ez esetben lehetetlen, a hálózatoknak olyan univerzális mechanizmusra van szüksége, amely képes átalakítani bármely információkereső nyelvet a befogadó rendszer (hálózat) nyelvére.

A külföldi hálózatok és kereskedelmi információs szolgáltatások (pl. Dialog, Data-Star, SCD stb.) adottan veszik az adatbázisok információkereső nyelv-

nek szemantikáját és struktúráját, és beérik a hálózatba tartozó adatbázisok szintaktikai struktúráinak formális átalakításával. Ezzel a felhasználók számára olyan univerzális lekérdező (parancs-) nyelvet alakítanak ki, amely alkalmas a hálózat valamennyi adatbázisában való keresésre. Maguk az adatbázisok változatlan formában működnek, így maradnak a kiküszöbölhetetlen párhuzamosságok, eltérő keresőnyelvi eszközök, sőt különböző természetes nyelvek. Ezek a hálózatok a jövőben – a teljes szövegek tárolásának (full-text database) megvalósulásával – teljesen ki akarják küszöbölni az adatbázisba bevitt dokumentumok szemantikai feldolgozását és azt szabad szövegszavas kereséssel akarják felváltani. A hálózatok fejlődésének ezt az irányát joggal nevezhetjük *extenzívnek*. Az ilyen típusú hálózatok könnyen képesek adaptálni új adatbázisokat, de képtelenek olyan szolgáltatásokra, amelyek minősége megköveteli az információfeldolgozás szemantikai módszereit.

A hálózatok fejlődésének másik iránya az *intenzív* fejlesztés: a rendszer működési intenzitásának növelése, amelyre nem az adatbázisok számának növelése, hanem a feladatok és a felhasználók körének kiszélesítése és a minőségi követelmények ebből következő növekedése a jellemző. Az extenzív és az intenzív fejlesztés közötti választás kérdésében a faktografikus rendszerek fejlődése az intenzív fejlesztés irányát erősíti. Természetesen a létező információs rendszerekben és hálózatokban a fejlődés extenzív és intenzív útja kombinálódik. A konkrét rendszerek esetében egyik vagy másik legfeljebb uralkodó jellegű. Az intenzíven fejlődő automatizált információs rendszereknél a vezető tendencia a rendszer "intellektusának" növelése. A fejlesztés iránya az adatbázisok tudásbázisúvá való átalakítása, az automatizált információs rendszerek mesterséges intelligencia rendszerekké való fejlesztése.

Az intenzíven fejlődő hálózatok nyelvi alapjai és kompatibilitási problémái

A hálózatok intenzív fejlesztése során elsősorban az ágazati és ágazatközi rendszerekben felismerték a szemantikai módszerek prioritását mind az információk feldolgozásában, mind az információkeresésben. Az ilyen hálózatok információkereső nyelvei explicite magukba foglalják a tárgyi világ fogalmi struktúráját, rendszerint osztályozási rendszer vagy teaurusz segítségével. Éppen ezért az ilyen információkereső nyelveket néha "fogalmi" nyelveknek nevezik, megkülönböztetve őket a "verbális" információkereső nyelvektől, amelyek a természetes nyelv lexikáját a szemantikai struktúra és a szemantikai viszonyok értelmezése nélkül veszik át. A "verbális" információkereső nyelvek fordítása az adatbázisokban megjelenő terminusok szó szerinti fordításának felel meg, amely minőségileg az idegen nyelvről való szó szerinti gépi fordításra emlékeztet. Ilyen fordítás mellett a torzulások hányada általában a fordított szöveg több mint 50%-át teszi ki. A verbális és fogalmi nyelvek

szempontjai közötti különbség gyakorlati megnyilvánulása az egyedi szavak vagy szóösszetételek, szókapcsolatok alkalmazásának ellentéte az információkereső nyelvek fejlesztésében.

A "fogalmi" információkereső nyelvek nehezen tűrik a közvetlen gépi fordítást, mivel szemantikai struktúrájuk bonyolultsága az összehasonlításkor valószínűtlenné teszi izomorfiájukat. Ezt támasztják alá azok a sikertelen kísérletek is, amelyekben az egyik osztályozási nyelvről (Nemzetközi Szabadalmi Osztályozásról) a másikra (az NTMIR és a szovjet automatizált információrendszer rubrikátorára) történő számítógéppel végzett átlépést, fordítást megfélemlések táblázatával végezték. A szemantikai struktúrák kongruenciájának mértéke a teauruszokban is igen alacsony.

Általánosan ismert, hogy bármely információkereső nyelv a természetes nyelvhez viszonyítva metanyelv, amelyben a természetes nyelv információkereső nyelvre való leképezésének szabályai az adott információkereső nyelv szemantikájának és pragmatikájának formájában jelennek meg. Ha az információkereső nyelv – természetes nyelvből átvett – lexikai egységeinek szemantikája többé-kevésbé invariáns a konkrét információs rendszerre nézve, akkor az alkalmazás gyakorlata (pragmatikus faktor) jelentős hatással van az információkereső nyelvre. Az indokolt azonban, hogy az információkereső nyelv olyan elvi tulajdonságait, mint a szótár terjedelme, a lexikai egységek prekoordinációjának mértéke, a grammatika típusa, az indexelés módszerei stb., az információs rendszer feladatainak jellege határozza meg.

A fentiekből következik, hogy a természetes nyelvű szövegek különböző információkereső nyelvek szerinti interpretációja jelentősen eltér, vagyis az eredeti szöveg még azonos típusú információkereső nyelveken is más-más interpretációban jelenik meg. Éppen ezért a szó szerinti fordítás helyett a *jelentés–szöveg* transzformációra van szükség. Az ilyen transzformáció legfontosabb követelményeit a gépi fordításban megfelelően számításba vették, ahol erre a feladatra egy olyan *közvetítő* nyelvet hoztak létre, amely biztosítja az összes fordítandó nyelv szemantikai invariáns kifejezéseinek egyértelmű reprezentálását, ábrázolását.

A jelentés reprezentációjának modelljei

A jelentést leíró nyelv – *lingua mentalis* – kidolgozásával főként azok a kutatások kísérleteznek, amelyek az ember és a társadalom intellektuális tevékenységének modellálásával foglalkoznak (nyelvelmélet, a mesterséges intelligencia kutatása, gépi fordítás stb.).

A jelentést leíró ismert modellek közül a legteljesebbnek és a legkorrektebbnek a "Jelentés–Szöveg" modell tűnik. Ennek legfőbb jellemzője (és egyesek szerint gyengéje), hogy tisztán nyelvi jellegű, kidolgozó tudatosan eltekintenek a "jelentés–való világ" viszonytól.

Ezzel szemben a mesterséges intelligencia rendszerek kidolgozói arra törekednek, hogy az extralingvisztikai ismereteket, információkat is ábrázolják. R. Shank – a mesterséges intelligencia elmélet egyik legnépszerűbb modelljének, a fogalmi függőség modelljének kidolgozója – kimondja, hogy a közvetítő nyelvnek azokat az információkat is le kell tudnia írni, amelyek a nyelvben világos, explicit formában nincsenek jelen. Ugyanezen az állásponton vannak a KRL (Knowledge Representation Language) néven ismert modell szerzői is.

E. V. Popov azt javasolja, hogy a gnoszeológiai (megismerési) folyamat különböző komponenseinek megfelelően a szemantikai nyelvek teljes készletét dolgozzák ki, megkülönböztetve a valóságról való tudás, a nyelvről való tudás és a viszonyokról való tudás rendszereit. A szemantikai nyelvek közül az ún. RX-kódok nyelve a bibliográfiai információkeresésre kidolgozott, bonyolultsága és költségessége folytán kevésbé elterjedt Perry–Kent-féle "clevelandi szemantikai kód" utóda. Ezzel szemben más tudásalapú rendszerekkel – nem az általában "keretekkel" [frame] ábrázolt típuszituációk elemzésén alapul, hanem a terminusokkal és fogalommeghatározásokkal ábrázolt fogalmak rendszerére épül.

Összegezve az elmondottakat: a "jelentések", amelyeket ezek a szemantikai nyelvek ábrázolnak, különböző ontológiai létezők, amelyek a "szöveg–jelentés–valóság" gnoszeológiai láncon helyezkednek el. A különböző nyelvek a láncon különböző fragmentumait írják le.

A "kölcsonös megértés" nyelv és a kompatibilitás kérdése

Ez az információkereső nyelv egy fogalmi háló, amely "atomáris" fogalmak, valamint irányított és értelmezett kapcsolatok halmazából áll. Atomáris fogalomról akkor beszélünk, ha két – atomáris – fogalom között egyetlen reláció áll fenn, vagy ha közöttük semmiféle kapcsolat nincs.

A fogalmi háló összefüggő részletét fogalmi blokknak nevezzük. Minden atomáris fogalom csak egy blokkba tartozhat. A blokkok leírásokká kapcsolódhatnak össze, és a leírás határain belül az atomáris fogalmak ismétlődhetnek. A fogalmi hálóban az irányított kapcsolatok inverze is megjelenik. Ugyanakkor semleges (irányítatlan) kapcsolatok is használhatók. A "kölcsonös megértés" nyelv szemantikai ereje főként az atomáris fogalmak "elemi" jellegétől függ, ami viszont pragmatikus alapú megállapodás kérdése.

Bár számtalan hasonlatosság mutatható ki a "Jelentés–Szöveg" modellek és a fentiek szerint modellált "kölcsonös megértés nyelve" között, lényegi különbség közöttük, hogy az utóbbi nem a természetes nyelv szemantikai struktúráját, hanem az információs hálózatban alkalmazott különböző információkereső nyelvek szemantikai struktúráját modellálja.

A szemantikai nyelvek felépítésében az egyik központi probléma a jelentéssel bíró "atomok" minimalizálásának kérdése. A "kölcsonös megértés nyelve" ezt a problémát tisztán pragmatikus kritériumok alapján oldja meg: az atomáris fogalmakat olyan mértékig kell szétbontani, hogy a hálózatba tartozó adatbázisokban lévő valamennyi információt meg tudják különböztetni. Hasonló a helyzet a fogalmak kapcsolatainak differenciálásával.

Az atomáris fogalmak blokkokba való összevonása biztosítja a különböző információkereső nyelvekben lévő lexikai egységek (deszkriptorok, kulcsszavak, osztályozási rubrikák) jelentésének megfelelő ábrázolását és a különböző információkereső nyelveken megadott keresőképek jelentésének átadását lehetővé tevő leírások felépítését.

A "kölcsonös megértés" nyelv pragmatikus irányultságát mutatja, hogy a rendszer felhasználóinak tematikai profiljába tartozó fogalmakat erősen részletezi, míg a határterületi fogalmaknál nagyfokú általánosítás figyelhető meg.

Következtetések

- ▶ Az információs hálózatok nyelvi eszközeinek kiválasztását két tényező határozza meg: a hálózat felépítésének induktív vagy deduktív jellege, a hálózat fejlesztésének extenzív vagy intenzív iránya.
- ▶ A deduktív elven felépülő hálózatokra jellemző, hogy a nyelvi kompatibilitás biztosítására egyetlen fogalmi szerkezetet dolgoznak ki, a nyelvi eszközök funkcionálisan hasonló komponenseit szabványosítják, egységesítik. Az induktív módszerrel felépülő hálózatoknál ezek az eljárások elégtelenek, szükség van fordítási mechanizmusok kidolgozására.
- ▶ A fordítási mechanizmusok az extenzíven fejlődő hálózatokban tisztán verbális jellegűek, ami a fejlődés során növekvő terheket hárít a felhasználókra.
- ▶ Az intenzíven fejlődő hálózatokban alkalmazott fordító mechanizmusok szemantikai jellegűek, közlések jelentésének modellálását tűzik ki célul és a mesterséges intelligencia rendszerek felé fejlődnek.
- ▶ A működő hálózatokra a deduktív és induktív módszerek kombinálása a jellemző, ezért a kompatibilitást különböző módszerekkel kell lehetővé tenni.

/ARTAMONOV, G. T.–ANTOPOL'SKIJ, A. B.: Problemy razrabotki lingvističeskogo obespečenija informacionnyh setej. = Naučno-tehničeska informacija, Ser. 2. 10. sz. 1986. p. 1–7./

(Környei Márta)