

Az elektronika a szakirodalom feldolgozásában

ETO 651.838.8 : 001.814

Elektronika és irodalom — e két egymástól lát-
szólag oly távol eső fogalom — a gyakorlatban a leg-
szorosabb kapcsolatba került egymással. Egyrészt
azért, mert a technika, különösen pedig annak ki-
emelkedő fontosságú ága, a híradástechnika csak
akkor fejleszhető tovább eredményesen, ha minden
kutatói, fejlesztési, gyártási feladat megoldásakor
a már elért, legfrisebb eredményekből indulunk ki.
Másképp, mert a folyóiratcikkek, szabadalmak,
könyvek száma ma már olyan óriásira nőtt, hogy
szinte elárasztja a szakembereket. Komoly károkat
okozhat a népgazdaságnak, ha olyan munkára pa-
zaroljuk erőnket, aminek eredménye valamely
könyvtárban porosodó folyóiratban már megtalál-
ható. Nincs azonban olyan mérnök, aki megenged-
hetné magának, hogy saját maga nézze át mindazt
a folyóiratot, amely akárcsak legszűkebb munkaterü-
letével kapcsolatosan értékes anyagot tartalmazhat,
nem is beszélve a különböző nyelvi nehézségekről,
valamint arról, hogy egy-egy gyárhoz sőt kutató-
intézethez is csupán a megjelenő szakirodalmi anyag
kis töredéke jut el. Ebből a zsákutcából a kiutat a
dokumentáció az elektronika segítségével hívásával
találta meg.

Az elektronika, az elektronikus digitális számító
gépek, valamint a speciális dokumentációs „cél-
gépek” ma már számos formában segítik az iro-
dalom feldolgozását. Géppel lehet nagytömegű anya-
got tárgy szerint csoportosítani, adott kérdéssel
kapcsolatos irodalomlistát készíteni, megoldották
a dokumentumok gépi olvasását, referátumok (pon-
tosabban kivonatok) gépi előállítását, a gépi in-
dexkészítést, automatizáltak hagyományos könyvtá-
ri munkafolyamatokat, mint a katalogizálás, köl-
csönzés, rendelés lebonyolítás stb. Ez a cikk példa-
ként csupán azt kívánja röviden bemutatni, hogy
hogyan oldható meg elektronikus számítógéppel
egy vállalat (vagy akár iparág) mérnökeinek a beér-
kezett, legfrisebb folyóiratanyagról történő gyors
tájékoztatása, a gépi témafigyelés, valamint a gépi
index-készítés.

1. A gépi témafigyelés

Kis könyvtár esetében elképzelhető, hogy a lelki-
ismeretes könyvtáros jól ismeri néhány olvasójának
érdeklődési körét, és rendszeresen felhívja figyel-
müket azokra a legújabb cikkekre, amelyeket fel-
tétlenül el kell olvasniok. Amint azonban gazdagabb
könyvtári anyagról van szó, amint a szabadalmakat,
könyveket, prospektusokat stb. egyaránt be akarják
venni a figyelt anyag körébe, és amint a tájékoz-
tatásra váró személyek száma már tízes vagy száz-
as nagyságrendű, a feladat megoldása gépesítés nélkül

elképzeltetlen. A vállalatok részéről a gyors és
hatásos gépi figyelőszolgálat megoldására vonat-
kozó igény kielégítésére dolgozta ki az IBM az in-
formációk szelektív terjesztésének rendszerét (IST
rendszer).

A rendszer legfőbb tulajdonsága, mint neve is
mutatja, a szelektivitás. Célja, hogy számítógép
felhasználásával biztosítsa: az információ a vállalaton
vagy intézményen belül azokhoz és kizárólag
azokhoz a személyekhez jusson el, akik azt munka-
körük és érdeklődési területük figyelembevételével
a legnagyobb valószínűséggel tudják felhasználni
és értékesíteni. A szelektivitásból hangsúlyozottan
következik tehát az is, hogy az információnak nem
szabad eljutni olyan személyekhez, akik számára
érdektelen, mert amennyiben valakit szükségtelen
információk tömegével árasztanak el, az a további-
akban már a lényeges adatokat sem fogja felhasz-
nálni.

A gépi témafigyelés rendszerének kidolgozásakor
két alapfeltevésből indultak ki. Egyrészt abból,
hogy van pontos módszer az egyes olvasók egymás-
tól való megkülönböztetésére, van valami, ami egy-
értelműen azonosítja őket a tömegben: az ún. profil.
Másképp pedig abból, hogy egyszerű eszközökkel,
automatikusan és megfelelő pontossággal meghatá-
rozható, miről szól egy cikk és kiválaszthatók azok
a dokumentumok, melyek egy-egy profilnak meg-
felelnek.

A profil lényegében nem más, mint azoknak a sza-
vaknak a gyűjteménye, melyeket az illető szakember
érdeklődési területének részletes leírásakor felhasz-
nál. Meghatározható egyrészt az érdekelten folyta-
tott személyes megbeszélés alapján, vagy oly módon
s, hogy feltételezik: az embereket jellemzi az, amit
olvasnak, és amit olvasnak, jellemezhető az olvasott
anyagban előforduló fogalmak jegyzékével. Ha
ugyanis ez fennáll, akkor — amint erre a későbbiek-
ben még kitérünk — a profil változását, bővülését,
módosulását automatikusan értékelni és követni
tudja az irodalmat feldolgozó gép.

A folyóiratcikkek tárgyának meghatározása a
bennük előforduló ún. értékes szavak felhasználá-
sával történik. Nyilvánvalóan lényegtelenek a mon-
danivaló szempontjából a névelők, segédigék, név-
mások stb., de a szerző által felhasznált műszaki
kifejezéseket kiválogatva képet kaphatunk arról,
hogy a cikk mivel foglalkozik.

A fentiekből kiindulva a digitális számítógéppel
végzett, automatikus figyelőszolgálat megvalósí-
tása a következőképpen történik.

Mindenekelőtt meghatározzák a rendszeres tájé-
koztatásra váró érdekeltek profilját, és a profilban

szereplő műszaki kifejezéseket (tárgyszavakat) az érdekelt nevével és címével együtt mágnesszalagra vagy mágneses tárcsára rögzítik. A rendszeres heti tájékoztató munka során a tulajdonképpeni feladat az, hogy meg kell találni az egyezéseket a profilok és a beérkező dokumentumok között.

A dokumentumok feldolgozása ma még általában emberi erővel, manuálisan történik, mivel a már említett gépi olvasás és kivonatkészítés egyelőre túl drága lenne. A dokumentációs iroda munkatársai tehát munkalapra feljegyzik a cikk szerzőjét, a címét, a folyóirat megjelenési adatait, valamint a cikk mondanivalóját röviden összefoglaló referátumot. A munkalapról a szöveget Hollerith lyukkártyákra lyukasztják, mégpedig úgy, hogy az egyes adatsoportok (szerző, cím, referátum stb.) külön kártyára kerülnek. Ha a szöveg hosszú, egy adatsoporton belül több kártyát lehet lyukasztani: így pl. a referátum rendszerint 5-6 lyukkártyát vesz igénybe. Egyszerűsíti a munkát, ha a munkalapokat nem közönséges írógépen töltik ki, hanem olyan elektromos írógépen (pl. Flexowriter vagy Telex írógép), mely a szöveg írásával egyidejűleg lyukszalagot lyukaszt. Ezt a lyukszalagot lyukkártyákra lehet konvertálni, vagy közvetlenül is alkalmas számítógép bemeneteként. Technikai nehézségek következtében kevésbé terjedtek el, de ismeretesebbek olyan írógépek is, melyek a szöveg írásakor lyukkártyákat lyukasztanak. A cikkekről készített lyukkártyákat vagy lyukszalagot beolvassák a számítógépbe, mely az adatokat memóriájában tárolja, majd megkezdi a tárolt szöveg (a cím és a referátum) elemzését. Mindenekelőtt leméri az egyes szavak hosszúságát, vagyis megszámlálja, hogy hány jel van két szóköz között, és a szavak hosszúságukkal

súlyozza. Ezután a szavakat abc sorrendbe rendezi, természetesen figyelmen kívül hagyva, hogy ugyanaz a szó esetleg többször is szerepel. Így egy mondat szavai az eredeti és az ártírt formában (abc sorrendben rendezve) a következőképpen követik egymást:

I ADMIT THAT MATHEMATICAL SCIENCE IS A GOOD THING BUT EXCESSIVE DEVOTION TO IT IS A BAD THING

IA 5ADMIT 3BAD 3BUT 8DEVOTION 9EXCESSIVE 4GOOD 1I 2IS 2IT S MATHEMATICAL 7SCIENCE 4THAT 4THING 2TO (S annyit jelent, mint 12, mivel 10-nél nagyobb számokat be-tűvel jelölik.)

A cikkről ily módon készült sűrített szójegyzék szavait kell az egyes profilokat képviselő szójegyzékekkel összehasonlítani. A feladatot azonban megnehezíti az, hogy a szerzők a szebb stílus érdekében ugyan arra a fogalomra lehetőleg több szót is használnak, és ha mind a profilban, mind a cikke vizsgálatkor az összes rokon és azonos jelentésű szót külön-külön tüntetnek fel, az a figyelőszolgálat eredményességét, a szolgáltatott anyag áttekinthetőségét nagyban rontaná. Ezen úgy segítenek, hogy még a rendszeres gépi tájékoztató munka elindítása előtt olyan szótárt készítenek, mely meghatározza, hogy a szinonimák közül melyiket fogadja el a gép szabványosnak. Ezt a szótárt a gép memóriájában tárolják. A profilok leírásában már eleve csak szabványos kifejezéseket használnak fel. A dokumentumok feldolgozásakor azonban a sűrített szójegyzékben szereplő, nem szabványos szavakat a gépi szótár segítségével először szabványos megfelelőjükkel kell helyettesíteni.

Key Words-In-Context Index

LIST OF ABBREVIATED AND FULL TITLES OF TECHNICAL AND PROOF SERVICES, TRING COUNTRY, MACHINES CARDS TO SORT INFRARED CARDS INDEXING INFRARED GRAPHIC SCHEME BASED ON TIC INFORMATION.* USING PUBLISHING MODERN COMPANY PHARMACEUTICAL A PUNCHED CARD THE RELATION OF AN FROM JOURNAL ARTICLE TO ID SYSTEM OF CODING AND SYMPOSIUM ON THE ORGANIZATION OF AN WABASH CUTS WAY BILL.	ABBREVIATED AND FULL TITLES OF TECHNICAL AND PROOF SERVICES, TRING COUNTRY, MACHINES CARDS TO SORT INFRARED CARDS INDEXING INFRARED GRAPHIC SCHEME BASED ON TIC INFORMATION.* USING PUBLISHING MODERN COMPANY PHARMACEUTICAL A PUNCHED CARD THE RELATION OF AN FROM JOURNAL ARTICLE TO ID SYSTEM OF CODING AND SYMPOSIUM ON THE ORGANIZATION OF AN WABASH CUTS WAY BILL.	INSTS-57-LAF PERREW-57-NIS BENSIC-55-DOI KUEML-51-NCH KEUNLE-52-CIW BISCH - -BSE SEWEN -57-RTI JELLEM-48-AAA WEILEH- -PMA SEWEN -54-PIC PATTID-55-PCA CORTE -55-ATR MALOQJ. -ATR DYSOG -51-RAI BIOLAB-56-BRJ KIRSH- 56-SRS CHEMEN-52-SAI MOGEJH- -OAS EASTWR-50-WCW BOWPR-56-YCS MILEJH-57-AIS GARITE-56-PCA JOHNRK-54-NSH JELLEM-48-AAA FAIRRA -ARS PARKRA-56-AT MIKHAJ-58-PAT BAILMP- -PCD JOURDO-56-BPC STONW -56-ACR GULLCP-53-ASI	THE INADEQUACY OF THE CORRELATIVE INDEXES 1. THE INFORMATION CENTER OF INTRODUCING DOCUMENTARY REPRODUCTION GATHERING HABITS OF SOME NOTES ON PROCEEDINGS CONFERENCE OF SOURCES- A CHALLENGE TO PROJECT ON THE SURVEY OF TROPICAL INSTITUTE. A METHOD OF KEEPING IEM 990 SPECIAL INDEX CORDING OF SURGICAL AND TION IN THE LIBRARY- AN AN USE OF ENT AND PROOF SERVICES. E DATA HANDLING SYSTEM.	ALPHABETICAL SUBJECT INDEX, ALPHABETICAL CORRELATIVE INDEXES, AMBIGUITY OF SYNTACTIC FUNCTION RESO AMERICAN BAR FOUNDATION PROJECT ON I AMERICAN CYNAMID-S STAMPORD LABORATO AMERICAN DOCUMENTATION. AMERICAN HISTORICAL ASSOCIATION. AMERICAN MEDICAL SCIENTISTS. AMERICAN PRACTICE IN DOCUMENTATION. AMERICAN SCIENTIFIC AND TECHNICAL AB AMERICAN SCIENCE AND INDUSTRY. AMERICAN STATUTORY LAW. AMSTERDAM-NETHERLANDS. ANAESTHESIA RECORDS AND ASSESSING RE ANALYZER. ANESTHETIC DATA IN MILITARY SERVICE, ANNOTATED BIBLIOGRAPHY. ANTIBIOTIC LITERATURE FILE FOR CHEMI APERTURE CARDS FOR THE CONSOLIDARION APC. APPENDIX A, MACHINE TRANSLATING OF I	POLLAR- -IAS BERNC -56-CII KOUTA -57-ASF MACK -57-ABF PUNKCE-57-ICA TATEVD- -IAD ORLIB - -CPC HERNS- -IHA COBILH -50-SNA COMPAS- -PCA SHERJH-58-IRC MACK -57-ABF HECKEP-55-DET NOSWOR-43-MXA MURPEW-58-19S WANGOP-41-BRS KENTA -56-ALA CHRMME-54-ALP SIATPM-58-IAC KENTP.-53-TIU RAMOWC-57-18T
PRESENT CONDITIONS OF CT HEADING LIST FOR THE ABSTRACT ARCHIVE OF AN THE FUNCTIONING OF THE FIELD OF PETROLEUM AND S ON CLASSIFICATION AND	AIR FORCE INFORMATION SERVICES IN IT AIR UNIVERSITY PERIODICAL INDEX- AN ALCOHOL LITERATURA. ALGEBRAIC REPRESENTATION OF STORAGE ALGEBRAIC THESAURUS, ALL-UNION INSTITUTE OF SCIENTIFIC TE ALLIED SUBJECTS, ALLIED SUBJECTS, ALPHA-NUMERIC CHARACTER READER, ALPHABETIC, SUBJECT INDEXES AND COORD	GARITE-56-PCA JOHNRK-54-NSH JELLEM-48-AAA FAIRRA -ARS PARKRA-56-AT MIKHAJ-58-PAT BAILMP- -PCD JOURDO-56-BPC STONW -56-ACR GULLCP-53-ASI	THE AMERICAN OF BABEL,* THE LANGUAGE AKING THE SUBJECT INDEX THE SERVICE, NAT. BU. STOS. INTELLIGENT AGES CARRIED OUT ON THE VAC COMPLES A COMPLETE	RACERIOLOGICAL PAPERS. BAILEY-S AFRICAN-ORIENTAL CATALOGUE BAR FOUNDATION PROJECT ON THE SURVEY BARRIER IN SCIENCE. BARRIER-A CITATION INDEX FOR CHEMICA. BARTER OF KNOWLEDGE.* AN ANALYSES OF BATEM SYSTEM, BEHAVIOR IN PROBLEM-SOLVING MACHINES B. E. S. M. BIBLE CONCORDANCE.	READWR-53-PCA BAILBS-55-BAC MACK -57-ABF VICKBC-55-BEL GARPE -57-BSI MOHRPE-56-DEA NAT BS-55-1RS GELEHL-58-18P MUKHIS-56-EMF MCCUWR-56-UCC

BIBLIOGRAPHY

- ABA OR-54-SPC A.B.A. OPERATIONS RESEARCH COMMITTEE
SUBCOMMITTEE ON PUNCHED CARDS AND LEGAL LITERATURE
SEARCHING.
NATURE AND SCOPE OF PROGRAM JUNE 1954
- ADAIWC-55-CIS ADAIR WC
CITATION INDEXES FOR SCIENTIFIC LITERATURES.
AMERICAN DOCUMENTATION, 6. (1). 1955.
- ADAMS -56-1NR ADAMS S
INFORMATION - A NATIONAL RESOURCE.
AMER DOC. V. VII. NO. 2 APR 1956
- ADIAWC-55-CIS ADIAR WC
CITATION INDEXES FOR SCIENCE.
AM. DOCUMENT. 6. 31 (1955).
- ADKIEW- -DPL ADKINSON BW
DATA PROCESSING AND LIBRARY OPERATIONAL PROBLEMS.
LIBRARY OF CONGRESS
- ADKIEW-56-IUR ADKINSON BW LIBRARY OF CONGRESS
INTERNATIONAL UTILIZATION OF RECORDED KNOWLEDGE.
CHAPT. VIII IN DOCUMENTATION IN ACTION REINHOLD
PUB CORP 1956.
-
- ANDRD- -MFG ANDREWS DD
MAJOR FACTORS GOVERNING THE SELECTION OF MACHINE
DATA HANDLING TECHNIQUE.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRD- -PEP ANDREWS DD
PATENT EXAMINING PAST, PRESENT AND FUTURE.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRD- -PRU ANDREWS DD
PROGRESS REPORT ON U.S. PATENT OFFICE MECHANIZED
SEARCHING.
U.S. PATENT OFFICE OF RESEARCH AND DEVELOPMENT
- ANDRJD- -DEP ANDRIOT JL
DOCUMENTS EXPEDITING PROJECT.
LIBRARY JOURNAL. LXXVII, 693-695.
- APPLPL-52-SHL APPLIED PHYSICS LAB. JOHNS HOPKINS UNIV. NOV. 1952
SUBJECT HEADING DIST FOR GUIDED MISSILES.
APPLIED PHYSICS LAB. JOHNS HOPKINS UNIV. NOV. 1952
- ARGECJ-56-MPT ARCEAUX CJ (ETHYL CORP. BATON ROUGE)
MOVIES PRESENT TECHNICAL PAPERS.
CHEM. & ENG. NEWS 34, 627-9 FEB. 1956.
- ARMEFM-56-PEM THE ARMED FORCES MEDICAL LIBRARY
PROPOSED BIBLIOGRAPHY OF MEDICAL CONGRESS PUBLICATIONS.
UNESCO BULL. FOR LIBRARIES. 10. 22. JAN JAN. 1956
- ARMST-57-MAS ARMES SERVICES TECHNICAL INFORMATION AGENCY WASHINGTON
FEB. 1957.
MULTIPLE ASPECT SEARCHING FOR INFORMATION RETRIEVAL.
ARMED SERVICE TECHNICAL INFORMATION AGENCY WASHINGTON
FEB. 1957.
- ARMYAF-55-N5V ARMY-AIR FORCE REGISTER. 77. 8. 15 OCT. 1955
NEW 50-LANGUAGE VARI-TYPER DEVELOPED.
ARMY-AIR FORCE REGISTER. 77. 8. 15. OCT. 1955
-
- BAILECF-46-PCI BAILEY CF
CASEY RS
COX GJ
PUNCHED CARDS FOR INDEXING SCIENTIFIC DATA.
SCIENCE 104. * 181. 1946
- BAILEM-53-MSU BAILEY M
MECHANIZED SEARCHING IN THE U.S. PATENT OFFICE.
JOUR. P. PATENT OFFICE SOCIETY. P. 566 Aug. 1953

Ügyelni kell továbbá arra is, hogy a sűrített szójegyzékben szereplő értéktelen szavak (a már említett névelők stb.) a további feldolgozásban ne vegyenek részt, hiszen a témafigyelés szempontjából semmiféle jelentőségük nincs. Ennek érdekében, ugyancsak a rendszeres munka megkezdése előtt jegyzéket készítenek az értéktelennek tekintett szavakról, és ezt is tárolják a gép memóriájában.

A dokumentum sűrített szójegyzékének feldolgozása a szabványos szavak szótárának és az értéktelen szavak jegyzékének segítségével hívásával a következőképpen történik:

Először az értéktelen szavak jegyzékével hasonlítják össze a dokumentum sűrített szójegyzékét, és egyezés esetén a szót törlik a sűrített szójegyzékből. Azután a szabványos szavak szótárával történik meg az összehasonlítás, mégpedig oly módon, hogy a cikk címének és a referátumnak értékes szavait egybevetik a szótár azon oldalával, amelyik a nem szabványos szavakat tartalmazza, és a szótár másik oldalán szereplő megfelelő szabványos szavakkal helyettesítik azokat. Az előzetes ábc rendezés következtében ez az összehasonlítási folyamat igen gyors, mert a gépnek csak egyszer kell végigfutnia a szótáron, vagyis a mágneses szalagon vagy tárcsán. Ha pedig a szótár szavait is hosszúságukkal súlyozták, a művelet elvégzéséhez szükséges idő még azzal is lerövidül, hogy a gép a dokumentum sűrített szójegyzékének szavait az azonos betűvel kezdődő szótári szavak közül is csak azokkal hasonlítja össze, melyek a szövegszóval azonos vagy annál kevesebb számú betűt tartalmaznak. Így pl. a 7INDEXES és a 8INDEXING szavaknak a szótárban egyaránt az 5INDEX szó felel meg. Ha a dokumentum szavai között olyan értékes szó fordul elő, mely a gépi

szótárban még nem szerepel (új eljárás, berendezés, alkatrész stb. neve), akkor ezt a számítógép a vezérlő pulton levő írógépen automatikusan kinyomtatja, s a szótárba szükség esetén be lehet építeni. Így a technikai fejlődéssel lépést lehet tartani, és a gépi szótár mindig időszerű lesz.

Az összehasonlítások eredménye végül is egy olyan szójegyzék, mely csupán szabványos értékes szavakat tartalmaz. Ezt a szójegyzéket kell sorra összehasonlítani az egyes érdekeltek profiljával. Ha a dokumentum szójegyzéke és a profilok között megegyező szavakat találnak, a számítógép gyorsnyomtató egysége kitölt egy témafigyelő kártyát, amit majd az érdekeltek megküldenek.

Az összehasonlítási folyamatokat minden dokumentum és minden profil esetében meg kell ismételni.

A számítógéppel készített témafigyelő kártyák mérete rendszerint kétszerese a Hollerith kártya méretének, és középen perforálással látják el azokat. A kártya egyik része a kiértékelés, a másik a válasz céljára szolgál. A kiértékelő rész a felhasználó nevét, címét, a dokumentum azonosítási számát, tárgyszavait vagy referátusát és a cikket tartalmazó folyóirat adatait tartalmazza, míg a válasz részen a dokumentum azonosítási számát, tárgyszavait, valamint a felhasználó nevét és címét tüntetik fel (ez utóbbi azonban itt már feladóként szerepel).

A felhasználó a témafigyelőkártya alapján nyilatkozhat, hogy

- Az anyag érdektelen volt számára
- Az anyagot érdekesnek találta, de a referátum elolvasása kielégítette, részletes információt nem kér
- A dokumentumot eredetiben kívánja elolvasni

A témafigyelő kártya válaszcészén a felhasználó a megfelelő rovatot bekarikázza, és a válaszcészt lépve visszaküldi a dokumentációs központnak. A kiértésítő részt megőrizve saját irodalmi kartotékot építhet ki magának.

Ha a dokumentációs központ a visszaküldött válaszcészelvény alapján értesül arról, hogy a dokumentum az érdekeltnak megfelelt, akkor egyrészt azonnal gondoskodik a cikk eredetiben vagy másolatban történő megküldéséről, másrészt a dokumentumot jellemző tárgyszavakat gépi úton ismételtlen egybeveti az illető profiljával, és a profilban addig még nem szerepelt tárgyszavakat a profilhoz hozzáteszi. Így a profilok változását jól követni lehet. Időszakonként természetesen az érdekelttel való megbeszélés alapján a profilból egyes szavakat törölni kell.

Olyan vállalatoknál, amelyeknél a folyóiratok rendszeres körözésének elvét fogadták el, a körözés „menetrendjét” is számítógéppel lehet elkészíteni. Ilyenkor folyóiratonként összegyűjti a gép azoknak a nevét, akiknek valamelyik cikk az érdeklődési körükbe esik, s a megfelelő oldalszámokkal együtt egy listában kinyomtatja.

Az IST rendszer üzemben tartásának költségeit három tényező befolyásolja: a profilok száma (ami az érdekeltek számával egyenlő, vagy annál több lehet), a feldolgozott dokumentumok száma és az igénybe vett számítógép típusa, illetőleg üzemköltségei. Tájékoztatásul közöljük, hogy egy olyan vállalat esetében, ahol a profilok száma 3000, a naponta feldolgozott cikkek száma 50 és az igénybe vett számítógép a rendkívül gyors működésű, nagy teljesítményű IBM 7044 és IBM 1401-es gépcsoport volt, az IST-rendszer költsége profilonként havi 6,50 \$-ra adódott, ami a naponkénti tájékoztatást figyelembe véve igen kis összeg.

Fel kell hívnunk a figyelmet arra, hogy a dokumentáció gépesítéséről beszélve sohasem arra gondolunk, hogy akár egy nagyvállalat is kizárólag dokumentációs célra vásárol számítógépet. De az ügyvitelgépesítés előrehaladásával egyre nő azoknak a vállalatoknak a száma, amelyek kifizetődőnek tartják bérelszámolásukat, anyagnyilvántartásukat, és általában teljes ügyvitelüket számítógéppel elvégezni. Ebben az esetben pedig mindig megoldható, hogy hetenként mindössze néhány órára a szakirodalom feldolgozásánál is segítségül hívják az elektronikát, ami a fejlesztés, szabadalmi vizsgálatok stb. terén sokszorosan megtérülő befektetést jelent. Az IST-rendszer alkalmazói között pl. ismeretes olyan vállalat, mely mindössze 120 mérnököt tájékoztat rendszeresen a könyvtárhoz beérkező 60 folyóirat alapján, az üzemgazdasági munkák elvégzésére beszerzett IBM 1401-es számítógép heti 1 óras kölcsönvétele útján.

A dokumentációs iroda az elektronikát még abban az esetben is felhasználhatja, ha csak egészen rendszeretlenül, esetlegesen jut hozzá a vállalati számítógéphez vagy csupán arra kap engedélyt, hogy valamely számítóközpont gépét néhány órára bérelje. Ez alatt a rövid idő alatt is olyan értékes segédeszközöket tud létrehozni, amelyeket akár minden

szakembernek állandó használatra odaadhat, és azok a segédeszközök felhasználásával maguk végzik el saját tájékoztatásukat. Ilyen elterjedten használt segédeszközök a legújabbban beérkezett vagy bizonyos időszak alatt feldolgozott irodalmi anyagból készített tárgy (esetleg szerző stb.) szerinti indexek. Az indexkészítés gépi módszerei közül a KWIC-rendszert használják a leggyakrabban, ezért ennek ismertetését adjuk meg egészen röviden.

2. A KWIC indexek

A KWIC-rendszer lényege, hogy az indexet a dokumentumok címében szereplő értékes szavakból állítja össze, mégpedig oly módon, hogy az index vezérszavának tekintett címszó körül a szövegrészt is feltünteti amelyben az illető címszó szerepel. Így lehetőség nyílik arra, hogy a felhasználó érzékelje a címszó pontosabb jelentését és fontosságát. Erre utal a rendszer elnevezése is, amely a Key Word In Context (kulcsszó a szövegben) kifejezésben szereplő szavak kezdőbetűiből tevődik össze és sajátos hangutánzó módon fejezi ki azt a gyorsaságot, amellyel a rendszer dolgozik (a hasonló hangzású angol „quick” szó jelentése: gyors). Mivel a vezérszavak a szövegben hol itt, hol ott vannak, az index áttekinthetővé tétele érdekében speciális megoldást alkalmaztak. A KWIC-index sorait a gép úgy nyomtatja ki, hogy a vezérszavak a szöveg megelőző részeitől elhatárolódó külön oszlopban állanak és így feltűnőkké válnak.

A KWIC-rendszer a dokumentumoknak csak címét vizsgálja, bővítve azt, ha nem utal eléggé a tartalomra, és szűkítve, ha túl terjedős. Elvileg semmi akadályja sincs azonban annak, hogy egy referátum teljes szövegéből készítsenek ilyen indexet.

A közlemény tartalmának jellemzéséhez felhasználható értékes és a tulajdonképpen jelentés nélküli, értéktelen szavak (pl. névelők stb.) két különböző módszerrel választhatók szét.

Az első módszer a címet alkotó szavakat az értékes szavak előre elkészített jegyzékével hasonlítja össze, míg a második módszer az összehasonlítás alapján az értéktelen szavak jegyzékét tekinti.

Nyilvánvaló, hogy az első módszer alkalmazása esetén értéktelen szó véletlenül sem kerülhet az indexbe, de egyben sok információ elvesztéséhez is vezet, mivel mindig újabb és újabb értékes szavak alakulnak ki a műszaki nyelvben. A második módszer elterjedtebb, de alkalmazása esetén elkerülhetetlen, hogy néha kis jelentőségű szó is vezérszóként szerepeljen az indexben.

A KWIC-index készítésének fázisai a következők:

A dokumentumok jellemző adatait lyukkártyákba lyukasztják. A jellemzőeknek tartott adatok rendszerint az alábbiak:

- a) szerző (vagy szerzők) családneve és keresztnévének kezdőbetűi,
- b) cím, esetleg olyan kifejezésekkel kiegészítve, amelyek nagyobb pontossággal írják körül a közlemény tartalmát,

- c) bibliográfiai adatok, mint: a kiadó neve, a kiadási hely, az időszakos kiadvány neve, kötete vagy száma, lapszám, az oldalak és ábrák száma, a megjelenés időpontja stb.,
d) esetleg a kiadvány rövid kivonata.

Az a), b) és c) pontokban foglalt információk összessége IBM 1401 típusú adatfeldolgozó gép alkalmazása esetén nem haladhatja meg a 660 betűt, kisebb teljesítményű gép alkalmazása esetén ez a határ általában 540 betűre csökken.

Az adatfeldolgozó gép elolvassa a lyukkártyákat, és önműködően előállítja a hivatkozási kódot.

A hivatkozási kód feladata a vizsgálat alá vett kiadvány azonosítása oly módon, hogy az azonos dokumentumból származó különböző információkat ugyanazzal a hivatkozási kóddal jelöli meg. A hivatkozási kód a KWIC-indexek minden tétele után megjelenik, és összefüggést biztosít a tárgy és a szerző szerinti index között. A kiadványhoz kijelölheti maga a dokumentátor, de hozzárendelheti automatikusan az elektronikus adatfeldolgozó gép is, igen pontos szabályok szerint, amelyek messzeemenően lehetővé teszik mind a különböző dokumentumok kódjai közti esetleges alaki hasonlóságok elkerülését, mind a nagy mnemotechnikai tartalmú kódok előállítását.

A hivatkozási kód előállításakor a következő elemeket célszerű figyelembe venni:

- a) a szerző család- és keresztnévét,
b) a kiadás évét,
c) a dokumentum címét.

A hivatkozási kód 11 jegyből áll, és az egyes számjegyeket az alábbi szempontok szerint határozzák meg:

a) az első hat jegyet az első szerző család- és keresztnévéből lehet kivenni, pl.

Rosetti C. = ROSETC
Rossetti C. M. = ROSSCM
Rossetti Carlo = ROSSCA
Rui Giovanni = RUIGIO
Bo Pietro = BOPIET

b) A következő két számjegy a kiadás évét jelzi (a tizes és egyes helyiértékű számjegyek) pl. 1958 = 58;

c) a három utolsó számjegy a cím első három szava kezdőbetűjének felhasználásával nyerhető, de alakilag különböznie kell az ún. értéktelen szavaktól. Pl.:

I Nuevi Sistemi per La Documentazione (Új Dokumentálási Rendszerek) = NSD, Documentazione Automatica: glo Indici KWIC (Automatikus Dokumentáció: a KWIC-indexek) = DAI

Un Sistema Ridotto di Documentazione Automatica (Automatikus Dokumentálás Redukált Rendszere) = SRD

A hivatkozási kód előállítása után a gép mágnesszalagon rögzíti a kártyákon lyukasztott bibliográfiai adatokat (bibliográfiai szalag) és a címeket (címszalag), majd

— elolvassa az így létrehozott bibliográfiai szalagot, és kinyomtatja a kiadványok jegyzékét.

Egyidejűleg mágnesszalagra rögzíti az összes szerző nevét (szerzői szalag);

- a család- és keresztnéveket rögzítő szerző szalagot feldolgozva az elektronikus számítógép a neveket ábécé sorrendbe rendezi, és kinyomtatja a KWIC-index név szerinti részét;
— ezután az adatfeldolgozó gép feldolgozza a címszalagot is, és a címben előforduló minden szóra nézve elvégzi a cím permutációját: a kapott permutációkat feljegyzi egy mágnesszalagra (vezérszó-szalag). Kivételt csupán a leggyakrabban előforduló értéktelen szavak képeznek, amelyeket a gép memóriájában tárolnak, és automatikusan törölnek a szövegből. Ezek a szavak már a permutációban sem vesznek részt.
— az adatfeldolgozó gép végül elolvassa az ábc sorrendben készült vezérszalagot, és kinyomtatja a KWIC-index tárgy szerinti részét. A kinyomtatás előtt azonban még összehasonlítja minden egyes tétel vezérszavát a kártyákra lyukasztott és azokról a gép memóriájába bevitt értéktelen szavak jegyzékével, és csak akkor végzi el a kinyomtatást, ha nem talált azonosságot. Ha az indexben annak kinyomtatása után kevéssé jelentős szavakra vett permutációkat is találnak, az utolsó fázis könnyen megismételhető oly módon, hogy a nem kívánt vezérszavak az indexből most már kimaradnak.

A feldolgozás különböző fázisainak végén tehát az alábbi jegyzéket kapják:

- a) a KWIC-index tárgy szerinti része,
b) a KWIC-index név szerinti része,
c) a kiadványok jegyzéke, amely a lyukkártyákra felvett összes bibliográfiai adatokat tartalmazza.

A különböző részek közötti összefüggést a hivatkozási kód biztosítja, amely megjelenik a tárgy szerinti rész minden tétele, minden szerző neve és a kiadványok jegyzékében idézett bibliográfiai adatok mellett.

Az adatfeldolgozó gép által kinyomtatott három jegyzéket azután fotográfiai eljárással kb. eredeti méretének 50%-ára lekicsinyítik, és fotorota eljárással sokszorosítják.

A KWIC-indexek felhasználásának módja röviden a következő: A kutató — miután eldöntötte, hogy melyik szó határozza meg legjobban azt a tárgyat, amelyre nézve tájékozódni kíván-, megvizsgálja, hogy a KWIC-index tárgy szerinti részében szerepel-e az illető szó vezérszóként. Ha a keresésnek nincs pozitív eredménye, akkor más hasonló jelentésű vezérszót kell keresni; ha viszont a keresett tárgyszó szerepel vezérszóként, akkor az azt körülvevő szövegrészből bizonyos mértékben látni lehet, hogy maga a dokumentum megfelel-e az érdeklődő igényének. Ha igen, akkor a tétel jobb oldalára nyomtatott hivatkozási kód megadja az útmutatást, hogy a kiadványok jegyzékében hol találja meg a dokumentum bibliográfiai adatait.

A KWIC-indexek mindenütt érdeklődést keltenek és tetszésre találtak. Nálunk is sokan hasz-

nálják pl. a Chemical Titles című kiadványt, amely már a megindulása utáni első évben 68 400 címet, 147 000 szerzőt dolgozott fel, az index összes tételeinek száma 410 000 volt. A munka időigényére jellemző, hogy egy 125 oldalas kiadvány elkészítéséhez 3 órára kell egy IBM 7044, valamint 2 órára az IBM 1401 típusú számítógépet igénybe venni. A Chemical Titles szerkesztése mindössze 4 személy munkáját igényli: a létszámban egy vezető, egy dokumentátor és 2 kártyalyukasztó szerepel.

A dokumentációs munkában felhasznált számítógépekkel szemben támasztott követelmények különböznek a számítástechnikai célokra igénybe vett gépek kiválasztásakor érvényesített szempontoktól. Ebben az esetben a nagy kapacitású belső és külső memória, a gyors, alfanumerikus be- és kimeneti egységek azok, melyek feltétlenül szükségesek, de viszonylag lassúbb működésű gép is megfelelő lehet. Gépi témafigyelés megvalósításakor nagyon előnyös, ha a külső memória közvetlen keresésű (random access) mágneses tárcsa vagy kártya, mert ezt az egyes dokumentumok, tárgyszavak, érdekeltek

stb. megtalálásának idősükségletét a mágnesszalag (sorrendi keresésű) memóriához képest töredékre csökkent. A legelterjedtebben az IBM 1401-típusú, mágnes tárcsával ellátott számítógépet használják, de gyakori a Honeywell, Univac, General Electric számítógépek dokumentációs felhasználása is.

Befejezésül tegyünk fel egy kérdést: Megoldható lenne-e hazánkban, jelenlegi adottságainkat figyelembe véve, kiemelt fontosságú területek, mint a híradástechnika, szakirodalmi tájékoztatásának gépesítése és gazdaságos lenne-e ez? A válasz igenlő kell hogy legyen. A beszerzett Elliott 803-típusú számítógépek, ha nem is a legalkalmasabbak, de azért felhasználhatóak lennének heti néhány órában dokumentációs célokra, és tekintettel arra, hogy a világszínvonal tartása, a legfrisebb eredmények ismerete az egész híradástechnikai ipar létfontosságú szükséglete, a gépek igénybevételének költségei sokszorosán megtérülnének már akkor is, ha csupán egyetlen újabb gyártmányunkat tudnánk külföldi piacokon bevezetni.

S Z E M L E

Összeállította: V Á S Á R H E L Y I P Á L

Október 2. és 10. között rendezik Ljubljában a 12. Nemzetközi Elektronikai, Távközlési, Automatizálási és Magfizikai Kiállítást. A hazai kiállítókön kívül számos külföldi cég is képviselteti magát ezen a nagyszabású kiállításon, melyen az elektronikai szekcióban ipari és orvosi elektronikus berendezéseket, elektronoptikákat, rádió- és televízió vevőberendezéseket, elektroakusztikai berendezéseket, lemezjátszókat és magnetofonokat, a távközlési szekcióban vezetékes és vezeték nélküli távközlő berendezéseket, telefonkészülékeket és telex-berendezéseket, az automatizálási szekcióban gépek és berendezések automatizálására, valamint számítások végzésére alkalmas elektronikus számítógépeket, távmérő berendezéseket, a mérés-technikai szekcióban elektronikus és elektromos mérőműszereket, valamint e berendezések felépítéséhez szükséges alkatrészeket és részegységeket, és az atommagfizikai kutatásokkal kapcsolatos mérő-, ellenőrző berendezéseket mutatnak be.

A kiállítás tartama alatt a távközlés és mikroelektronika kérdéseivel kapcsolatban szaktanácskozásokat is rendeznek.

*

1964-ben Angliában 2,2 millió TV-készüléket gyártottak, melyből a belső piacon 1,9 milliót értékesítettek. A rádiókészülékek és zeneszerekények termelése ezzel szemben 1964-ben az előző évhez képest gyengén csökkenő tendenciát mutatott: a 2,7 millió db-os termelés 10%-kal maradt el az előző évi színvonalától. A lemezjátszó-forgalom több mint 400 000 db volt, de ugyancsak elmaradt az előző évi termeléstől. Magnetofonok területén a forgalmat kb. 450 000 db-ra becsülik. (Nachrichten für Aussenhandel, 1965. IV. 20.)

Az angol rádióipart meglehetősen aggasztja a nagyvolumenű honkongi és japán rádióimport. A Radio and Electronic Component Manufacturers Federation évi jelentése szerint 1964-ben Honkongból 1,4 millió db, Japánból pedig 180 000 db rádiókészüléket importáltak.

*

Az Egyesült Államokban a fekete-fehér TV-készülékek gyártása 1964 első 11 hónapjában 1963 hasonló időszakához

hasonlítva 13,7%-kal növekedett. A televíziógyártás igazi növekedését azonban a színes vevőkészülékek jelzik. A közölt időszakban kibocsátott 8,6 millió készülékből 1,3 millió, azaz 15% volt színes készülék. A Radio Corp. of America főnöke szerint, cégének 1965. évi beruházási költségeiből a legtöbbit a színes televíziós készülékek gyártására fordítják, amely szerinte Amerika leggyorsabb növekvő iparága. (Iron Age, 1965. jan. 21. 195. k. 21. p.)

*

Az USA elektronikai ipara az elmúlt években jelentős mértékben előretört a többi iparághoz képest; 1964-ben 16 200 millió értékben adtak el elektronikai gyártmányokat. Eddig az évi növekedés mintegy 8%-ot tett ki, de az irányzat most visszafordulóban van, és 1975-re az évi növekedés előrelátás szerint már 3%-ra csökken.

Az elektronikai iparon belül a leggyorsabban a számítógépek gyártása fejlődött. Az alkatrészeket tekintve az integrált áramkörök gyártása van és lesz a legnagyobb fejlődésben, évi 30%-os emelkedéssel. Az integrált áramkörök és félvezetők gyártása volt az, amely az USA elektronikai iparát az ország harmadik legfontosabb termelési ágává tette a termelési érték alapján. (A teljes termelést alapul véve az iparág a gépkocsigyártás, az acélipar, vegyipar és repülőgépgyártás után az ötödik helyet foglalja el.)

A katonai elektronika továbbra is dominál a elektronikai iparon belül, de már nem a leggyorsabban növekvő terület. Az ipari és fogyasztói elektronika a piac 39%-át foglalta el. Mintegy 5000 cég foglalkozik az elektronikával, de a koncentrációra jellemző, hogy a 20 legnagyobb cégre esik az évi eladások 50%-a, és csak igen kicsi az újonnan alakuló cégek száma. (The Financial Times 1965. ápr. 12. p. 87.)

*

Spanyolországban a rádió és televízió készülékek gyártása Madrid és Barcelona térségében koncentráldott. A foglalkoztatottak száman ebbe az iparágban jelenleg kerekén 14 000. Az alkalmazottak száma 5 vállalatnál van 500 felett, 15 üzemenél 100 és 500 között, míg a többi üzem 100 munkaezrőnél