

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 fejlesztendő. 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.

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-

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 technikával, 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)

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 kerekén 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 ún. 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

évek végére zömében befejeződtek. A fenti tezauszok kivétel nélkül ágazati orientáltságú információszervezetekben kerültek alkalmazásra. Speciális dokumentumtípusra orientált információszervezetek, illetve szolgáltatások esetében több, a bázistezauszhoz kapcsolódó tezauszt alkalmaznak, ezzel megteremtve az egységesség minimális feltételeit.

A jelenlegi tezauszok kézi működtetésű és automatizált rendszerekben egyaránt használatosak.

A tezauszok minősége metodikai előírások szempontjából, illetve a bázistezauszokhoz és az egyéb kooperáló információszervezetekhez való illeszkedést figyelembe véve megfelel a követelményeknek. Az általuk elérhető hatékonyságra vonatkozóan (a keresés eredményének teljessége és pontossága, hatása a tárolási és keresési költségekre stb.) csak általános jellegű nyilatkozatokra támaszkodhatunk, mivel ezideig ilyen irányú értékelést nemigen végeztek.

A tezauszokat leggyakrabban használó szakemberek véleménye szerint ezek "lényegében megfelelnek a követelményeknek". Mindezek ellenére az utóbbi években több tezauszt továbbfejlesztettek. Ezek a munkák a mindenkori nemzetközi ágazati tezauszokhoz való alkalmazkodást figyelembe véve folynak. A szocialista gazdasági integráció kiszélesedésével párhuzamosan nőtt a többnyelvű tezauszok jelentősége is. Ezek kidolgozása a tezauszok továbbfejlesztési munkálataival párhuzamosan történik.

Speciális dokumentumtípusokkal foglalkozó információszervezetek területén nem folyt semmilyen említésre méltó munka.

A többnyelvű tezauszok kidolgozása orosz-német nyelven indult, és kezdeti szakaszában tart. Egy bizonyos kidolgozottsági szint elérése után a kétnyelvű változatot kiterjesztik a KGST információszervezetében részt vevő többi ország nyelvére is. Az újonnan létrehozandó nemzetközi és nemzeti tezauszok közötti szükségszerű egyeztetés legtöbbször esetében szükségessé teszi az utóbbiak alapos átdolgozását. Éppen ezért a gyakorlatban úgy járnak el, hogy a többnyelvű tezausz kidolgozásával párhuzamosan a megfelelő szakterület egynyelvű tezauszának újrafogalmazása is megtörténjen. A tárgyszavas rendszereket elsősorban a hagyományos könyvtári munkaterületeken használják. Ezt alkalmazzák a szakkatalógusok egyik legelterjedtebb típusának, a tárgyszó-katalógusoknak az előállításánál is. Az egyes információszervezetek által kiadott referátumok, bibliográfiák, jegyzékek stb. mutatói is többségükben tárgyszavas rendszerek. Az NDK-ban jelenleg alkalmazott tárgyszavas rendszerekről pontos áttekintésünk nincs. Valószínűleg széles körű az alkalmazásuk, így pl. a tezauszok

előállításának is fontos forrásai, illetve betűrendes deskriptorjegyzékként részévé válnak a tezauszoknak.

A tezauszok mellett a rubrikátor a leggyakrabban használt információkereső nyelv. Különösen az utóbbi években növekedett meg a jelentősége, amit az NTMIR keretében végzett ilyenirányú munkálatok is elősegítettek.

Az NTMIR információkereső nyelveinek létrehozásával foglalkozó alapidokumentum a rubrikátorok alkalmazását kötelezően előírja. E dokumentum szerint az NTMIR rubrikátorrendszere az alábbi részekből kell, hogy álljon:

- ◆ az NTMIR rubrikátora,
 - ◆ az un. specializált (dokumentumtípusra orientált) információszervezetek rubrikátorai,
 - ◆ az ágazati információszervezetek rubrikátorai.
- A teljes rendszer ugyan még nem készült el, de pl. az NTMIR-rubrikátora felhasználásra már készen áll.

A Szovjetunió Állami Tudományos-Műszaki Információszervezetének kiépítése szükségessé teszi egy információkereső nyelv-komplexum létrehozását. Ennek magvát rubrikátorok és az egész rendszert lefedő bázistezauszok (makrotezauszok) együttese képezi majd. A rubrikátorok 1982-es kiadása jelentős lépést jelentett a rendszer megvalósítása felé. A Szovjet Tudományos-Műszaki Információszervezet és az NTMIR által alkalmazott megnevezések és kódok első szinten teljes mértékű egyezést mutatnak. Az átmenet az egyikből a másikba a rubrikák tartalma, megnevezése és kódolása tekintetében a második szinten kb. 90%-os.

Az NTMIR rubrikátora nemzeti célra történő felhasználását elősegíti a német nyelvű kiadás, ami tartalmilag és formailag pontosan megfelel az eredetinek.

Az NTMIR rubrikátorának német megfogalmazása szolgált az ágazati rubrikátorok létrehozásának alapjául, melyek csaknem valamennyi ágazat számára elkészültek. Gyakorlati alkalmazásuk főleg bizonyos részfunkciókra terjed ki (pl. egy adott információszervezet tematikus körülhatárolása és elhatárolása más rendszerektől). A későbbiekben egyes ágazati tezauszok tematikus terjedelmének meghatározására, illetve a tezauszlexika elsődleges durva felosztására alkalmazzák majd. A rubrikátorok információkeresésre való alkalmazásáról tapasztalat hiányában nem beszélhetünk még. A nemzetközi ágazati rubrikátorok német nyelvű változatait az NDK ágazati információszervezeteinek mintegy 25%-ában használják. A faktorgrafikus információkereső rendszerekben alkalmazható keresőnyelvekkel kapcsolatos eddigi munkák alig érdemelnek említést.

Az információkereső nyelvek fejlesztési irányai az NDK-ban

A tudományos-műszaki tájékoztatásban alkalmazott információkereső nyelvek továbbfejlesztésénél figyelembe kell venni, hogy azok fontos interdiszciplinális területeket is ki kell, hogy szolgáljanak. E területek szabványait, műszaki-gazdasági jellemzőit az információkeresés számára elérhetővé kell tenni. Az 1990-ig továbbfejlesztendő információs rendszer adattárak, adatbankok információs szolgáltatásaiából, illetve speciális dokumentumokból nyújtott információs szolgáltatásokból áll majd. Az információs szolgáltatások tovább szakosodnak és hatékony, terminálos hozzáférést biztosító hálózattá szerveződnek. Mindez a nemzetközi együttműködéssel összhangban történik majd. A gépi adatfeldolgozás és lekérdezés megvalósítása érdekében az információkereső nyelvek kidolgozásánál az alábbi szempontokat kell szem előtt tartani:

- ◆ Az IKNY-et kezdettől úgy kell kialakítani, hogy magas színvonalú információellátást biztosítson a felhasználó számára.
- ◆ Minden munkát a Szovjetunió és az NTMIR vonatkozó szándékainak figyelembevételével és azzal összhangban kell végezni.
- ◆ Az IKNY-ekkel kapcsolatos minden munkát a kapcsolódó nemzetközi metodikák szem előtt tartásával, a megfelelő KGST-szabványokkal, illetve egyéb előírásokkal összhangban kell végezni.

Az IKNY-ekkel kapcsolatban megoldandó legfontosabb problémák és feladatok a következők:

- ◆ Az automatizált információs központok hálózata számára alkalmazható információkereső nyelvek kidolgozása, melyeknek alapvető feladata a bevitt információk tartalmi rendszerezése és az adatok strukturálása.
- ◆ IKNY-ek kidolgozása a párbeszédés üzemmódhoz.
- ◆ Automatizált rendszerek kidolgozása az egy- és többnyelvű indexeléshez, melynek minőségileg a jelenlegi szintet messze túl kell szárnyalnia.
- ◆ Nyelvi eszközök kidolgozása a kérdések rövid szövegek formájában való bevitelére.
- ◆ Automatizált információs rendszerek nyelvi eszközeinek összekapcsolása gépi fordítást végző rendszerek nyelvi eszközeivel.
- ◆ Nyelvi eszközök és algoritmusok kidolgozása a primér dokumentumokban elhelyezkedő faktografikus adatok jellemzésére és kiválasztására.
- ◆ Normatív dokumentumok kidolgozása az információkereső nyelvek egységesítésére, különös tekintettel a hálózatokban történő felhasználásukra.

Automatizált tudományos—műszaki információs rendszerek információkereső nyelvei

Az NDK-ban kifejlesztendő információs rendszerek szervezési szempontból 3 különböző szinten álló részrendszer összességéből fognak majd állni:

- ◆ országos szintű részrendszerek,
- ◆ ágazati szintű részrendszerek,
- ◆ üzemi szintű részrendszerek.

Ennek megfelelően az IKNY-eknek a következő feladatokat kell ellátniuk:

- ◆ részrendszerek tematikus határainak megállapítása,
- ◆ beérkező információk szétosztása az alrendszerek számára,
- ◆ dokumentumok és/vagy kérdések indexelése és rendszerezése,
- ◆ az állami automatizált osztott adatbankok és decentralizált adatbank-rendszerek felépítése. Itt egymáshoz illesztett információkereső nyelvek egész komplexumára lesz szükség, ami lényegében két összetevőből kell, hogy álljon:
 - ◆ általános rendszerszemlélettel épült IKNY-ből,
 - ◆ decentralizált IKNY-ekből

Az általános IKNY két részből, a központi és periférikus változatból fog állni. Az utóbbi főleg az ágazati rubrikátorokból és tezaurusokból épül fel.

A központi rész az alábbi elemeket tartalmazza:

- ◆ az állami rubrikátort (egészen a 3. szintig),
- ◆ a szükséges konkordanciataláblázatokat a rubrikátor rovatai és az egyetemes osztályozások (ETO, NSZO stb.) jelzetei között,
- ◆ a tudományos-műszaki tájékoztatás makrotezaurusát mint bázistezauruszt,
- ◆ minden segéd- és szakszójegyzék teljes anyagát,
- ◆ szabványosított nyelvtani eszközöket,
- ◆ a dokumentumok koordinált indexelésére és lekérdezésére vonatkozó típuseljárásokat,
- ◆ normatív dokumentumokat, melyek az IKNY kidolgozását, alkalmazását és fejlesztését kötelezően szabályozzák.

Egy önálló automatizált információs rendszer IKNY-inek komplex rendszere

A rendszer összetételét a vele szemben támasztott mindenkori követelmények határozzák meg. Egy nagy rendszer — mondjuk automatizált ágazati rendszer — esetében a munkát a részrendszerekkel való együttműködés figyelembevételével kell megszervezni.

Az információ egyszeri, komplex feldolgozása után az eredményeket többszörösen hasznosítják. Ezt figyelembe véve az IKNY komplex rendszerének kialakítása az alábbi folyamatokból áll:

- ◆ az információs bázisba kerülő dokumentumok tartalmi feldolgozása (indexelés, rubrikálás, osztályozás),
- ◆ faktografikus információk feldolgozása és kódolása,
- ◆ a kérdések kidolgozása a dokumentumok és faktografikus adatok keresésénél,
- ◆ a rendszer információs eszközeinek előállítására felhasználandó dokumentumok kiválasztása és rendszerezése.

A mélyebb tartalmi feltárára posztkoordinált információkereső nyelvet (tezaurusz) használnak.

Hálózati üzemmódban működő automatizált információs rendszerek IKNY-ei

Mivel a hálózat egyes tagjai egymással állandó kommunikációs kapcsolatban állnak – különösen számítógépes információfeldolgozás és keresés esetén – így szükséges, hogy ezek információkereső nyelvei egymáshoz illeszkedjenek.

A hálózat egy tagjának az alábbi nyelvi eszközökkel kell rendelkeznie:

- ◆ hálózati rubrikátor,
- ◆ hálózati bázistezaurusz,
- ◆ a hálózati IKNY grammatikai eszközeinek nomenklatúrája,
- ◆ a hálózati rubrikátor és bibliográfiai osztályozási rendszerek közötti konkordanciátáblázatok,
- ◆ a dokumentumok indexelési metodikája az információcsere és a visszakeresés figyelembe vételével.

IKNY a párbeszédés üzemmód megvalósítására

A párbeszédés üzemmód speciális lehetőségeinek minél jobb kiaknázására jól kidolgozott tezauruszra van szükség, melynek lexikai egységei deskriptorokból és faktografikus adatokból állnak. Ezeket egymással a hagyományos tezauruszokénál bonyolultabb logikai kapcsolatba hozva formulázhatjuk meg a keresőképet illetve stratégiát.

Az IKNY-ekkel kapcsolatos távlati feladatok 1990-ig az NDK-ban

A dokumentumorientált IKNY-ekkel szerzett eddigi tapasztalatok alapján arra lehet következtetni, hogy a jövőben egy adott információs rendszeren belül a tezauruszok és rubrikátorok kombinált használata lesz célszerű. A tezaurusz eddigi feladatának megfelelően továbbra is a koordinált indexelés legfőbb segédeszköze marad, míg a rubrikátor egy un. rendező funkciót is ellát majd.

A tezauruszokat alkalmazni kell a fent vázolt többszintű információs rendszer különböző szintjein (makrotezauruszok, vagy bázistezauruszok a felső szinten és ágazati, vagy szaktezauruszok alsóbb szinteken). Ezekhez járulnak még a politematikus dokumentumállományokhoz alkalmazható tezauruszok.

A nemzetközi információs rendszerek számára szükséges többnyelvű tezauruszokat tervszerűen tovább kell fejleszteni.

A rubrikátorok területén a következő feladatok vannak:

- ◆ az NTMIR rubrikátorát a Szovjet Tudományos-Műszaki információs Rendszer rubrikátora alapján tovább kell fejleszteni a 3. szintig,
- ◆ az ágazati rubrikátorok továbbfejlesztése és fokozottabb felhasználásuk az információs gyakorlatban.

Várható, hogy a tezauruszok és rubrikátorok egy- és többnyelvű felhasználhatóságát segítő munkák a 80-as évek második felében folytatódni fognak. Ugyanakkor a faktografikus IKNY-ekkel és az automatikus indexelést célzó hivatkozási rendszerekkel kapcsolatos munkák a 80-as évek végéig nem fognak látványos előrelépést felmutatni, mivel ezen a területen még jelentős kutatómunkára is szükség van.

/ ELLMANN, D.: Stand und Entwicklungsrichtungen der Informationsrecherchesprachen für die wissenschaftlich-technische Information in der DDR = Informatik, 30. k. 6. sz. 1983. p. 12–16./

(Péteri György)