

bőlésére újabban kifinomultabb stratégiákat alkalmaznak (automatikus indexelés tezaurusszal, szakértő rendszerek alkalmazása a dokumentumok elemzésére, indexelésére és a visszakeresésre).

A javasolt programcsomag egy nagyobb projekt része. Célja az elmúlt tíz évben a dokumentumok szabad indexelésével kialakított PRODOC adatbázis átszervezése, majd egy későbbi szakaszban automatikus újraindexelése. A PRODOC adatbázis jellemzői: több mint 350 000 rekordot tartalmaz; tematikailag átfogja a természettudományok, az alkalmazott tudományok, a technika és a közgazdaság egész területét; az indexelés természetes nyelven történik, az indexelő által szabadon létrehozott szintagmákkal; az alkalmazott hardver és a két lemezegységes INDEPENDENT miniszámítógép 1–15 Mb-át belső memóriával.

Az adatbázis újraszervezése két feladat megoldását jelenti: a rekordok szövegében található hibák kijavítását, mivel az adatbevitelnél nagyon sok hiba került be; a rekordok terjedelmének csökkentését a tárolókapacitás jobb kihasználása céljából.

A megoldás: azokból a szavakból (szón értve minden karaktersort, amelyet szóközök határolnak), amelyek már szerepelnek a rendszerben, szótárt állítanak össze, s a szótár szavait numerikus kóddal kódolják. A rekordokban a deskriptorokat ezekkel a kódokkal helyettesítik, s ezáltal lényegesen tömörítik őket.

A szótárt ugyanakkor fel lehet használni egy inverz fájl elkészítésére, amelyben a szavakhoz rendelve előfordulna minden dokumentum, amelyben az illető szó szerepel.

A szótár elkészítésénél a legnagyobb probléma a különböző nyelvtani alakok felismerése, mivel ezeket nem lehet külön-külön mind beírni, hiszen ez nagyon felduzzasztaná a szótárt, nagyon sok memóriát igényelne, megnehezítené a keresést.

Ezt a problémát oldja meg a "nyelvi analízátor", amely a főnevek és mellénevek elemzésével biztosítja a lexikális és morfológiai szempontból a helyességüket. Megvizsgálja, hogy azok a szavak, amelyek a kulcsszót alkotják, a szótárhoz tartoznak-e, és a kulcsszón belül ellenőrzi a szavak hajlítását és a szintaxist. Az analízátor alkalmazása nem korlátozza az indexelést, de a lexikális része szótárt épít fel, és benne keresi a szavakat.

A nyelvi analízátornak két alkotórésze van: a lexikális és a nyelvtani analízátor. Ez utóbbinak szintén két része van: a morfológiai és a szintaktikai analízátor.

A szintaktikai analízátor

Ellenőrzi a szavak kapcsolódását, és egy fadiagramot hoz létre, amely az alárendeltségi viszonyokat tükrözi. A szavakat szópárokká alakítja (szótő + toldalék), s a gráfot úgy építi fel, hogy a szópárok alkotják a szövegpontokat, az élek pedig az alárendeltségi viszonyokat mutatják. A szintaktikai elemzést nem mindig lehet teljesen elvégezni, mert a szavak alakja nem mindig határozza meg egyértelműen a köztük levő kapcsolatot.

A lexikális analízátor

A szavak tövét azonosítja úgy, hogy összehasonlítja egy különleges szótárral, amelyben a szavakat alkotó betűk fadiagram formájában vannak elrendezve, s a szótövek végének megfelelő szögpontokban információk találhatóak a szótőre vonatkozólag. Az analízátor összeállítja a lehetséges szótövek listáját. Természetesen előfordulhat, hogy ilyen módon nem lehet a helyes szótövet meghatározni. Ilyen esetekben csak a szemantikai elemzés segíthetne.

Morfológiai analízátor

A lexikális elemzés után a szó a morfológiai analízátorba kerül, amely a szó végződését (a szótőhöz csatlakozó betűket) elemzi. Az elemzés úgy történik, hogy a végződés végighalad egy hálódigramon, amelynek szögpontjai a toldalékokban előforduló betűk. Ha a végződés (a szótő morfológiai sajátosságaihoz viszonyítva) szabályos toldalék, az út egy végponthoz vezet, amelyben a toldalékhoz hozzárendelődik a megfelelő kód.

Az elemzés végül egy kételemű numerikus kódot eredményez, amelynek az elemei a szótő és a toldalék kódja. Ez a kód helyettesíti a rekordban a kulcsszót. Az eljárás előnyei: a rekordok legalább 50%-os tömörítése; a változó hosszúságú alfanumerikus mező átalakítása állandó hosszúságú mezővé, s ezáltal a feldolgozás egyszerűsítése; a szavak alakjától független visszakeresés lehetősége.

Az analízátort alkalmassá lehet majd tenni más beszédrészek elemzésére is, s akkor felhasználható lesz teljes szövegek feldolgozására.

/PETCU, B.: Analizor lexical și gramatical pentru substantive și adjective. = Probleme de Informare și documentare, 24. kötet. 1. sz. 1990. p. 3–6./

Nyilvános adatbankok

Romániában a helyzet nagyon rossz. Nem beszélhetünk egyetlen működő adatbankról sem, s a kilátások a közeljövőt illetően nem biztatók. Az ok: az illetékesek nem értik meg az adatbankok szerepét és fon-

tosságát, s ezért nem biztosítják a létesítésükhöz és működtetésükhöz szükséges pénzügyi, műszaki és emberi erőforrásokat. A lemaradás még a többi ex-kommunista országhoz képest is nagyon nagy.

Az információgyűjtemények akut hiánya mellett hiányoznak a kommunikációs lehetőségek is (számítógép-hálózat, Teletext/Videotex típusú nyilvános szolgáltatások, a külföldi hálózatokhoz csatlakozó kapuszolgálat, a gateway), amelyek megkönnyítenék a hozzáférést a létesítendő hazai és a létező külföldi adatbankokhoz.

Romániában az átalakulás időszakában azért lenne nagyon fontos a nyilvános adatbankok létrehozása és a már potenciálisan létezők hozzáférhetővé tétele, mert meggyorsítaná az áttérést a piacgazdálkodásra, segítené a szakemberket, hogy az adatbankokból szerzett információk felhasználásával gyorsabban tudják helyettesíteni a régi centralizált vezetési mechanizmust egy új, rugalmas mechanizmussal. Ezt a megállapítást támasztja alá a Közös Piac elhatározása, hogy Közép- és Kelet-Európában, valamint a Szovjetunióban felleltározzák az információs erőforrásokat, s ebből kiindulva dolgozzák ki a segítségadás módozatait.

A felgyorsítás útjai

Az adatbankok létesítését és használatát a következő műveletekkel lehetne felgyorsítani:

- ▶ az országban létező információs források, a külföldi nagy központoktól kapott mágnesszalagok

(AGRIS, INIS, CAS, MEDLARS, INPADOC) maximális hasznosítása. Ennek legnagyobb akadálya a nagy kapacitású számítógépek hiánya;

- ▶ a már létező *automatizált adatgyűjtemények* hasznosítása. Azonnal lehetne hasznosítani a különböző területeken létrehozott dokumentációs adatgyűjteményeket (könnyűipar, kohászat, energetika stb.);
- ▶ a *pénzügyi alapok* és a *műszaki feltételek* megteremtése, hogy a nyugati adatbankok információs hozzáférhetőkké (beszerezhetőkké) váljanak;
- ▶ *állami szubvenció* olyan adatbankok létesítéséhez, amelyeket nem lehet külföldről átvinni (törvényhozás, igazságszolgáltatás, szállítás, kereskedelem, erőforrások stb.);
- ▶ a *törvényes keretek* megteremtése az információs intézmények működésének szabályozására.

A szerző a továbbiakban részletesen ismerteti az adatbankok elvi alapjait, használatuk módját, az adatbank-katalógusokat, a kritériumokat, amelyek alapján ki lehet választani az igényeknek legmegfelelőbb adatbankot vagy bankokat.

/TEODORESCU, B.: *Banci de date publice. = Probleme de informare și documentare*, 24. köt. 3. sz. 1990. p. 91–102./

A tudományos-műszaki tájékoztatás és a nemzeti bibliográfia kapcsolata

A felhasználók minél teljesebb, pontosabb és gyorsabb tájékoztatása a mai körülmények között, amikor minden területen hihetetlen mértékben elszaporodtak a legkülönbözőbb kategóriájú folyóiratok és kiadványok, szükségessé teszi az információfeldolgozó és -forgalmazó intézmények közötti együttműködés stratégiájának új alapokra helyezését.

Romániában a tudományos-műszaki tájékoztatási hálózatban a különböző szakosított egységek (dokumentációs intézetek, irodák, könyvtárak stb.) megszervezése és működése az egyszeri feldolgozás-több csatornán való terjesztés elvén alapul, azaz arra törekednek, hogy a primer dokumentumokat csak egyszer dolgozzák fel, s az információkat azután minél több csatornán juttassák el a fogyasztókhoz.

Az országos hatáskörrel felruházott intézmények közül a legfontosabbak az *Institutul National de Informare și Documentare* (INID = Országos Információs és Dokumentációs Intézet) és a *Biblioteca Națională* (BN = Nemzeti Könyvtár). Az előbbi feladata a tudományos-műszaki tájékoztatás, az utóbbié többek között – s a jelen dolgozat szempontjából ez a legfontosabb – a hazai nyomtatványok bibliográfiai számbavétele. A két intézmény közötti együttműködésnek régi hagyományai vannak, több síkon is. Szoros az együttműködés egyebek mellett az információ és dokumentáció általános módszertanának a kidol-

gozásában, eszközeinek kimunkálásában (szabványok, előírások, útmutatók, az ETO román nyelvű kiadása stb.). A számítógépes rendszerek létrehozása a két intézményben új lehetőségeket teremtett az együttműködésre, és jelentősen módosította a céljaikat. Jelenleg a szakemberek figyelme a következő feladatokra összpontosul: a feldolgozásban a párhuzamosságok kiküszöbölésének szigorúbb ellenőrzése, a dokumentumok leírásának (adatstruktúrájának) egységesítése, a két intézményben létrehozott információállomány kölcsönös átvétele.

Ezeknek a céloknak a megvalósítása azért lehetséges, mert bár a két intézmény informatikai rendszerének más-más a feladata, sajátosak a funkciói, mégis vannak közös vonásaik is. Elsősorban az, hogy a hardver azonos, román gyártmányú, s így egymással kompatibilis. Az informatikai együttműködés első szakaszában az volt a cél, hogy a hazai tudományos-műszaki irodalmi termés bibliográfiáját, amelyet a BN állít elő, a közösen megállapított szerkezetben mágnesszalagon átadják az INID-nek.

A BN több sorozatban adja ki a nemzeti bibliográfiát. Ebben a vonatkozásban két sorozata érdekes, a havonta kétszer megjelenő *Cărți, albume, hărți* (Könyvek, albumok, térképek) és az évente megjelenő *Teze de doctorat* (Doktori disszertációk). 1983-tól mindkettőt Felix számítógépen készítik, eredeti programok segítségével.