

## INFORMÁCIÓKERESŐ RENDSZEREK

Többnyelvű információkereső rendszerek:  
integráltsági szintek és nyelvek

Az utóbbi években számos nemzetközi információkereső rendszer jött létre vagy volt születőben. Közülük a legjelentősebb az UNISIST-tervezet, a KGST-tagországok Nemzetközi Tudományos és Műszaki Információs Rendszere, valamint az EURONET keretében kidolgozott rendszer. A cikk néhány olyan működő vagy kísérleti szakaszban levő információkereső rendszert mutat be, amely a kooperáció valamely lehetséges formáját hordozza magában.

A nemzetközi információkereső rendszerek fő jellemzője az *integráltság szintje*. Ez a következőket jelenti: tételezzük fel, hogy egyazon tematikában a különböző országokban több információs központ működik. Mind-egyikük ellátja felhasználóit nemzeti nyelvű információval. A központok közös jellemzője az általuk *feldolgozott dokumentumok tematikája* (ezen belül előfordulhat, hogy maguk a dokumentumok is ugyanazok) és a *nyújtott szolgáltatások formái*. Általában az egyes központok feladatai a következők:

1. a nemzeti és idegen nyelvű primer dokumentumok gyűjtése és tárolása;
2. a primer dokumentumok leírása és ennek alapján szekunder dokumentumok készítése;
3. rendszeres tájékoztatás a megjelent dokumentumokról a legkülönbözőbb bibliográfiák és indexek formájában;
4. a szelektív információterjesztés megszervezése;
5. retrospektív keresés;
6. a dokumentumok másolása és fordítása;
7. adatbankok szervezése.

A feldolgozásra kerülő dokumentumok és a megoldandó információs feladatok azonossága az ilyen központok rendszerbe foglalásának alapfeltétele. A rendszer létrehozásához elengedhetetlenül szükséges még a közös nyelv és a kommunikációs hálózat megléte.

A működő nemzetközi információkereső rendszerek vizsgálata arról tanúskodik, hogy közöttük az *együttműködés különböző formáiban valósul meg*. Ennek megfelelően különböző integráltságuk szintje is, ami elsősorban az információellátás területén kezdeményezett közös vállalkozásaik számától függ. Nyilvánvaló, ha két információkereső rendszerben az integrálás mindössze az egységes dokumentumleírási formátum használatára korlátozódik, a másik kettőben viszont ezzel együtt a rendszeres tájékoztató szolgálat is kifejlődött, azaz a bibliográfiákat, katalógusokat, referáló lapokat, tematikai kiadványokat stb. közösen állítják elő, az integráltság minőségileg más-más szintjeiről van szó.

Az integráltsági szint másik mutatója, hogy a kooperatív információellátás érdekében sikerült-e és *mennyire sikerült kiküszöbölni a nyelvi akadályokat és a nemzeti sajátosságokat*. Ezt a mutatót viszonylag nem régen használják, csupán az új európai információkereső rendszerek kialakításával kapcsolatban bukkant fel, mégpedig e rendszerek ún. belső és ún. külső nyelveinek problematikájából kifolyóan. A *belső nyelvek* (rendszerint több van belőlük) formalizáltak, a dokumentumok és kérések egységes leírását szolgálják. A *külső nyelv* általában valamely nemzeti nyelv. Ezért a nyelvi akadályok leküzdése egyet jelent a belső nyelvek és a külső nyelv kompatibilitásának megteremtésével.

### Az integráltság első szintjén levő nemzetközi információkereső rendszerek

Az első nemzetközi információkereső rendszerek létrehozásának az volt a fő oka, hogy a több nyelven folyamatosan beérkező anyagról operatív és centralizált szelektív tájékoztatást lehessen adni. Az ilyen rendszerek legjellemzőbb példájaként a MEDLARS (*Medical Literature Analysis and Retrieval System*) és a CAS (*Chemical Abstracts Service*) rendszerét lehet megemlíteni. Ezeket a rendszereket nevezhetjük teaurusz előttiékné is, mivel belső nyelvként a tárgyi osztályozást használják. Jellemzőjük a visszacsatolás mint irányítási forma hiánya, mivel egy külső nyelvük van (az angol) és egy belső, a központi szervezet által kidolgozott rubrikátor.

Az ilyen nemzetközi információkereső rendszerben részt vevő nemzeti szervek feladata csupán a saját dokumentumaik indexelése és egy külső nyelven – a rendszer munkanyelvén – való referálása. Nyilvánvaló, hogy sem a nemzeti szervezetek sajátossága, sem a nyelv nincs hatással az ilyen nemzetközi rendszerek működésére. Az egyetlen belső és egyetlen külső nyelv használata következtében a nyelvi akadályok problémája fel se vetődik.

### Az integráltság második szintjén levő nemzetközi információkereső rendszerek

A 60-as évek közepén kezdtek megjelenni az első, többnyelvű deskriptorokkal operáló információkereső rendszerek, amelyek belső nyelvként a teauruszt használták. Ezekre az információkereső rendszerekre *egyetlen belső nyelv és néhány külső nyelv használata jellemző*. E rendszerek tipikus példájaként az INIS (*International Nuclear Information System*) és az euró-

pai geológiai és ásványtani információkereső rendszer említhető.

Az európai geológiai és ásványtani információkereső rendszer a dokumentumok decentralizált gyűjtésére és elemzésére, a mutatók centralizált összeállítására és kiadására épül. A rendszer a szekunder dokumentumokat valamennyi tagja számára a közös tematikában és egységes formában állítja elő. Valamennyi résztvevő az egységes adatlapot használja, a központi szervezet – jelen esetben a Francia Nemzeti Tudományos Kutató Központ (*Centre national de la recherche scientifique, CNRS*) – végzi az információ bevitelét, ellenőrzését és terjesztését. Az adatlapok tárgyi rubrikákat jelölő speciális kódjai teszik lehetővé a szelektív információterjesztést. A retrospektív keresésekhez a BRGM (*Bureau de recherches géologique et minière*) az európai rendszer francia nemzeti központjával együtt dolgozta ki a báziszezsaurust. A más nemzeti nyelven történő információbevitel lehetővé tétele érdekében egyes információkereső rendszerek – a rendszer tagjai – elkészítették a *tezsaurusz nemzeti változatait*. Így készült el pl. a francia–spanyol és a francia–német kétnyelvű tezsaurusz. A kétnyelvű szótárak rendszere lehetővé teszi a dokumentumok és kérések keresőképek fordítását: mindkettőnek franciául kell szólnia. Mivel azonban a tezsauruszok különböző változatai eltérőek, az ezek alapján összeállított keresőképek is nagyon eltérhetnek egymástól.

Megjegyzendő, hogy az INIS-ben hasonló a helyzet. Az INIS tezsaurusz 1976. júliusi kiadásában az angol változat 14 649 deskriptort és 4272 tiltott kifejezést tartalmaz, míg a francia változatban 13 115 és 5338 a két mennyiség.

E hibák kiküszöbölésére az európai geológiai és ásványtani rendszer 1970-ben új hatnyelvű tezsaurusz kidolgozását határozta el. Alapjául az Amerikai Geológiai Intézet kb. 33 ezer kifejezést tartalmazó szótárát választotta. Minden tagországnak joga van saját kifejezésjegyzékének összeállítására, és részt vesz az egységes paradigmátika kidolgozásában is.

Ennek során először a többnyelvű tezsaurusz szókészletét vizsgálják át, mivel az egyes országok elsődlegesen az őket érdeklő fogalmakat illetően tartoznak állást foglalni. Második lépésként a belső nyelv egységes szemantikai struktúráját dolgozzák ki úgy, hogy az a rendszer minden résztvevőjének egyformán megfeleljen.

A belső és külső nyelv kapcsolatát tekintve optimális esetben a rendszernek biztosítani kell:

1. a dokumentumok elemzését és indexelését az indexelő szakember anyanyelvén;
2. a kérés kifejezését a felhasználó nyelvén;
3. a válasz kifejezését a felhasználó nyelvén.

Ebből a szempontból gyakorlati előnyük van azoknak a rendszereknek, amelyek munkanyelvüként több nemzeti nyelvet használnak. Az ilyen információkereső rendsze-

rek nemcsak demokratikusabbak és a használatban kényelmesebbek, de integráltsági szintjük is magasabb.

#### Az integráltság harmadik szintjén levő nemzetközi információkereső rendszerek

E szinten levő rendszerek legtipikusabb példája az OECD (*Organization for Economic Co-operation and Development*) mellett működő Nemzetközi Közúti Dokumentációs Központ által kidolgozott IRRD (*International Road Research Documentation*) rendszer. A központ 1965-ben kezdte meg működését, kezdetben az autópályaépítésről szóló információk szolgáltatásával. Ezt 1967-ben kiegészítették az autópályák használatáról szóló információkkal. A központban 3 egyenrangú (angol, francia, német) változatot tartalmazó többnyelvű tezsauruszt használnak, és folyamatban van a spanyol változat kidolgozása is. (Az integráltság második szintjén levő nemzetközi információkereső rendszerekben kétnyelvű szótárak vannak, amelyek egy bázisszótárra „mennek vissza”, illetve annak igénybevitelét segítik elő.)

A rendszernek 3 koordinációs központja van: Nagy-Britanniában, Franciaországban és az NSZK-ban. Ezek a központok további 17 nemzeti információkereső rendszer tevékenységét irányítják. Az egyes országok felelősek saját dokumentumtermésük és néhány rendszeren kívüli ország dokumentumainak referálásáért és indexeléséért. A referátumok a rendszer 3 hivatalos nyelvéné valamelyikén készülnek, az indexeléshez a háromnyelvű tezsaurusz deskriptorait használják. Ezután az információ valamelyik koordinációs központba kerül, itt mágnesszalagra viszik, majd az OECD veszi át az információ terjesztését (mágnesszalagon, számítógépen nyomtatott formában vagy speciális bibliográfiai kártyákon). Az információhoz hozzájutó országoknak joguk van a belátásuk szerinti hasznosításra.

Az OECD mint vezető szerv a következő funkciókat látja el: ellenőrzi, hogy a beérkező információ megfelelő a formátumnak, kiszűri a duplumokat, a dokumentumokat nyilvántartási számmal látja el, az információt mágnesszalagra viszi és azt havonta egyszer megküldi előfizetőinek. A formátumot az input és output információ fajtájától függően választják ki, valamennyi alkalmazott formátum kompatibilis egymással.

A rendszer egyik fő sajátossága, hogy *lehetőség van a manuális és automatizált alrendszer együttes használatára*. Ezt azzal érik el, hogy az információ kétféle hordozón – mágnesszalagon és bibliográfiai kártyán – kerül kihozatalra. A rendszerben az információfeldolgozás is kétféle: manuális és automatizált.

Az automatizált és manuális rendszerek ilyen kombinációja, melynek során a dokumentumleírás egyidejűleg mágnesszalagra és lyukkártyára kerül, a keresés pedig bármely állományban folytatható, egyszerű és kényelmes megoldásnak látszik.

Azokban a nemzetközi információkereső rendszerekben, amelyekben a dokumentumokat és a kérdéseket különböző nemzeti központokban indexelik, nagyon fontos az indexelési szabályok egységesítése. Egyik ilyen jellegzetes szabálygyűjtemény az ISONET indexelési útmutatója (*Manuel d'indexation ISONET. ISO/INFO/GT I 228*) c. kiadvány. Az ISONET is egyik példája az integráltság harmadik szintjén levő rendszereknek.\* Az ISO (*International Organization for Standardization*) információs központjában dolgozták ki a nemzeti és nemzetközi szabványok retrospektív keresése céljából.

Kezdetben úgy tervezték, hogy az információs központ által feldolgozandó dokumentum-mennyiség kb. 15 ezer tétel, az évenkénti gyarapodás pedig 1500–1700 dokumentum lesz. A valóságban azonban az iparilag legfejlettebb országok szabványainak felvételével a feldolgozandó dokumentumok száma kb. 50 ezerre nő, az évenkénti gyarapodás pedig 5000–6000 tételt fog kiadni. Nyilvánvaló, hogy egyetlen központ ezt a mennyiséget a bonyolult tartalom és a nyelvi különbségek miatt nem lett volna képes feldolgozni. Ez vezetett az ISONET rendszer decentralizált formában való létrehozásához.

A rendszernek biztosítania kell

- a központok közötti információcserét manuális és automatizált információfeldolgozással;
- a dokumentumok és a kérések indexelését;
- az üzemeltetőtől függetlenül a dokumentumok feldolgozásához egységes adatlapok használatát;
- a szabványokról és a vállalatok műszaki feltételeiről szóló információ mellett a normatív jellegű kormányhatározatokról szóló tájékoztatást is.

Az ISONET működését főként az ISO teaurusza és az ISONET indexelési mutatója orientálja, határozza meg.

A normatív–műszaki dokumentumokat egységes követelmények szerint speciális adatlapon indexelik. Az adatlapon 38 mező van, ebből 15 kötelezően, 23 fakultatívan töltendő ki. Így sikerül figyelembe venni az információs rendszerek nemzeti sajátosságait. Az indexeléshez *deszkriptorokat* vagy *alfanumerikus kódokat* használnak. Az indexelés célja:

- a dokumentum formájának leírása;
- a tartalom leírása;
- a más dokumentumokkal való kapcsolatainak leírása.

A *szabvány formájának leírása* a következő elemeket tartalmazza:

- a nyilvántartásba vétel azonosítója;
- a dokumentum azonosítója;
- a dokumentum adminisztratív jellemzői;
- bibliográfiai leírás.

A formális bibliográfiai leíráshoz csak alfanumerikus kódokat használnak.

\* Vö.: VÖLGYES Tamás: Az ISONET, a szabványosítás nemzetközi információs hálózata = TMT, 27. köt. 7–8. sz. 1980. p. 280–287.

A szabvány tartalmi leírásának a szelektív információterjesztést kell lehetővé tennie. Mivel az összehasonlító elemzésekhez a szabványok faktografikus leírását is tervezik, a szabványok tartalmát az ISO teaurusz deszkriptoraival, a tárgyi szakrend kódjaival, az ETO-jelzések segítségével és a szabvány alkalmazási területét jelölő speciális kódokkal egyaránt leírják.

Végül az adatlap utolsó mezői a dokumentumok közötti kapcsolatokat tükrözik: egyik dokumentum helyettesítheti vagy kiegészítheti a másikat, lehet egy nemzetközi szabvány továbbfejlesztése, közel állhat ugyanazon ország szabványaihoz, vagy más országok nemzeti szabványaihoz.

A szabványoknak még két olyan jellemzőjét kell megemlíteni, amelyek pótlólagos követelményeket jelentenek az indexeléssel szemben. Ezek egyike a *formális jellemzők változékonysága* (pl. a szabvány lehet kidolgozás alatt, elfogadva, felülvizsgálat alatt, végül érvénytelenítve). A másik a *szabványok kimunkálásának rendszeresítése*, ami formailag a szabvány szövegében levő utalásokban tükröződik. Ezért egy-egy szabványban történő bármilyen változást az összes vele kapcsolatos szabványban át kell vezetni. Mindez a szabványállomány különleges, dinamikus jellegét tanúsítja és magas szintű követelményeket támaszt az ISONET szervezésével és működtetésével szemben.

A korszerű nemzetközi információkereső rendszerek többségét a manuális indexelés jellemzi. Az utóbbi időben azonban több, különböző nyelvekkel operáló automatikus indexelésű nemzetközi információkereső rendszert is kifejlesztettek. Az automatikus indexelésnek két szintjét lehet megkülönböztetni:

1. a referátumból a keresőkép kialakítása a cél,
2. az indexelést a referátum teljes szövegének a rendszer másik hivatalos nyelvére fordítása egészíti ki.

Az első szintű indexelést alkalmazza a kétnyelvű – angol és francia – *Alliance rendszer* (az Institut du Verre, a Norsk Senter for Informatikk és az Institut de Ceramique Français fejlesztése). A rendszer a következő feladatokat látja el: a számítógép segítségével készült rövid francia referátum automatikus angolra fordítása, az automatikus indexelés során a deszkriptorok kiválasztása, betűrendes mutatók összeállítása, szelektív információterjesztés; egyes témákra bibliográfiai jegyzékek készítése.

A rendszer működésének elve a következő: egységes szabályok szerint manuálisan készül a referátum, számítógépbe vitele során ellenőrzésen megy keresztül. A számítógép a tiltott szavakat kihagyja, egy szó különböző formáit azonos alakra hozza, a nem deszkriptorokat deszkriptorokkal helyettesíti. Ezután a francia nyelvű deszkriptorokat az angol megfelelővel váltja fel, így két egyforma keresőkép jön létre. Ellenőrzési célokra a számítógép kiadja azoknak a referátumbeli szavaknak a jegyzékét, amelyeket nem talált a teauruszban, valamint a tiltott szavak jegyzékét.

Az Alliance rendszerben alkalmazott automatikus indexelési módszer, amelyet a referátum szó szerinti fordítása kísér, nem hibátlan. Az ily módon készült referátumok referáló lapokban nem közölhetők és a display-n is nehezen olvashatók. Ezért egyre inkább terjednek azok a rendszerek, amelyekben az automatikus indexeléssel egyidejűleg megtörténik a referátum másik nyelvre fordítása is. Az ilyen rendszerek a célnyelven is összefüggő szöveget adnak.

Az ilyen feldolgozást végző nemzetközi információkereső rendszerek tipikus példája a Francia Textilipari Intézet által kidolgozott *TITUS rendszer*. A rendszer célja a szigorú szabályok szerint írt referátum vagy a dokumentum teljes szövegének automatikus fordítása. Segítségével a német, angol, francia és spanyol nyelvű dokumentumokból egységes állományt lehet kialakítani, illetve információt szolgáltatni e nyelvek bármelyikén.

A rendszerben Franciaország mellett részt vesz az NSZK, Belgium, Spanyolország, Olaszország, Argentína és az USA. A rendszerbe évenként 25 ezer dokumentum kerül bevitelre. Lehetőség van keretében a szelektív információterjesztésre és az eseti kérések megválaszolására is. A TITUS I. rendszer 1969–1970-ben működött. 1972-ben kezdett működni a referátumok automatikus fordítását végző TITUS II. rendszer.\* 1974-ben indult a TITUS III. alrendszer, amely *Franciaország és az NSZK számára biztosítja a párbeszédes üzemmódot*.

A TITUS II–III. rendszer részei:  
 speciálisan strukturált szótár,  
 általános nyelvészeti modell,  
 4 genitív nyelvtan (az egyes nemzeti nyelvek számára),  
 keresési modell,  
 4 transzformációs nyelvtan a fordító alrendszer számára.

A keresés szempontjából az ilyen rendszer azért értékes, mert lehetővé teszi a rendszer bármely nyelvén írt kérés bevitelét, elemzését és „klasszikus”, deszkriptorokból és szintaktikai relációkból álló profilokra fordítását.

A referátum szövegének fordítása és a többnyelvű szótárban való keresés annyira hatékony, hogy az átlag 10 mondatból álló referátum lefordításához, amelynek mindegyik átlag 10 szót tartalmaz, mindössze 2 másodperc szükséges.

Nagyon fontos kérdés a *többnyelvű teaurusz karbantartása*. Általában ez a következőket kívánja meg:

1. az új, eddig a teauruszban nem szereplő fogalmat jelölő kifejezés bevitelét;
2. az új relációk bevitelét a szükséges fogalmak keresésének megkönnyítésére;
3. a teauruszban már szereplő deszkriptorok újfajta hasznosíthatóságát jelölő megjegyzések bevitelét;

4. a deszkriptorok jelölésének módosítását;
5. a szükségtelenné vált deszkriptorok törlését.

A többnyelvű teaurusz karbantartásához célszerű számítógépet igénybe venni, mivel a teaurusz különböző nyelvi változataiban átvezetendő módosítások száma igen nagy lehet. Így pl. az IRRD teaurusza 350 deszkriptorának vizsgálata a teauruszban 2500 módosítást vont maga után.

A többnyelvű teaurusz számítógépes karbantartása a következő műveletek elvégzését teszi szükségessé:

- az indexelő szakember által használt kifejezések ellenőrzése a szótár alapján;
- a tiltott fogalmak automatikus helyettesítése;
- a homográfok kinyomtatása a manuális ellenőrzéshez;
- a generikus fogalmaknak a specifikus fogalmakhoz való automatikus hozzárendelése;
- egyazon deszkriptor többszöri használatakor a szövegben csak egyetlen egynek való megtartása;
- a leírásban különböző nyelveken szereplő fogalmaknak a bázisteaurusz fogalmaival helyettesítése;
- az új fogalmak kinyomtatása a kereséshez.

A teaurusz karbantartása során nagy szerepe van a *terminológiai ellenőrzésnek*. Ennek okai:

- az ekvivalencia hibás megállapításának lehetősége;
- a szótárban új fogalmak megjelenése;
- a deszkriptorok jelentésének természetes változása a nemzeti változatokban, ami az ekvivalencia felbomlásához vezet.

A terminológiai ellenőrzéshez valamennyi eszköz használható: a terminológia szabványok, értelmező szótárak és enciklopédiák, többnyelvű és egynyelvű teauruszok. Pl. az új fogalomnak az INIS teauruszba vitelekora feltüntetik azt a dokumentumot, ahol a fogalmat találták. Ez a lehetőség elősegíti az új fogalom pontos megértését és az alkalmazási területek meghatározását.

A néhány működő nemzetközi információkereső rendszer példáján látható, milyen követelményeknek kell az ilyen rendszereknek megfelelniük. Ezek a következők:

1. már a tervezés szakaszában meg kell határozni az integráltság szintjét;
2. biztosítani kell a nyelvi akadályok leküzdését;
3. ki kell dolgozni a rendszer egységes adatlapjait és azok kitöltési szabályait;
4. a rendszer karbantartásához egységes szabályokat kell elfogadni;
5. előzetesen létre kell hozni a kompatibilis műszaki és a kommunikációs csatornák hálózatát.

/PIGUR, V. A.: *Mnogojazücsnue IPSZ: urovni integracii i jazükovoe obeszpecsenie = Naucsno-Tehniczeszkaja Informacija*, 2. sor. 1. sz. 1979. p. 22–28./

(Viszocsekné Péteri Éva)

\* Vö.: Tapasztalatok a TITUS II. információs rendszerrel = TMT, 26. köt. 12. sz. 1979. p. 531–535.