

3. Nihon Denshi Kogyo Shinko Kyokai/Japan Electronic (Kiad.) Industry Development Association (JEIDA) Kikai honyaku shisutemu no chosa kenkyu (Vizsgálódás és tanulmányok az automatikus fordítási rendszerek terén). Tokió: 1982, 281 p. (1-228. p. japán szöveg, 229-281. p. részben rövidített angol változat.)

/GALINSKI, Ch.: Die Informationsfluss zwischen Japan und Europa = Nachrichten für Dokumentation, 35. köt. 3. sz. 1984. p. 152-158./

(Papp István)

Az információkereső nyelvek jelene és jövője Csehszlovákiában

A Csehszlovákiában használt információkereső nyelvek (továbbiakban: IKNY) a következő célokat szolgálják:

- ◆ az információs rendszerekben feldolgozott dokumentumok bibliográfiai és tartalmi leírásának szabályozása,
- ◆ a dokumentumleírások gépi tárolási formájának szabályozása,
- ◆ az információkereső kérdés lefordítása természetes nyelvről IKNY-re,
- ◆ a releváns és pertinens dokumentumok kiválasztása a gépi memóriából,
- ◆ az információs rendszerből nyert válaszok szétosztása a felhasználók között.

Az IKNY-re közvetlenül vagy közvetve van szükség a fent felsorolt funkciókhoz, a feltárás, tárolás és visszakeresés munkafolyamataihoz.

Szakrendi felépítésű információkereső nyelvek helyzete és fejlesztési elképzelések

A szakrendi (hierarchikus) információkereső nyelvek a segítségükkel osztályozott dokumentumokat besorolják az emberi tudás egészének egy meghatározott – információkereső nyelv rendszerében különálló egységként nyilvántartott – helyére. A dokumentum tartalmának verbális megfogalmazását az őt magába foglaló osztály (vagy osztályok) tartalmának verbális megfogalmazása veszi át, de az osztályokat az esetek többségében nem leírásuk, hanem jelzetük képviseli.

A legtöbbet használt hierarchikus információkereső nyelvek Csehszlovákiában: az ETO, a szovjet könyvtári-bibliográfiai osztályozórendszer, és a NMTIR Rubrikátora.

Az ETO

Az információkereső nyelvek egységesítésére vonatkozó jelenlegi elképzelések szerint az ETO-nak az automatizált információkeresésben is jut szerep.

A Szlovák Nemzeti Könyvtárban az ETO-t az információkereső nyelvek egyikeként használják. Az ETO a dokumentum osztályba sorolására, míg a másik alkalmazott tárgyszavas IKNY a tartalom mélyebb feltárására, idő- és helybeli lokalizálására szolgál a Szlovák Nemzeti Bibliográfia számítógéppel kezelt rendszerében.

Az ETO-nak számítógéppel kezelt információkereső nyelvként való másik alkalmazását az ARDIS-ESZ elnevezésű, ESZ-1030 számítógépen futó rendszerben láthatjuk. A rendszert a nyitrai Agrártudományi Egyetem könyvtára használja. Érdekes tapasztalatokra tettek szert különböző információkereső nyelvek összekapcsolása terén, az ETO-t metanyelvként használják a kereső kérdések első megfogalmazásához, az ETO-ról a gépi tároláshoz használt tárgyszavas nyelvre már géppel fordítják a kérdést. A gépi fordítás a nyelvek belső kapcsolatait is figyelembe tudja venni, e munka tapasztalatai a bibliográfiai adatbázisok építésénél jól használhatók lesznek.

A Szlovák Műszaki Könyvtár a válogatott külföldi irodalom szolgáltatását egy 500 ETO-jelzet és egy deskriptoros kereső nyelvet kombináltan alkalmazó rendszerben oldja meg. Az ETO automatizált keresőrendszerben való alkalmazására kiemelt példa lehet még az EKO-INDEX rendszer, a Szlovák Központi Közgazdasági Könyvtárban. Az ETO-t kiegészítik az EKO-INDEX saját, szintén hierarchikus keresőnyelvével, így teszik alkalmassá gépi kezelésre.

A szovjet könyvtári-bibliográfiai osztályozórendszer

A hierarchikus osztályozórendszerek alapelvein készült szovjet osztályozórendszert a pozsonyi Egyetemi Könyvtár alkalmazza. Egyetemi könyvtárak számára azért jó ez a – számítógépes kezelésre is alkalmas – rendszer, mert különböző speciális gyűjtemények és katalógustípusok általában ilyen fajta tudományos könyvtárakban fordulnak elő, és a részletekbe menő keresés igénye is itt szokott felmerülni.

Az NTMIR Rubrikátorának rendszere

A hierarchikus információkereső nyelvek új típusa a Szovjetunióban az információkeresés gyakorlati tapasztalatai alapján kifejlesztett, egyetemes, a tudomány, technika és a népgazdaság egészét, valamint az interdiszciplináris területeket átfogó rubrikátor. A rubrikák jelzetei kifejezik tartalmukat és a hierarchiában elfoglalt helyüket. A rubrikátor három szintű, az egyes szintek hierarchikusan kapcsolódtak.

Az NTMIR Rubrikátorának főbb funkciói:

- ◆ az NTMIR alrendszerek tematikus határainak megállapítása,
- ◆ a más alrendszereknek továbbítandó információk csoportosítása,
- ◆ az információk szolgáltatás céljára való csoportosítása az alrendszereken belül
- ◆ a rubrikák szerint csoportosított információk kereshetőségének biztosítása,
- ◆ az ágazati rubrikátorok egyeztetésének szabályozása.

Ezt a legutóbb említett egyeztető funkciót azért szükséges külön kiemelni, mert az általánostól a speciális felé haladva könnyű az egyeztetés. Egyes ágazati alrendszerekben azonban a saját tematikájukat lefedő rubrikátor korábban készült el, mint az NTMIR Rubrikátora. Az egyeztetés "lentől-fölfelé" irányba csak kompromisszumok árán, és többéves munkával sikerült, például a VODOINFORM alrendszer esetében is. A naprakész vízügyi ágazati rubrikátor végül is 1984-ben készült el. Három részből áll: bevezetés, szakrendi rész és alfabetikus mutató. Teljesen kompatibilis a központi rubrikátorral, annak "70. Vízgazdálkodás, meliorizáció" rubrikájának tartalmát bontja tovább harmadik szintig (a harmadik hierarchikus szint hatszámjegyű jelzetet jelent). Ugyanolyan utalórendszert használ, nevezetesen a "lásd" és "lásd meg" utalókat, ezek a vízgazdálkodásnak más diszciplinákkal való kapcsolatait fejezik ki. Az alfabetikus mutató szlovák és orosz nyelven is elkészült.

Az NTMIR Rubrikátorának további alkalmazása a gépipar terén jelentős: az INFORMAŠ ágazati alrendszer rubrikátora három részrendszert szolgál, a gépipari technológiai, a mezőgazdasági gépipari és a kőolajipari információs rendszereket. Az INFORMAŠ rubrikátora az NTMIR-éből indul ki, és a negyedik hierarchikus szintig lebontva tagolja a gépipar tematikáját.

Tárgyszavas információkereső nyelvek jelenlegi helyzete és a velük kapcsolatos fejlesztési tervek Csehszlovákiában

A természetes nyelven alapuló IKNY-ek két típusba sorolhatók

- ◆ prekoordinált IKNY-ek,
- ◆ postkoordinált IKNY-ek

A prekoordinált nyelv a dokumentum tartalmát (vagy a kereső kérdést) olyan szólánccal fejezi ki, amelynek tagjai meghatározott szabályok szerint kapcsolódnak egymáshoz és csak együtt értelmezi a keresés az előre összekapcsolt szavakat.

A posztkoordinált IKNY-ek is a természetes nyelv szavaival írják le a dokumentumot, de egymástól független szavak ezek, kötött szókincsből választva. A posztkoordinált IKNY-ek tipikus képviselői a deskriptor-nyelvek. Ha a természetes nyelvből származtatott szavakkal írják le egy dokumentumot, akkor a homonímia mindig gondot okoz a visszakeresésnél.

A dokumentumok címéből vett szavakon alapuló IKNY-ek

A címből vett kulcsszavakon alapuló visszakeresést a KWIC és KWOC típusú indexek segítségével végzik. A számítógépek tájékoztatási célra való alkalmazásának hőskorában, 1966-ban kezdtek ilyen munkákat végezni Csehszlovákiában. Az AGROINDEX nevű referáló folyóirat mutatórendszere jelenleg is KWIC típusú.

Tárgyszavas osztályozórendszer

A tárgyszavakat használó IKNY-ek egyik példája a mező- és erdőgazdaság információs rendszerében található, ezt nem nevezhetjük tipikusnak, mert a kulcsszavas és deskriptor-típusú IKNY-ek sajátosságait egyesíti, az előbbi a feltárás, az utóbbi a visszakeresés folyamán jelentkeznek.

Deskriptor típusú IKNY-ek

A deskriptor típusú IKNY-ek kidolgozása, bevezetése egy országon belül komoly koordinációt kíván, különben a teauruszokra vonatkozó szabványok elkészültével javult a helyzet.

Csehszlovákiában az NTMIR normatív előírásain kívül a 01-0193 számú nemzeti szabvány szabályozza a teaurusz-készítést.

Egynyelvű (angol) tezauszot használnak Csehszlovákiában az import mágnesszalagokról való kereséshez. Példák erre: a MALIMET, az EX-CERPTA MEDICA egészségügyi információs rendszer tezauszja, az NEVIROTAPES adatbázis mágnesszalagjainak fogadásához használt tezausz, amellyel 21 témacsoportból lehet információt keresni, és a CAC és CAS kémiai információs rendszerek tezauszja.

Nemzeti használatra készült egy kétnyelvű (cseh és szlovák) kereskedelmi tárgyú tezausz, a cseh eredetű áruk a cseh elnevezést deszkriptorként, a szlovákot szinonimaként veszik fel, és megfordítva.

Többnyelvű tezauszok építése a jelenlegi fejlődési irány, természetesen ebben a munkában elfogadott nemzetközi szabványok segítségével lehet csak eredményt elérni.

Az NTMIR a környezetvédelmi a VODOINFORM a vízügyi és építőipari alrendszerének tezauszja készül olyan módon, hogy többnyelvűvé legyen fejleszthető. A tárgyszavas IKNY-ek közös problémája a szabad szövegből kötött szótárra való fordítás szabályozása. Ez a szabályozás vezet a többnyelvű tezauszok belső kapcsolatai kidolgozásának megoldásához is. A szabályozástól lehet várni azt is, hogy az egyes KGST-tagországokban kialakult ágazati információkereső rendszerek IKNY-ei összekapcsolhatók legyenek.

Következtetések

A csehszlovák tudományos-műszaki információs rendszer munkái során a modern információs technológiák alkalmazásával a fentebb érintett információkereső nyelvi problémákat is meg kell oldani, erősebb koordinációval össze kell fogni a különböző szakterületek e téren végzett munkáit.

A kialakuló IKNY-eknek az NTMIR már elkészült információkereső nyelvi eszközein kell alapulniuk. A jövőben kísérletezni kell az új gépi technikákkal, automatikus szövegelemzéssel, a szemantikai elemzésen alapuló szövegkiválasztással, az automatikus osztályozással, hogy az információszolgáltatás minősége javuljon. A bibliográfiai keresés céljára készült IKNY-ek faktográfiai célú keresésre való kiterjeszhetősége is további kutatást igényel, és munkálkodni kell az NTMIR nyelveinek más rendszerek nyelveivel való összekapcsolhatóságán.

/STANČIKOVÁ, P.: Present state and future development of information retrieval languages in Czechoslovakia = International Classification, 11. köt. 3. sz. 1984. p. 133-138./

(Domokos Miklósné)

A tudományos-műszaki tájékoztatás információkereső nyelveinek helyzete és fejlődési irányai a Német Demokratikus Köztársaságban

A tudományos-műszaki tájékoztatás információkereső nyelveinek (IKNY) kidolgozottsága és alkalmazásuk az NDK-ban

Az NDK-ban jelenleg a következő IKNY-eket alkalmazzák:

- ◆ tezauszok,
- ◆ tárgyszavas rendszerek,
- ◆ rubrikátorok, szakmai osztályozórendszerek,
- ◆ faktográfiai információs rendszerek speciális keresőnyelvei,
- ◆ egyetemes osztályozások, mint pl. az NSZO, vagy az ETO.

A továbbiakban röviden jellemezzük ezeket az IKNY-eket, részletesebben kitérve a tezauszokra és rubrikátorokra.

A tezauszok fejlettségi színvonala az NDK-ban viszonylag magas. A tudományos-műszaki terü-

leten folytatott tezausz-munkákra hatással van a Nemzetközi Tudományos és Műszaki Információs Rendszer (NTMIR) keretében végzett ilyenirányú tevékenység. Csaknem valamennyi diszciplína rendelkezik megfelelő tezausszal, így a természettudományok és a műszaki tudományok is. 1983-ban az országban kereken 100 német nyelvű tezausz volt használatban. Ezek eloszlása az egyes tudományterületek között azonban egyenetlen, ami összefügg az ágazati információs rendszerek eltérő struktúrájával.

A tudományos-műszaki tezauszok többsége a 70-es években keletkezett, zömében 1974 és 1977 között, ami alapvetően az 1973 végén kiadott un. bázistezausz megjelenésével (Thesaurus Naturwissenschaft und Technik) magyarázható. Tekintettel arra, hogy ennek használatát kötelezővé tették, sok olyan szakterületen is dolgoztak ki tezauszot, ahol ez korábban nem volt. Ezek a munkák a 70-es