

A GÉPI NYELV SZEREPE A TUDOMÁNYOS ÉS MŰSZAKI TÁJÉKOZTATÁSBAN

Virághalmy Ferenc

A tudományos publikációk óriási számszerű növekedése miatt az irodalom-áradattal eredményesen megbirkózni "kisipari" módszerekkel ma már lehetetlen. Elkerülhetetlen szükségesség az információkat tartalmazó dokumentumok keresésének, szelektálásának gépesítése, automatizálása.

Ezt a feladatot az elektronikus logikai gépek vannak hivatva megoldani.

Ahhoz azonban, hogy a gépet ilyen célra felhasználhassuk, szükség van az anyag megfelelő előkészítésére. Az összes információknak tájékoztató központba kell befutniuk, amely el van látva olyan elektronikus logikai géppel, amely memóriájának befogadóképessége jelentékenyen nagyobb, mint a jelenlegi hasonló gépeké, melyekben még nincs lehetőség minden tudományos és műszaki területet felölelő "szótár" tárolására. Ezért a közeljövőben a tudomány és technika minden egyes területén külön ön információ-szolgáltató gép felállítása látszik indokoltnak.

A központba beérkező dokumentum előkészítése három lényeges munkafázisból áll:

1. Mikrofilmezés

Erre feltétlenül szükség van a férőhellyel és munkaerővel való takarékoskodás miatt. /1 millió könyv tárolásához 2500 - 3000 m² helyre van szükség. Ugyanezt az anyagot mikro-formában 30-60 m²-en el lehet helyezni! A mikro-anyaggal való foglalkozás a személyzetnek 1/10 részére csökkentését teszi lehetővé - a hagyományos kiadványfajtákkal szemben! /

A mikro-kiadványok megjelenése révén lehetőség lesz arra is, hogy a szakember szakterületével kapcsolatos teljes információs anyag birtokában legyen - "magánkönyvtár" formájában.

2. Referálás

A mikrofilm mellett a kiadvány referátumát is tárolják. Az előb-

bi kiadására csak akkor kerül sor, ha a referátum átolvasása után erre igény van.

A referálás jelenleg intellektuális és bizonyos tekintetben szubjektív folyamat. Ez is gépesíthető - ezáltal a szubjektivizmus kizárható. A gépbe kódolt formában /perforált kártya vagy szalag - mágnesszalag/ viszik be a referálandó cikket. A gép kiválasztja a legfontosabb kulcsmondatokat, ezeket a cikkben lévő előfordulási sorrendbe szedi - ezek alkotják a referátumot.

Vizsgáljuk meg a kulcs-mondatok kiválasztási módszerét. A kiindulási alap: a kulcs-szavak. /Meghatározásuk statisztikusan történik./ A gép a szöveget először szavakra bontja, kizárja a nem-informatív szavakat, a megmaradókat alfabétikus sorrendbe szedi. Megállapítja a szó-ismétlődés gyakorisági sorrendjét, majd megállapítja a kulcs-mondatok fontossági sorrendjét /a bennük előforduló kulcs-szavak száma alapján/. Tekintettel van a szavak mondatbeli "távolságára", elhelyezkedésére is. Minél gyakrabban fordul elő egy mondatban valamely kulcs-szó, annál fontosabb az illető mondat.

A legfontosabb kulcs-mondatokat mágnesszalagra jegyzi, majd papírra írja ki a referátumot.

3. Elemzés

Ennek eredménye a dokumentum alapos tartalmi analízise útján nyert index. Az index a dokumentum megfelelő terminusai a l történi jelölése. Ez a terminus szám, betű vagy betű-szám kombináció lehet, melynek alapján a memória-berendezésből értesülést szerzünk a dokumentumról.

Az indexelést gyorsműködésű elektronikus gép is végezheti, akár kulcs-szó /descriptor/, akár leíró indexről van szó.

Az előbbi esetben a kulcs-szavakat a dokumentum tartalmát meghatározó koordinátáknak fogjuk fel. Ezek megállapítása a gépi referátum-készítésnél említettéhez hasonlóan történik. Meghatározza a szavak elhelyezkedési rendjét a dokumentumban, kizárja a nem-informatív szavakat /kötszókat, előljárókat stb. és az olyan - esetünkben semmitmondó - kifejezéseket, mint "beszámoló", "felsorolás", "jelentés" stb./. A megmaradt informatív szavakat gyakorisági sorrendbe szedi. Az első 20-30 szót használják fel a több ezer dokumentum hasonló analízise útján nyert koordinációs információkeresési szótár összeállításakor. A szótár állományába tartozó kulcs-szavakat viszik be a visszakereső gép memóriájába.

A leíró indexelés az információs anyag elemzésének másik útja, amikor a kulcs-szavak kontextusait állapítják meg. A leíró index néhány kulcs-szóból áll - benne az a t t r i b u t í v viszony kifejezésre jut. Az ilyen indexelés végezhető a dokumentum teljes tartalmának, referátumának vagy címének elemzése útján. Ez a folyamat is automatizálható. Az elektronikus

logikai gép olvasó-berendezésébe helyezik a dokumentumot. A gép regisztrálja a szavak ismétlődési gyakoriságát. Minden legalább hat-szor előforduló informatív szó kulcs-szónak számít. Ezeket a szavakat a meglévő memóriatartalommal összehasonlítva a gép meghatározza az indexet, és azzal megjelöli a dokumentumot.

Az információ megtelelése céljából a fenti indexek valamelyike mellett a dokumentumnak "m u n k a - s z á m a" is van, melynek alapján maga a dokumentum megtalálható.

Az előkészítés után kerül sor a t á r o l á s r a , amely két munkafázist tételez fel:

1. Kódolás

A kód szerkezete az indexelési módszertől függ, és messzemenően figyelembe veszi az illető nyelv statisztikai összetételét. A kódolás nyelvészeti is érdekes esete a s z e m a n t i k a i k ó d , melynek segítségével meghatározott nemű terminusok és összetételeik nagy mennyiségéből kereshető vissza a kívánt információ. Alkalmazásához gondosan kidolgozott k ó d - t e r m i n u s s z ó t á r r a van szükség. Jellegetessége - és előnye - hogy tekintettel van a kódolandó mondat kulcs-szavai közti é r t e l m i v i s z o n y r a . A kódolt szócsoport "alcsoportokra" bomlik, amelyek kód-terminusból és néhány s z i m b ó l u m b ó l állnak. A szimbólumok feladata, hogy a terminus mondatbeli szemantikai szerepét jellemezzék.

A természetes nyelvet olyan formalizált k ó d - n y e l v é alakítják át, amely szemantikai koefficienseket, képzőket, jeleket stb. tartalmaz.

Minden egyes jelet bináris számokká alakítanak, amelyeket meghatározott lyukrendszerben helyezik el a perforált szalagon. Ebben a formában viszik be az információt a gépbe.

2. Bevitel a memória-berendezésbe

A már említett perforált kártyák, -szalagok és mágnes-szalagok szolgálnak e célra. A perforált szalagokra tetszőleges hosszúságú szöveg rávihető.

A következő lépés a k e r e s é s , amely két részre tagolódik:

1. A kérdés elemzése;
2. Gépi keresés.

A gépi keresés eredményeként a gép a dokumentum referátumát vagy teljes szövegét adja ki. Az előbbi esetben írógéppel irt formában vagy perforált szalag alakjában, az utóbbiban pedig pozitív fotokópia képében fogunk az információhoz jutni.

Az információ kiszolgáltatása három munkafázist foglal magába:

1. A kiválasztott dokumentumok elemzése;
2. Az információ szintézise;
3. Az információ kiadása.

A továbbiakban az információ-visszakereső elektronikus logikai gépek memóriájában történő adattárolás néhány dokumentációs és nyelvészeti kérdésével kívánunk foglalkozni.

Első kérdés: Mit tartalmazzon a gép memóriája?

Tartalmaznia kell a kutató /igénylő/ részéről várható valamennyi kérdésre a választ. /Pl. milyen irodalom áll rendelkezésre a félvezetőkről? Mely anyagok alkalmasak szigetelésre? stb./.

Második kérdés: Milyen formában történik az adatok rögzítése a gép memóriájában?

G é p i k ó d /egy adott rendszerre - abc-re - épül fel; segítségével ezen abc betűiből álló szöveg kódolható, akár bináris, akár egyéb elv alapján és bevihető a gép memóriájába/ vagy g é p i n y e l v formájában.

Ez utóbbinak jellegzetességeit, életrevalóságát fogjuk megvizsgálni.

A gépi nyelv, mint neve is mutatja, m e s t e r s é g e s és a b s z t r a k t nyelv, amely tudományos információk rögzítésére szolgál. Gépi alkalmazásra születik. Létrehozása feltétlenül indokolt, mivel a természetes nyelvek bizonyos tulajdonságaik /formalizálatlanság, kivételek és rendhagyó alakok, "lexikai szubjektivitás" stb./ miatt nem alkalmasak a gép számára.

A gépi nyelvek jelrendszeren alapulnak. A jeleket rendszerint a matematikai logikából kölcsönzik. Ez az absztrakt nyelv úgy viszonylik a természetes nyelvekhez, mint a sakk-parti lefolyását rögzítő egyezményes jelek "nyelve" a partit é l ő b e s z é d b e n előadó egyén által használt - és anyanyelvű felfogó számára érthető - nyelvhez.

A gépi nyelv nullák és egységek formájában kódolva nyer alkalmazást a gép memóriájában. Egy adott gépi nyelv-konstrukció csak e g y meghatározott tudományág vagy ismeretkör számára alkalmazható.

A gépi nyelv összeállításakor a következő főbb követelményeket kell szem előtt tartani:

1. Minden lejegyzésnek csak e g y és kizárólag csak egy fogalmat vagy fogalomcsoporthoz szabad jelölnie, illetve minden egyes fogalom vagy fogalomkör csak egy jel formájában jelenhet meg;

2. A gépi nyelv jól f o r m a l i z á l h a t ó legyen;

Milyen funkciók formalizálásáról van szó?

- a/ az információ kiválasztása;
- b/ a különböző megformálásu tényrögzítések azonosítása;
- c/ a logikai következtetések levonása.

3. A gépi nyelv mindenegybes mondata "szabvány-formában" legyen kifejezhető.

Az ilyen nyelvben alkalmazott jelek - mint már korábban említettük - a matematikai logikából is származnak /pl. olyan általános fogalmakat kifejező jelek mint az "és", "vagy", "vagy" stb./. A jelekhez a predikativ viszonyok és a funkciók jelölése céljából állandó és változó terminusok jelölései is hozzátartoznak. A gépi nyelv minden mondata logikai alanyból /S/ és logikai állítmányból /P/ áll:

S est P.

E nyelv mondatait az előbbi viszonyoknak megfelelően - állítás formájában fogalmazzák meg. Pl.: "A víz fagyáspontja 0°C" - ezt az állítást a gépi nyelv szerkezetének megfelelően így fogalmazzák meg: "állítás /jelölése: // /, hogy a víz /jelölése: /, hőmérséklete /jelölése: t / 0°C-nál kevesebb, egyenértékű azon állítással, hogy a víz szilárd állapotban van /jelölése: Szv/".

A gépi mondat képe: /t / 0°C Szv / /.

A gép memóriája hierarchikus elv szerint épül fel, amit így szemléltethetünk:

ország - országban - városok - városokban - utcák - utcákban - házak - házakban - szobák - szobákban - szekrények - szekrényekben - fiókok - fiókokban - dobozok.

Minden tárolt információnak a megfelelő ország - város - utca - ház stb. sorszámából álló "címe" van. A gép memóriájába bevihető táblázat vagy olyan képlet, amelyből behelyettesítéssel megkapható a táblázat. Ez utóbbi esetben kisebb a memória befogadóképessége, viszont szükséges, hogy a gép behelyettesítési képességgel rendelkezzen.

A memóriában az információkat úgy helyezik el, hogy az biztosítani tudja a meghatározott program szerinti keresés eredményességét.

Ahhoz, hogy az ember és a gép "információcseréje" megkezdődhessen, még egy feladatot kell megoldani. Ki kell dolgozni az absztrakt nyelvről a természetes nyelvre /és fordítva/ történő fordítás formális szabályait.

Elsődleges szempont, hogy a fordítás a természetes nyelvről a gépi nyelvre a u t o m a t i k u s legyen, hiszen a természetes

nyelven kifejezett tények gépi nyelven való a u t o m a t i k u s
lejegyzése /a gépben/ csak így valósítható meg.

A gépi nyelvről történő automatikus fordítás lehetővé teszi az
információk reprodukálását a k i v á n t természetes nyelven.

A gépi nyelv értelemszerűen felhasználható k ö z v e t i t ő
n y e l v k é n t egyik természetes nyelvről a másikra történő for-
ditáshoz, amiben a tudományos és műszaki szövegek gépi fordításának
kulcsát kell látnunk.

Az itt vázolt információ-visszakereső rendszer egyike azon rend-
szereknek, amelyekkel a jövőben a tudományos és műszaki tájékoztatás
megvalósítása történhet.

.o.

BIBLIOGRÁFIA

- INFORMATION retrieval management. Ed. L.H. Hattery, E.M.McCormick.
Detroit. American Data Processing, Inc. 1962. 151 p.
/Data processing library series./ Bibliogr.: 136-147.p.
- ISZAEV,É.A.: Izobrazenie na magnitnoj lente. Moszkva. Izdat. "Zna-
nie". 1964. 16 p. /Tehnika 24./
- KARPILOVSZKIJ,E.B.: "Elektronnij mozg". Moszkva. Izdat. "Znanie".
1962. 40 p. /Tehnika 8./ Bibliogr.: 39.p.
- REJNBERG,M.G.: Buduscsee vücsiszlitel'nüh masin. Moszkva. Izdat.
"Znanie". 1964. 48 p. /Tehnika 6./
- SAROL',I.I.: Kiberneticeszkie masini vücssego ranga. Moszkva. Izdat.
"Znanie". 1963. 38 p. /Tehnika 24./

o o o
o

ВИРАГХАЛМИ Ф.: Роль машинного языка в научно-технической информации

Рост литературы требует механизации и автоматизации поиска информации. В статье говорится о перспективах этого процесса. Все информационные материалы будут поступать в информационный центр, оснащенный электронно-логической машиной. Документ поступивший в центр, после микрофильмования, реферирования, анализа, кодирования годится для хранения в памяти машины. Здесь информации хранятся в форме машинного языка. Структура машинного языка и характеристика машинного предложения позволяют осуществление "обмена информацией" между человеком и машиной. Машинный язык можно рассматривать в качестве языка-посредника при переводе с одного реального языка на другой.



VIRÁGHALMY, F.: Rolle der maschinellen Sprache in der technisch-wissenschaftlichen Information

Das Anschwellen der Literatur erforderte die Mechanisierung und Automatisierung des Nachweises der Informationen. Der Artikel behandelt die Perspektiven dieses Prozesses. Alle Informationen laufen im Informationszentrum ein, welches über eine elektronische logische Maschine verfügt. Das eingegangene Dokument eignet sich - nach Mikroverfilmung, Referieren, Analysieren, Kodieren - zur Speicherung in der Memorie der Maschine. Hier wird die Information in der Form der maschinellen Sprache gespeichert. Die Struktur der maschinellen Sprache und die Charakteristik des maschinellen Satzes machen den "Informationsaustausch" zwischen Mensch und Maschine möglich. Die maschinelle Sprache lässt sich als Zwischensprache bei der Übersetzung aus einer natürlichen Sprache in die andere anwenden.

.§.