

30. évf. 12. sz. 1983. december

Tudományos és Műszaki Tájékoztatás

A FAKTOGRAFIKUS INFORMÁCIÓKERESŐ RENDSZEREK NÉHÁNY ELMÉLETI KÉRDÉSE

Dr. Kovács Árpád—Nagy Péter

Ipari Informatikai Központ

A tudományos—műszaki információs tevékenység jelenlegi fejlődési iránya az információs szolgáltatások formáinak és módszereinek bővülését tükrözi. A szakirodalmi információk szolgáltatásának igénye mellett egyre több használói igény vonatkozik közvetlenül a tényekre és adatokra, a primer dokumentumok igénybevétele nélkül. Ilyen típusú információs igény a tudományos és műszaki, a társadalomtudományi és a gazdasági területen egyaránt jelentkezik. Az igényeket közvetlenül tényekkel és adatokkal kielégítő rendszereket általában *faktografikus információkereső rendszereknek* nevezik.

Cikkünk célja az alapvető idevágó problémák bemutatása egységesen értelmezett fogalmak alapján, fenntartva a létszám jogát, hiszen az információkereső rendszerek elmélete még koránt sincs kiforvva. Ezen belül is a *faktografikus információkereső rendszerek* egzakt elméletének a létrehozása egyelőre az informatika távlati adatai közé tartozik.

és tényszerű (faktografikus) információ

„tény”-t (faktumot) filozófiai, konkrét tudományos és informatikai szempontból határozhatjuk meg. Informatikai szempontból a kérdés az, hogy a tény maga objektív jelenség, avagy „csak” az objektív jelenséget tartalmazó információ(k). Az informatikában célszerűnek tűnik a véleményen az *objektív valóság azon jelenségeit érteni, amelyekről információink vannak*. Ebből a meghatározásból következik, hogy

a tény az objektív jelenség, és nem a róla szóló információ;

nem minden jelenség tény, hanem csak azok, amelyekről információink vannak.

A második megállapítás tartalma az, hogy a számunkra ismeretlen tények nem rendelkeznek azzal a tulajdonsággal, hogy az ember felhasználhatná őket döntések meghozatalához, tudományos elméletek felállításához, megalapozásához, illetve megdöntéséhez stb.

A faktografikus információ fogalmának bevezetése azt vonja maga után, hogy egyidejűleg meghatározzuk a „nem-faktografikus” információt is, amely körbe ebből a szempontból minden más információfajta beletartozik. Választ kell adnunk arra a kérdésre is, hogy hogyan viszonyul a faktografikus információ az információ más típusaihoz.

Faktografikus információn egyszerűen konkrét ismereteket, adatokat értünk, függetlenül attól, hogy azok ténylegesek, hipotetikusak, illetve prognosztikai jellegűek. A közös bennük az, hogy konkrét tárgykörre, és nem az ezeket tükröző dokumentumokra vonatkoznak. A konkrét dokumentumokat leíró ismereteket *dokumentografikus* információnak nevezzük, formái pedig a bibliográfiai leírás, referátum, annotáció stb.

Gyakorlatilag faktografikusnak nevezhető az az információ, amelyhez a felhasználók döntő többsége nem közvetve a dokumentumokat leíró adatok révén akar hozzájutni, hanem a dokumentumok hordozta információ közvetlen használatával.

Felvetődik a kérdés: célszerű-e így megalkotni a faktografikus információ fogalmát, ha ennek következtében bizonyos adatszerű ismeretek (mint például a bibliográfiai leírás) nem-faktografikusnak minősülnek, viszont sok, látszatra nem-faktografikus ismeret faktografikusnak? A faktografikus információ szó szerinti jelentése: tényleíró információ, de a nem-tényszerű adatok (tervadatok, hipotézisek stb.) szintén leírhatók, mint tények. Tehát a faktografikus információ fogalma

kitágul: tartalmazza nemcsak a tények leírását, hanem minden olyan mást is, ami tényszerűen leírható. Szem előtt kell azonban tartani egy gyakorlatias követelményt: nem minősíthető faktografikusnak a dokumentografikus információ, amely közvetítő szerepet játszik a kommunikációs folyamatban.

A faktografikus információk osztályozása

A faktografikus információkat többféle szempontból lehet osztályozni, egybevégoan az információosztályozás elveivel. Ennek megfelelően a faktografikus információt megkülönböztetjük tárgya szerint (ágazati, tudományi stb.), beszélünk hipotetikus, nem hipotetikus és prognosztikai információkról, továbbá potencionális és tényleges, mennyiségi és minőségi, valamint egyedet, folyamatokat, rendszereket stb. leíró információkról.

A feldolgozottság foka szerint a faktografikus információ lehet elsődleges (pl. amelyet a műszerek szolgáltatnak) vagy másodlagos (amelyet az információs rendszer közvetítésével lehet megszerezni).

Más osztályozási szempont szerint lehetséges *a priori* jellegű faktografikus információ, amelyet az információs rendszer eleve tartalmaz (pl. „faktografikus” teaurusz formájában) még az üzemeltetés megkezdése előtt, vagy *a posteriori* jellegű, amely az információs rendszer működése közben halmozódik fel.

A fenti osztályozási kategóriák a faktografikus információk tartalmával függenek össze, de lehetséges a *forma szerinti* osztályozás is (pl. az információhordozók típusai, az információhordozó nyelv stb. szerint).

Fontos gyakorlati szerepe van az *erősen strukturált* és a *gyengén strukturált* információkra történő felosztásnak. Az erősen strukturált faktografikus információkhoz sorolhatók mindenekelőtt az adott struktúrájú mennyiségi és minőségi jellegű ismeretek: paraméterek és értékeik, megnevezések, címek stb. Ezek általában kérdőív, táblázat, vektor stb. formájában szervezhetőek meg; az ilyen típusú információkat *parametrikus információknak* is nevezik.

Gyengén strukturált faktografikus információk azok az ismeretek, amelyek nem adhatók meg határozott szóbeli konstrukciók formájában. Ilyenek például a koncepciók és elméletek kifejtései természetes nyelven.

Az erősen strukturált információk feldolgozását könnyű formalizálni, és ebből következően automatizálni. Ehhez a típushoz tartozik a termelést leíró adatok túlnyomó többsége. Ezzel magyarázható az a tény is, hogy ezen a területen történtek először lépések *faktografikus információkereső rendszerek (FIR)* létrehozására.

Talán meg lehet kockáztatni azt az általánosítást, hogy a működő FIR-ek erősen strukturált információkat dolgoznak fel. A gyengén strukturált információkat

feldolgozó FIR-ek létrehozása még távoli jövő; ez abból következik, hogy a természetes nyelvű szövegek formalizálása még megoldatlan.

A faktografikus információk feloszthatók *potenciális* és *tényleges* információkra is. Potenciális információ az, amely eljutott a konkrét felhasználóhoz. A potenciális információ akkor válik tényleges információvá, amikor az igénylő felhasználja. A tényleges információ fogalma relatív, mivel az információ értelmezése függ a felhasználó teauruszától, valamint időszerű információszükségletétől, amelyet teauruszára támaszkodva kérdés formájában fogalmaz meg. Az információkeresés eredményeként válik el, hogy a potenciális információ ténylegesen megfelel-e az adott kérdésnek.

Összefoglalva a fentieket, a faktografikus információkereső rendszer olyan rendszerként határozható meg, amely faktografikus információkat gyűjt, tárol, feldolgoz, visszakeres; azokat az információkat nyújtja, amelyek a felhasználónak a rendszerhez intézett kérdését ténylegesen kielégítik a primer dokumentumok igénybevétele nélkül.

A faktografikus információkereső rendszerek osztályozása

A FIR-ek rendeltetésük szerint problémára orientált vagy általános, információs bázisuk megoszlása szerint egy vagy több helyre elosztott, a felolelt információk típusa alapján kizárólag faktografikus információkat feldolgozó vagy faktografikus és szakirodalmi információkat egyaránt feldolgozó rendszerek lehetnek.

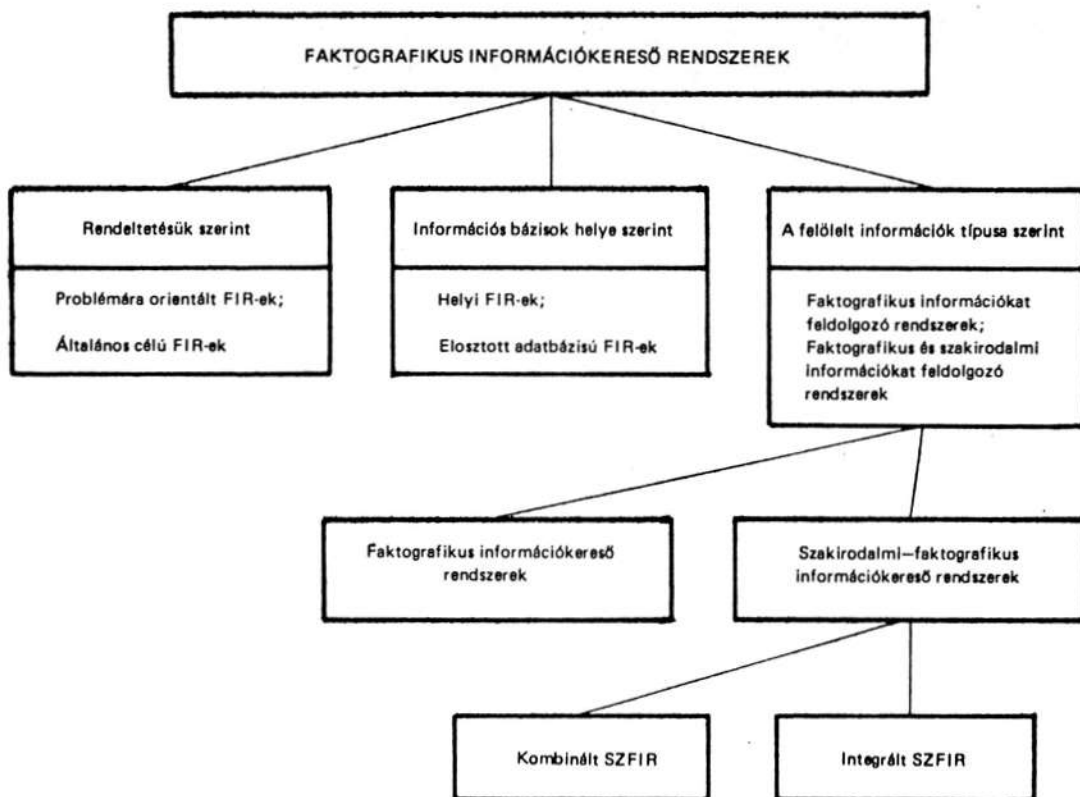
A fenti ismérvek alapján beszélhetünk „tisztá” FIR-ekről és vegyes, szakirodalmi–faktografikus információk rendszerekről (1. ábra).

A megoldandó feladatok alapján, az információkereső nyelvek felépítése, a szolgáltatott információk, valamint a tartalmazott információs elemek bonyolultsági foka alapján történő osztályozást a 2. ábra szemlélteti.

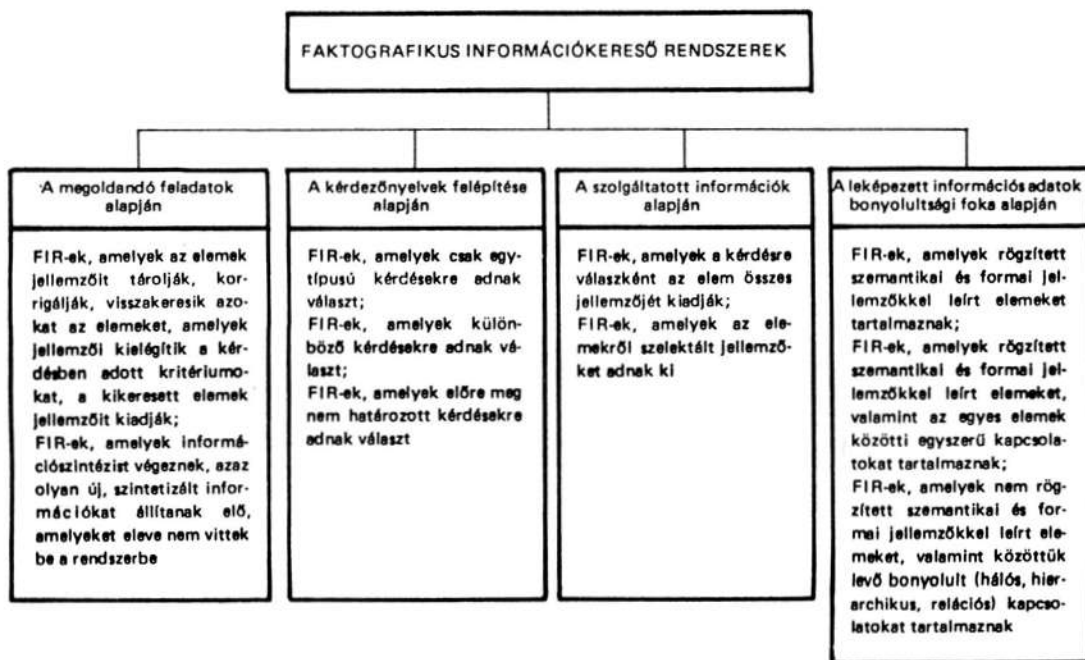
A FIR-ek kifejlődése során ún. generációkat különböztethetünk meg.

Első generációs FIR-ek azok, amelyek az elemeket és az őket leíró jellemzőket csak tárolják, karbantartják, egyetlen típusú kérdésekre végeznek keresést, s a kérdésre az elem teljes információs spektrumát adják válaszul. A kért elemek egy osztályba tartoznak, rögzített szemantikai és formai jellemzők halmazával írhatóak le.

A *második generációs* FIR-ek tárolják, karbantartják az elemek jellemzőit, rögzített kérdéstípusra adnak választ, de válaszukban válogatnak a jellemzők szerint, s képesek elvégezni az információsintézis meghatározott feladatait is. A bennük levő elemek különböző osztályokat alkothatnak, s minden elemtípushoz rögzített jellemzők tartoznak. Az input–output, a keresés, a korrekció folyamatai automatizáltak.



1. ábra A faktografikus információkereső rendszerek felosztása



2. ábra A faktografikus információkereső rendszerek osztályozása

A harmadik generációs FIR-ek az elemek jellemzőit úgy tartalmazzák, hogy képesek változtatásokat átvezetni az egész elemsoron, s képesek előre meg nem határozott kérdéstípusokra is keresést végezni, s ugyancsak előre meg nem határozott módon információszintézist végrehajtani. A bennük levő elemek különböző osztályokat alkotnak, amelyeket nem rögzített formájú tulajdonsághalmaz ír le. Az input–output, a keresés, a korrekció, az információszintézis folyamatai automatizáltak, hasonlóképpen a felhasználóknak a rendszerrel való együttműködése is.

A szakirodalmi–faktografikus információkereső rendszerek (SZFIR) két típusba sorolhatók:

1. kombinált SZFIR-ek, amelyekben

- a szakirodalmi és a faktografikus információk különböző típusokhoz tartoznak, de az elemtípusok között a rendszer felhasználói számára lényeges logikai kapcsolat van;
- a faktografikus információk feldolgozása független a szakirodalmi információktól;

2. integrált szakirodalmi–faktografikus információkereső rendszerek (ISZFIR), amelyekben

- a szakirodalmi és a faktografikus információk egyaránt vonatkoznak egy-egy elemre;
- a faktografikus információk feldolgozása nem választható el a szakirodalmi dokumentumok feldolgozásától;
- a faktografikus információk további lehetőségeket nyújtanak a szakirodalmi dokumentumok keresésénél.

Adatbázis, adatbank, automatizált FIR*

A FIR-ekkel kapcsolatban felvetődik az a kérdés is, hogy mit értünk „adatbázison”, „adatbankon”, mivel ezeket a fogalmakat is eltérően értelmezik a szakemberek.

A számítástechnika kezdeti időszakában az adatokat szekvenciális fájlban tárolták a mágnesszalagokon, s a feldolgozó programok tartalmazzák az adatléírásokat; emiatt a programok és az adatok közötti kapcsolat erős volt: bármilyen változás az adatokban maga után vont a programok változását is. Az adatok módosítása a mágnesszalagok többszörös másolását vont maga után. Egy-egy fájl egy-egy feladat megoldását szolgálta, s ezért ugyanazon adatokból különböző fájlokat kellett létrehozni, más-más struktúrában, más-más feladat megoldására. Mindez sok nehézséget okozott az adatfeldolgozásban, nem is szólva a tárolóeszközök gazdaságtalan kihasználásáról.

* Az adatbank fogalmát eltérően értelmezik a szakirodalomban. Így pl. a szerzők itt közölt meghatározása nem azonos NOVÁK T.–ROBOZ P. a TMT e számában közölt cikkében javasolt definícióval – A szerk.

Ebből a helyzetből csak egy kivezető út volt: el kellett különíteni az adatokat a felhasználói programoktól, és úgy megszervezni az adatok és a köztük levő kapcsolatok tárolását, hogy lehetővé váljék *ugyanazon adatok többféle felhasználása*. Ennek az elképzelésnek a realizálása eredményeképpen jelent meg az *adatbázis* fogalma. Ezen olyan, egymással összefüggő adatok halmazát értjük, amelyet minimális redundancia jellemz, s az adatokat fel lehet használni egy vagy több feladat megoldásához; az adatok tárolása független a feldolgozó programoktól, s az új adatok bevitele, a régiek módosítása, valamint a visszakeresés általános adatkezelési rendszerrel történik.

Az adatbázisokat általában feladatok vagy a felhasználók típusai szerint hozzák létre. Az adatbázis struktúrájából a felhasználónak szükséges struktúrába történő átmenetet automatikusan végzi el az adatbáziskezelő rendszer. Az adatbázisnak alkalmazkodnia kell a más-más felhasználó képében jelentkező változásokhoz, s éppen ez a folyamatos változásra való képesség az adatbázisok legfontosabb jellemzője. A módosításokat gyorsan végig kell vinni az adathalmazon, s lehetőleg úgy, hogy az alkalmazói programokat ne kelljen megváltoztatni.

Az adatbázisoktól jutunk el az *adatbank* fogalmához, amelyen az adatbázisok halmazát és a felhasználásukhoz szükséges eszközöket értjük. Az adatbankok három alapvető összetevője a következő:

adatbázis(ok);

az információknak az adatbázis(ok)ba történő bevitelét, módosítását, visszakeresését, átalakítását biztosító programok összessége;

katalógusok, szótárak, amelyek az információk állomány leírását tartalmazzák.

Az adatbank tehát az automatizált FIR szervezési módja, amely biztosítja az információk állományok létrehozását, karbantartását és a kérdésekre történő válaszok kialakítását.

Az automatizált FIR-ek az *automatizált irányítási rendszerek* szerves részei. A vezetés szempontjából a FIR tevékenysége önmagában nem zárt: az információszolgáltatást döntéshozatal követi. Az automatizált irányítási rendszerekben létrehozott FIR-ek csak a vezetés információellátását biztosítják, ezért bennük elsődlegesen jól strukturált faktografikus információk feldolgozása folyik. Bemenetüket olyan dokumentumok képezik, melyeknek tartalma és struktúrája előre meghatározott (táblázatok, kérdőívek stb.), s előre ismert kérdések megválaszolására kell felkészülniük.

Az automatizált irányítási rendszerekben bizonyos típusú FIR-ek információk állományai a feltett kérdésektől függően a működés során állandóan változnak (pl. helyfoglalás repülőgépen). Az automatizált irányítási rendszerekben létrehozott FIR-ekre jellemző még a folyamatos, tartalmi jellegű adatváltozás, amely az irányított objektum jellemzőinek változásától függ.

A FIR-ek működésének egyes kérdései

Nyelvi–logikai alapok

A korszerű FIR-ek lingvisztikai alapjait az úgynevezett elem- és attributum (jellemző tulajdonság) típusú információkereső nyelvek alkotják. A faktografikus információban általában (az erősen strukturáltak esetében mindig) a tárgykör elemeit jellemzők (attributumok) halmaza írja le, amelyek különböző értékeket vehetnek fel az egyes elemekre vonatkozóan. A faktografikus információ minimális egységét triadaként lehet meghatározni: elem – attributum – az attributum értéke.

Az elem–attributum típusú információkereső nyelvek egyszerű példájául szolgálhatnak a táblázatok. Az erősen strukturált ismeretek mindig ábrázolhatók táblázatok formájában, például a sorok az elemeket, az oszlopok az attributumok megnevezéseit, találkozási pontjaik pedig az attributum értékeit tartalmazzák az adott elemre vonatkozóan.

Az elem–attributum típusú nyelvek eléggé szegényesnek bizonyultak, mivel nem teszik lehetővé minden szükséges információ visszatükrözését, például az ugyanarra az elemre vonatkozó különböző attributumok közötti kapcsolatokat, illetve az attributumok értéke közötti kapcsolatokat az elemre vonatkozóan. Továbbá, az elem–attributum típusú nyelvek alkalmatlanok azokban az esetekben, amikor a különböző elemeket leíró attributumok halmaza nem esik egybe. Ezekben az esetekben célszerű olyan elem–attributum típusú nyelveket használni, amelyek lehetővé teszik változó hosszúságú leírások (rekordok) létrehozását. Az ilyen típusú információkereső nyelvek szintakszisa jóval bonyolultabb, de alapjául akkor is a minimális triáda szolgál: elem–attributum–attributum értéke. Ez megfelel a természetes nyelven megfogalmazott kijelentéseknek is: „*X* elem *A* attribútuma *a* értéket vesz fel”.

Az a követelmény, hogy az információkereső nyelv segítségével eléggé bonyolult faktografikus információkat lehessen leírni, szükségessé teszi az ilyen típusú információkereső nyelvek szemantikájának mélyebb kidolgozását. Bármely információkereső nyelv lexikai–szemantikai gazdagságát annak a rendszernek a teaurusza tükrözi, amely ezt a nyelvet használja.

Ha összehasonlítjuk a FIR-ek és a *szakirodalmi információkereső rendszerek (SZIR)* nyelveit, akkor a következő alapvető különbségeket látjuk: a SZIR-ekhez általában deskriptor típusú nyelveket használnak, a FIR-ekhez pedig többnyire elem–attributum típusú nyelveket.

Az automatizált FIR teaurusza

Egy automatizált FIR elképzelhetetlen teaurusz, valamint egyéb terminológiai eszközök nélkül.

Az automatizált FIR teaurusza struktúrája jóval bonyolultabb, mint egy SZIR-é; míg az utóbbi esetében a bemenő dokumentumok feldolgozásakor a dokumentumok tartalmi tömörítéséről van szó, azaz az alapvető fogalmi tartalom meghatározásáról, addig az előbbi esetében – mivel a felhasználó számára lényegesek lehetnek a legapróbb, a legrészletezettebb, a tárgyhoz képest periférikusabb ismeretek, adatok is – az információ-tartalom lehető teljes feltárásáról.

A faktografikus teaurusz jellegzetessége még az is, hogy információkereső nyelvének szemantikájában a lexikai egységeknek nem egy, hanem többféle lexikai–szintaktikai osztálya található. (A deskriptor típusú nyelvekben ezzel szemben csak egyféle lexikai osztály van: deskriptorok; a nem-deszkriptorok nem képezik a nyelv lexikai egységeit.) Az elem–attributum típusú nyelvekben legalább három típusú lexikai egység található: az elemek megnevezései, az attributumok megnevezései és az attributumok értékeinek megnevezései.

A FIR teauruszának hármas szerepe van:

1. *Információkereső teauruszként funkcionál.* Általában az információkereső teaurusz rögzített kapcsolatokkal rendelkező deskriptorok – kulcsszavak – normatív szótára, amelynek rendeltetése a dokumentumok és az információk kérdések koordinált indexelése. A faktografikus információkereső rendszerekben nem használják tiszta formájában a koordinált indexelést, ezért ilyen értelemben egy FIR-ben nem lehet beszélni az információkereső nyelv deskriptorairól és kulcsszavairól. Ellenben, ha a deskriptorok meghatározását egyszerűen úgy tekintjük, mint a lexika normalizálását és a szinonimák megszüntetését, akkor az elem–attributum típusú nyelvek minden lexikai egysége deskriptornak vehető. A faktografikus teauruszban is kötelező a lexikai egységek közötti legfontosabb fogalmi relációk megteremtése.

2. *A felhalmozott tudáshalmaz, az ismeretek rendszerének leírását adja.* A faktografikus teauruszt úgy is lehet tekinteni, mint az ismeretek azon halmazát, amellyel a FIR „rendelkezik”.

3. *A FIR értelmező szótára.* Ahhoz, hogy az információkereső nyelv teljes és átfogó leírása legyen, szükséges az is, hogy tartalmazza nemcsak a nyelv nyelvtanát, hanem az értelmező szótárát is.

A faktografikus információk indexelése

A FIR-ekben a dokumentumok indexelésén a faktografikus információk kijelölését, az információknak az információkereső nyelvre való lefordítását és adatlapra történő rögzítését értjük. Legcélszerűbb, ha az indexelés az adott szakterület (targykör) magasan képzett szakemberei végzik.

Az indexelés függ a feldolgozandó dokumentumok típusától, maguktól a feldolgozandó információktól, valamint a felhasznált információkereső nyelvtől.

Attól függően, hogy milyen típusú dokumentumok adják a FIR bemenetét, különböző nehézségű lehet a faktografikus információk indexelése. Az adatokat szervert formában tartalmazó dokumentumokból (pl. kérdőívekből) könnyű indexelni, míg szövegeket tartalmazó publikációk (cikkek, könyvek stb.) faktografikus információinak indexelése szükségessé teszi maguknak a szövegeknek a szemantikai elemzését is.

Az indexelendő faktografikus információk fajtájától függően is különbözik egymástól az indexelési folyamat. Ha a feladat például a szövegben kifejezett mennyiségi jellegű parametrikus adatok, berendezések típusai, földrajzi elnevezések stb. feltárása, akkor ez esetben az indexelés elvileg nem különbözik a SZIR-ekre jellemző koordinált indexeléstől. Ha azonban az a feladat, hogy minden szükséges információt ki kell emelni a dokumentumokból, akkor az indexelőnek a szöveg szemantikai elemzését is el kell végeznie, azaz meg kell állapítania minden információt hordozó elemet, azok tulajdonságait és az elemek közötti kapcsolatokat.

Az alkalmazott információkereső nyelvtől is függ az indexelés művelete. Ha deskriptor típusú nyelvet használunk — ami eleve meghatározott elemek indexelésénél alkalmazható — akkor megfelelnek a koordinált indexelés módszerei. Gyakoribb az az eset, amikor elem-attributum típusú nyelvet használnak, ami az indexelőtől azt követeli meg, hogy maga állapítsa meg az elemeket, attributumaikat és azok értékeit. Ilyenkor szükség lehet az alkalmazott információkereső nyelvtől eltérő nyelven történő szövegértelmezésre. Ezt a közbülső nyelvet közvetítőként is alkalmazzák a felhasználó és az információkereső nyelv között, ami lehetővé teszi a felhasználóknak, hogy az információkereső nyelv ismerete nélkül is hozzáférhessenek a rendszerhez. Ilyen közvetítő nyelvként alkalmazzák például a szabványosított kifejezéseket.

A dokumentumokból kigyűjtött információs elemek keresőképét az adott információkereső nyelven alkotják meg. Eltérően a dokumentumok keresőképétől — amelyet egyszer s mindenkorra szerkesztettek meg, hiszen a dokumentum publikálása után már nem változik — az elemeket leíró faktografikus információk módosulhatnak a FIR működésének egész ideje alatt, mint hogy az elemekre vonatkozó ismereteink egyre mélyülhetnek, gazdagodhatnak, vagy idővel az elemek paramétereit és jellemzőit is megváltozhatnak.

A FIR ugyanazon elemét jellemző információk különböző forrásokból is származhatnak, ezért szükséges az „információs szegmensek” összekapcsolása.

Gyakorlati szempontból, elsősorban a felhasználók igényeinek kielégítésére célszerű, hogy az elemek keresőképében legyen utalás arra a dokumentumra, amelyből az információk erednek.

Az információs kérdések feldolgozása a FIR bemenetén

Az információs szükségletek kielégítése érdekében arra kell törekedni, hogy a FIR-hez intézett kérdések — amennyire csak lehet — szabályozottak legyenek. Ezért nagy a jelentősége a lehetséges kérdések előzetes osztályozásának (1. táblázat).

A rendszerben tárolt faktografikus információk kijelentések (logikai ítéletek) halmazának tekinthetők (pl. az attributumok milyen értékei vagy az értékek milyen kombinációi jellemzőek az adott elemre). A FIR-ek esetében a kérdés feltevésekor nem elhanyagolható, hogy milyen egyedekről, mely attributumaiknak mely értékeiről, milyen lexikai kombinációban van szó. Hasonló a helyzet az indexeléskor is. Vagy fordítva: helyesen megérteni és leírni bizonyos ismereteket sokkal fontosabb, mint minél több ismeretet rögzíteni. Ugyanúgy a kérdésekre adott feleletek esetében is helytelen válasz helyett célszerűbb egyáltalán nem adni választ. Ezért is van nagy szerepe az interaktív rendszerek alkalmazásának a FIR-ek esetében is.

1. táblázat

A FIR-re jellemző információs kérdések főbb típusai

A kérdés	
Logikai formája	Tartalma
$A(E) = ?$	— Bizonyos attributum milyen értéket vesz fel adott elemre vonatkozóan
$A(?) \left\{ \begin{array}{l} = \\ \neq \\ < \\ > \end{array} \right\} \text{É}$	— Mely elemek veszik fel az attributum adott értékeit
$?(E) \left\{ \begin{array}{l} = \\ \neq \\ < \\ > \end{array} \right\} \text{É}$	— Melyek azok az attributumok, melyek az adott elemre vonatkozóan a megadott érték-halmazt veszik fel
$?(E) = ?$	— Az adott elemet jellemző összes attributum minden értéke
$A(?) = ?$	— Milyen értékei vannak az adott attributumnak minden elemre vonatkozóan
$?(?) \left\{ \begin{array}{l} = \\ \neq \\ < \\ > \end{array} \right\} \text{É}$	— Melyek azok az attributumok, melyek adott elemre vonatkozóan az adott értéket veszik fel

A keresési kritériumok

Az előbbieket alapján a faktografikus információállományt úgy foghatjuk fel, mint adott elemek és az őket jellemző ismeretek összességét. Minden elemet meghatározott attributumok értékeinek halmaza jellemez, ami

semmi más, mint az elem keresőképe. A kérdést viszont az információkereső nyelven megfogalmazott keresőkép jellemzi.

Az *elemek keresőképét* két mezőből szerkeszthetjük meg: az első az elemmező, amely tartalmazza az elem megnevezését; a második mező az attributummező, amely két részből tevődik össze: az attributum megnevezését és az attributum értékét (értékeit) tartalmazó almezőkből.

Az *információs kérdés keresőképe* hasonló szerkezetű, de lényegéből következően kétféle lehet:

adott elemekhez tartozó attributumok értékét vagy értékeit kell meghatározni;

adott attributumokhoz, illetve ezek értékeihez tartozó elemet (elemeket) kell megkeresni.

Az első esetben egyszerű a keresési kritériumok megadása: alapvető, hogy az elem megnevezése az elem és a kérdés keresőképében egybeesik.

Lényegesen bonyolultabb a keresési kritériumok megadása, ha az elemhalmazokban keresünk. Az elemhalmaz keresőképei altípusokra bonthatók:

a) mely elemeket jellemez az adott attributumhalmaz;

b) melyek azok az elemek, melyeket meghatározott attributumhalmazok adott értékei jellemeznek. A kérdés keresőképének altípusától függően a keresési kritérium egy- vagy többlépcsős lehet.

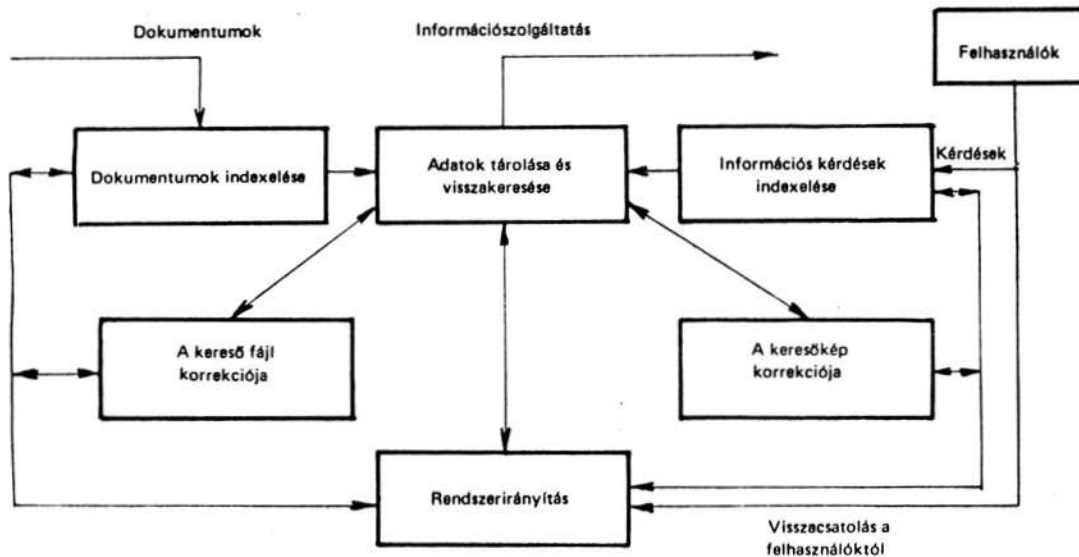
Az automatizált FIR-ek működési elvei

Az automatizált FIR alapvető funkciója: a faktografikus adatok gyűjtése, tárolása, információk szolgáltatása a felhasználók kérdései alapján, valamint az információs állomány naprakészen tartása.

A rendszer funkcionális blokkjait a 3. ábra szemlélteti.

A rendszer bemenetét dokumentumok adatai, információs kérdések és a felhasználóktól származó visszacsatolási információk képezik, kimenetere pedig a felhasználók kérdéseire adott válaszok.

A bemenet alapjául szolgáló dokumentumokból az adatokat indexelik, azaz lefordítják őket az információkereső nyelvre. Ezt követően az adatok a *Rendszerirányítás* nevű blokkba kerülnek, ahol döntenek afelől, növelik-e az újonnan indexelt adatok a meglévő állomány információtartalmát. Ha nem, akkor elutasítják az új adatok bevitelét a rendszerbe, ha pedig a szükséges informativitással rendelkeznek, akkor a megfelelő blokkba kerülnek (az adatok tárolását és visszakeresését, illetve kereső fájl korrekcióját végző blokkba). A korrekciót végző blokkban módosítják a már tárolt, az új információkkal kapcsolatosan módosítandó adatokat, amelyek aztán visszakerülnek a tárolási és visszakeresési blokkba.



3. ábra Az automatizált FIR-ek funkcionális sémája

Az információk keresőkérdések a kérdéseket indexelő blokkba kerülnek, itt alakul ki az indexelés eredményeképpen a kérdések keresőképe (keresőprofilja). A keresőkép a tárolási és visszakeresési blokkba kerül, ahol a keresés eredményeképpen megszületik a rendszer válasza.

A felhasználónak a választ értékelő reagálása (visszatolatása) az irányítási blokkba kerül, ahol döntenek a keresést javító intézkedésekről. Ha szükséges, akkor módosítják a keresőképet, azaz új profilt szerkesztenek, amely valószínűleg jobb keresési eredményekhez vezet. Ez az új, módosított keresőkép ismét a tárolási és visszakeresési blokkba kerül.

A dokumentumok és a kérdések indexelését végző blokkok jelenleg még nem automatizáltak: ezekben minden műveletet manuálisan végeznek, speciális módszertani anyagok és teauruszok segítségével.

A FIR-ek működésének értékelése

Mind a faktografikus, mind a szakirodalmi információkereső rendszerek esetében keresőkérdésekkel és a rájuk adott válaszokkal van dolgunk, azonban a válaszok jellege különbözik a két rendszerben.

A SZIR-től kérdező megtudja, hogy milyen dokumentumokat kell használni ahhoz, hogy választ kapjon az őt érdeklő kérdésekre, míg a FIR-ben közvetlenül az őt érdeklő tényeket kapja meg. Ezért teljesen más jelentősége van itt és ott a teljességnek és a pontosságnak. A SZIR-es esetében elfogadható a nem 100%-os pontosság, azaz ha olyan dokumentumokat is kap a felhasználó, amelyek számára nem relevánsak – ha ezáltal növekszik a keresés teljessége.

A nem megfelelő pontosság viszont a FIR-ek esetében *helytelen adatok szolgáltatását* jelenti. A felhasználó azonban azért fordul a rendszerhez, hogy bizonyos, számára ismeretlen tényekről ismereteket szerezzen; ezért nem lehet feltételezni róla, hogy képes észrevenni, ha a kapott adatok között hamisak is vannak. Következésképpen a FIR-rel szemben *abszolút követelmény a pontosság*. A teljességet jellemző mutató iránt, ami egy FIR esetében azt jelenti, hogy hány kérdésre tud válaszolni, engedékenyebbek lehetünk.

A FIR-ekhez intézett információk kérdést – betáplálása előtt – tökéletesen meg kell érteni, mert ha nem, pontatlan, azaz helytelen lesz a válasz, vagyis nem egyszerűen irreleváns, hanem téves, hamis, megtévesztő.

Felhasznált irodalom

1. ARTAMONOV, G. T.—CISTÁKOV, V. M.: Centry analiza informacii. Itogi nauki i tehniki. „Informatika” sor. 6. köt. Moszkva, VINITI, 1981. p. 3–159.

2. Avtomatizirovannye faktografičeskie informacionno-poiskovyje sistemy. (Osnovnye položenija i trebovanija). Moszkva, Gosudarstvennyj komitet SSSR po nauke i tehnike, 1979. 21 p.
3. Faktografičeskij poisk. Voprosy informacionnoj teorii i praktiki, 28. köt. 1. sz. Moszkva, VINITI, 1976. 119 p.
4. Gosudarstvennaja avtomatizirovannaja sistema naučno-tehničeskoi informacii: Katalog avtomatizirovannyh sistem. Moszkva, VINITI, 1980. 224 p.
5. KRISTAL'NYJ, B. V.: Faktografičeskij poisk v avtomatizirovannyh sistemah naučno-tehničeskoi informacii. (Metodičeskie posobie). Moszkva, 1982. 44 p.
6. KRISTAL'NYJ, B. V. i dr.: O ponatij „faktografičeskoi informacii”. Voprosy informacionnoj teorii i praktiki, 28. köt. 1. sz. Moszkva, VINITI, 1976. p. 7–16.
7. KRISTAL'NYJ, B. V. i dr.: O proekte tipovoj avtomatizirovannoj faktografičeskoi IPS. Voprosy informacionnoj teorii i praktiki, 28. köt. 1. sz. Moszkva, VINITI, 1977. p. 80–83.
8. Lingvističeskoe obespečenie faktografičeskogo informacionnogo poiska. Voprosy informacionnoj teorii i praktiki, 42. sz. Moszkva, VINITI, 1979. 116 p.
9. NAD', P. Analiz nekotoryh obših voprosov sozdanija avtomatizirovannyh faktografičeskih informacionno-poiskovyh sistem (AFIPS) i obobšenie opyta proektirovanija meždunarodnyh AFIPS v ramkah MSNTI. (Analitičeskij obzor). Moszkva, MCNTI, 1980. 219 p.
10. NAD', P.: Ispol'zovanie vnesnih po odnoseniju k MSNTI masinocitaemyh faktografičeskih baz dannyh = Problemy MSNTI/MCNTI, 1981. 3. sz. p. 26–41.
11. NAD', P.: Kommunikativnyj format predstavlenija dannyh dia vycislitel'nyh sistem. (Informacija o proekte Standarta ISO – specifikacija opisatel'nogo fajla dia obmena informacii) = Problemy MSNTI/MCNTI, 1982. 3. sz. p. 68–82.
12. Problemy avtomatizirovannoi obrabotki naučno-tehničeskoi informacii. NTO SSSR. Moszkva, 1982. 243 p.
13. TEREŠENKO, S. S.: Proektirovanie sistem naučno-tehničeskoi informacii: Analitičeskij obzor. Itogi nauki i tehniki. „Informatika” sor. 4. köt. Moszkva, VINITI, 1980. 263 p.
14. Tezisy dokladov 1-go Vsesoūznogo naučno-tehničeskogo simpoziuma „Dialogovye i faktografičeskie sistemy informacionnogo obsluživania”. Moszkva, NTO SSSR, 1979. 254 p.
15. Tezisy dokladov 2-go Vsesoūznogo naučno-tehničeskogo simpoziuma „Dialogovye i faktografičeskie sistemy informacionnogo obsluživania”. Moszkva, NTO SSSR, 1981. 254 p.

KOVÁCS Árpád—NAGY Péter: A faktografikus információkereső rendszerek néhány elméleti kérdése

A cikk a faktografikus információk, ill. információkereső rendszerek elméletével foglalkozva ezek terminológiai, osztályozási, információkereső nyelvi és működési (információkeresési) kérdéseit tárgyalja.

* * *

KOVÁCS, Á.—NAGY, P.: Theoretical problems of numeric information retrieval systems

The article deals with some theoretical problems of numeric information and information retrieval systems, discussing problems concerning their terminology, classification, indexing languages and operation.

КОВАЧ, А. — НАДЬ, П.: Несколько теоретических вопросов о фактографических информационно-поисковых системах

Статья занимается теоретическими вопросами фактографических информаций и информационно-поисковых систем, обсуждает вопросы, связанные с терминологией, классификацией, информационно-поисковыми языками и функционированием.

* * *

KOVÁCS, Á.—NAGY, P.: Zu einigen theoretischen Fragen der faktografischen Informationsrecherchen-Systeme

Die Arbeit behandelt — im Laufe der theoretischen Analyse von faktografischen Informationen bzw. Informationsrecherchen-Systemen — die Fragen der Terminologie, der Klassifizierung, sowie der Programmiersprachen und der Funktion von Informationsrecherchen.

Új szabványok

1983 végén megjelent az MSZ 3410 *Országnevek Kódjai* c. szabvány 2. kiadása. E szabványt kell alkalmazni a belföldi és nemzetközi tudományos és műszaki információközlésben, illetve cserében, ha az országokat és az egyes más földrajzi egységeket kódokkal jelölik meg.

A szabvány tartalmazza az országok nevének két-, illetve háromjegyű latin betűs, és két-, illetve háromjegyű cirill-betűs kódjait, továbbá számkódjait.

E 2. kiadás tartalmilag abban különbözik a szabvány 1. kiadásától, hogy az ENSZ Statisztikai Hivatala által kidolgozott és használt számkódokat is tartalmazza.

A szabvány használatának megkönnyítése céljából a függelék önálló táblázatos formában a kódok betűrendjében, illetve számsorrendjében is közli a kétjegyű latin- és cirillbetűs, a háromjegyű latin- és cirillbetűs, valamint a számkódokat a kódolt ország nevének feltüntetésével.

1983 végén megjelent az MSZ 193/1 *Mágnesszalagos bibliográfiai adatszere formátuma. A rekordok szerkezete* c. szabvány. A szabványt az azonos című és tárgyú ISO 2709-1981 nemzetközi szabvány, valamint a hasonló tárgyú, rövidesen KGST szabványként is megjelenő normatív-műszaki dokumentum alapján dolgozták ki.

A szabvány meghatározza a mágnesszalagos adatszere szánt bibliográfiai információkat tartalmazó rekordok szerkezetét. Nem határozza meg a rekordok hosszát és tartalmát, valamint a hívójelek, indikátorok és almezőazonosítók értékeit.

A szabvány szerinti rekordszerkezet csak a gépi adatfeldolgozást alkalmazó rendszerek közötti információcserét szabályozza (az adatszere-, más néven kommunikációs formátum rekordszerkezete), és nem terjed ki az egyes rendszerekben alkalmazott belső feldolgozó rekordszerkezetre. Ismerteti a bibliográfiai rekord általános és részletes szerkezetét (rekord-fej, a mutató és adatmezői, rekordhatár jel, illetve adatmezők, almezők, indikátor használata stb.)

*

*