

## A tudományos-műszaki tájékoztatás Romániában

Romániában a tudományos-műszaki tájékoztatás vezető intézménye a Műszaki Dokumentációs Intézet (Institutul de Documentare Tehnica = IDT), amely 1949. január 1-jén alakult meg, s keretében épült ki az országos műszaki könyvtár. Néhány éven keresztül ez az egyetlen intézmény gondoskodott – a természettudományok és a technika minden ágát felölelő referáló folyóiratai révén – a román szakemberek tájékoztatásáról. Az ötvenes évek közepén azután sorra alakultak az ágazati dokumentációs központok (vegyipari, gépipari, közlekedési stb.), továbbá az orvosi, pedagógiai dokumentációs központ, majd a hatvanas években a Tudományos Akadémia mellett a Tudományos Dokumentációs Központ (Centrul de Documentare Stiintifica). E fejlemények következtében az IDT feladatköre, neve is megváltozott: Központi Műszaki Dokumentációs Intézet (Institutul Central de Documentare Tehnica = ICDT) lett belőle. Miközben tájékoztató funkciója a műszaki területre és határterületeire szűkült le, új funkcióként rábízták az országban folyó információs tevékenység összefogását, szervezését és irányítását. Ezt a szerepét azonban sohasem tudta igazán betölteni, részben szervezési hiányosságok, részben az ágazati központok presztízsfeltétele és a hatásköri "villongások" miatt. Pedig az intézet akkori igazgatója nagyon mutatós tervet dolgozott ki egy háromlépcsős rendszer kiépítésére.

A rendszer legfelső fokán az időközben megint nevet változtatott Országos Információs és Dokumentációs Intézet (Institutul National de Informare și Documentare = INID) állt volna, a középső lépcsőn az ágazati központok, a legalsón pedig a vállalati, intézményi irodák foglaltak volna helyet. Ez a rendszer sajnos nem valósulhatott meg, mert 1974-ben egy pénzügyminiszeri rendelet szinte teljesen felszámolta az országban folyó dokumentációs tevékenységet. Az ágazati központokat irodákká (Centruból Oficiuvá) fokozták le, költségvetésüket és személyzetüket drasztikusan csökkentették, a devizakereteket jóformán zéróra redukálták, sőt: a rendelet már-már abszurdnak nevezhető módon szabta meg a könyvtárak által előfizethető folyóiratok

számát is. (Az üzemi könyvtárak például csak egy hazai folyóiraatra fizethettek elő.)

A rendelet természetesen az INID-et is alaposan "megtépázta". A munkatársak számának csökkenése, a pénzhány, a berendezések fizikai és erkölcsi kopása természetes módon vezetett az információs-dokumentációs tevékenység soradásához. Emellett a kutatásban, termelésben dolgozó szakemberek szakmai érdeklődése is a mélypontra süllyedt. A tudást, az elért eredményeket szinte egyáltalán nem értékelő javadalmazási rendszer, az akkoriban módosított szabadalmi törvény, amely a feltalálókat mondhatni kisémmizte, alaposan csökkentette az érdeklődést a szakmai újdonságok iránt.

Természetesen a mostoha körülmények ellenére mindenütt akadtak lelkes emberek, akik igyekeztek nemcsak fenntartani, hanem tovább is fejleszteni, korszerűsíteni a dokumentációs tevékenységet. Szép – de távolról sem kielégítő – eredmények születtek például az tájékoztatás számítógépesítése terén (az INID számítógépes szelektív információterjesztési rendszere, a Nemzeti Könyvtárban a nemzeti bibliográfia számítógépesítése stb.).

A nyolcvankilences decemberi fordulat óta eltelt idő még túl rövid, az ország gazdasági – és politikai – helyzete még nagyon bizonytalan ahhoz, hogy gyökeres változások mehessenek végbe a tudományos-műszaki tájékoztatás terén. Némi javulás azonban már mutatkozik. Bővült például az INID szolgáltatásainak köre, s a minősége is javult valamelyest. A további fejlődés minden bizonnyal a nemzetgazdaság fejlődésétől függ.

A tájékoztatásügy román folyóirata az immár huszonhatodik évébe lépő Probleme de informare și documentare (Az Információ és Dokumentáció Kérdései). A két nyelven – románul és angolul – megjelenő folyóirat valamikor havi folyóirat volt, az utóbbi másfél évtizedben azonban csak negyedévenként jelenik meg. Tartalma: két-három tanulmány és tallózás a külföldi szakirodalomban.

Az alábbiakban az elmúlt évben megjelent tanulmányok közül ismertetünk hármát. Ezek bizonyos mértékben tükrözik a román szakemberek érdeklődési körét is.

### Programcsomag főnevek és melléknevek lexikális és morfológiai elemzésére

Az információtároló és -kereső rendszerekben fontos szerepük van azoknak a keresési stratégiáknak, amelyek kötetlen szókészleten alapulnak, főleg ott, ahol nagyon specifikus tényinformációk és adatok keresésére van szükség, mert velük le lehet küzdeni az eleve korlátozott szókészletű ellenőrzött

szótárak vagy tezauruszok "felbontóképességének" hiányosságait. Az ilyen stratégiák alkalmazását az is indokolja, hogy az automatikus indexelés kötetlen szókészlettel sokkal olcsóbb, mint az ellenőrzött szótárral vagy tezaurusszal végzetté. A szemantikai pontatlanság miatt fellépő veszteségek kiküszö-

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 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 szópárokra 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. köt. 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.