

KWIC-INDEX ELŐÁLLÍTÁSA PERFOTECHNIKÁVAL*

PATAKY ERNŐ – OROSZ GÁBOR – BALÁZS SÁNDOR

Országos Műszaki Könyvtár és Dokumentációs Központ

Az automatizált indexkészítés egyik eljárásának terméke a KWIC-index. Nevét a Keyword-in-Context megjelölés rövidítéséből nyerte, jelentése: szöveggörnyezetes, szöveggörnyezetbe helyezett index. KWIC-index készítése kedvelt és világszerte mindjobban elterjedő eljárás, mivel nem igényel nagyobb manuális előkészítő munkát; az index gépi uton, automatikusan készül.

A teljes index három részre tagolódik. Ezek: a bibliográfiai rész, a szerzői névmutató és a vezérszavas-index, a tulajdonképpeni KWIC rendszerű index.

Eddig az ilyen munkákat nagyobb vagy egészen nagy elektronikus adatfeldolgozó berendezésekkel készítették, olyan berendezésekkel, amelyeknek vezérlése tárolt programmal történt, s amelyekhez nagykapacitású külső memóriaszerkezetek /mágnesszalag vagy mágnes-tárcsaszerkezetek/ tartoztak.

PERFOTECHNIKÁVAL TÖRTÉNŐ ELŐÁLLÍTÁS

Megvizsgáltuk, lehet-e előállítani, és ha igen, milyen módon KWIC-indexet olyan adatfeldolgozó berendezéssel, amely nem rendelkezik külső memóriaszerkezetekkel, valamint nem dolgozik tárolt programmal. Vizsgálatunk pozitív eredményre vezetett, és kidolgoztunk egy új eljárást. Ezzel az eljárással olyan indexet készítettünk, amely a KWIC-indexek legfejlettebb típusának összes jellemzőivel rendelkezik.

FELHASZNÁLT BERENDEZÉS

Az index kísérleti előállítására az ügyviteli munkák elvégzé-

* A varsói "PERFO-66" szimpozionon /1966.dec.5-9./ tartott előadás nyomán készült tanulmány.

sére Magyarországon rendelkezésre álló adatfeldolgozó gépparkok közül olyat választottunk, amelyben egy UNIVAC 1004 Card Processor gép is volt. Az indexkészítés leglényegesebb munkáit ezzel a géppel végeztettük.

A UNIVAC 1004 Card Processor /Model C/ fontosabb jellemzői:

- ferritgyűrűs beépített tárolójának kapacitása 961 pozíció. Berendezésünkben még egy második beépített mágnesmagos tároló is rendelkezésre állt, ugyancsak 961 pozíciós kapacitással;
- a műveletek vezérlése kapcsolótáblával történik;
- a gép gyorsíró szerkezete percenként 600 sort ír. Sorszálessége 132 pozíció;
- jelkészlete 64 jel: 26 betűjegy, 10 számjegy, 28 írásjegy és másféle jegy;
- lyukkártyás és lyukszalagos bemenet;
- feldolgozó teljesítménye óránként 36 000 kártya.

A UNIVAC 1004 géphez kábellel csatlakozik egy kártyabeolvasó--lyukasztó gép /Read--punch/, mely percenként 200 kártyát dolgoz fel.

Az említett gépparknak még a következő gépeit vettük igénybe munkáinkhoz:

- 8-csatornás lyukszalagírógép,
- kártyarendezőgép /szorter/,
- kártyamásológép /dopler/,
- kártyaválogatógép /kollátor/.

ELŐKÉSZÍTŐ MŰVELETEK

Az indexelendő szöveg kiválasztásánál tekintettel kellett lennünk arra a korlátozó körülményre, hogy az írószerkezet 26 betűjele nem volt elegendő a speciális magyar betűk írására. Ezért választásunk az CMKDK egyik angol nyelvű kiadványára esett. A Hungarian Technical Abstracts c. referáló lap 1965. évfolyamában szereplő referátumok címleírásait vettük indexelendő szövegnek. Az írásjeli korlátozások miatt ezekből kénytelenek voltunk kihagyni azt a néhány címet, mely kémiai képletet tartalmazott. Célszerű volt továbbá néhány egyéb rövidítést is teljes kiírással helyettesíteni.

A lyukszalagírógép és a UNIVAC 1004 gép jelkészleteinek eltérése miatt csupán hat írásjelet használtunk: pont, vessző, kötőjel, ferde törtvonal /ami egyuttal zárójel is/, "cím vége" jel és spácium.

A szövegeknek a gépi feldolgozáshoz való előkészítése nem igényelt semmiféle kódolást. A vezérszavak manuális kijelölésére sem volt szükség.

Egyes alcímeket terjedelmességük és információtartalmuk miatt önálló indextételeként szerepeltettünk, de az azonosító jelzetben feltüntettük a főcímmel való kapcsolatot.

A SZÖVEGNEK LYUKSZALAGRA RÖGZÍTÉSE

Az indexelendő címeket azonosító jelzeteikkel együtt 8-csatornás lyukszalagra lyukasztottuk. A lyukszalagra semmiféle vezérlőjel nem került, csupán az egymásután következő címeket választotta el egymástól a "cim vége" jel. A pontot nem használhattuk a cím végének jelölésére, mivel némelyik cím egynél több mondatot is tartalmazott. A lyukszalag hibátlanágát egyszerűen kiírással ellenőriztük.

AZ IRRELEVÁNS SZAVAK SZÓTÁRA

Kézi úton előre egy kis szógyűjteményt állítottunk össze, mely a feldolgozandó szövegben található ama rövid szavakat tartalmazta, amelyekre az indexben keresés nem volt várható. Ezek a szavak keresés szempontjából irrelevánsak.

A szótár összeállításánál figyelemmel kellett lennünk bizonyos szempontokra, elsősorban arra, hogy a szótár elhelyezésére rendelkezésre álló belső tároló kapacitása mindössze 961 pozíció. A tárolóban minden szó számára egyenlő terjedelmű mezőt kell felvenni, tekintet nélkül arra, hogy a vonatkozó szó rövidebb, mint a mező. Vigyáznunk kellett arra is, hogy olyan szavak kerüljenek a szótárba, amelyeknek felvétele hatékonyan csökkenti a KWIC-index terjedelmét. Az optimalizálási vizsgálatok azt a 115 szót eredményezték, mely az 1.sz. ábrán látható. A szótárba felvett szavak hosszát 7 betűjegyben limitáltuk. A UNIVAC-gép kiírószerkezetével készített ábra bemutatja, hogy a szótár a 961 pozíciós tárolóban miként van elhelyezve. Az ábra elég jól szemlélteti a 31 x 31 pozíciós ferritgyűrűs tárolómátrix feltöltését a szótár szókészletével.

Amint az ábrán is megfigyelhető, a szótárba fel kellett venni a szavaknak írásjelekkel előforduló kombinációit is. Erre azért volt szükség, mert a gép belső tárolásu program hiányában az írásjeleket a szavakról nem tudta leválasztani. Az írásjeles szavakat így az eredeti szavaktól eltérő szavakként kezelte volna és az alaki különbözőség következtében a KWIC-indexbe vezérszavakként is felvette volna azokat.

Az irreleváns szavak így összeállított szótárának jóságát 38,6 %-os sormegtakarítás jellemezte a szöveg minden szavát tartalmazó indexszel szemben. Az irreleváns szavak kiszűrésének haté-

A ABILITY ABOVE AFFECT
 AFTER AGAINST AN AND
 AND/OR ANOTHER ARISING AS
 AT BASED BASIC BASI
 S DE BETTER BETWEEN BON
 DEU BUILT BY CASE CA
 USE CAUSED CAUSING CERTAIN C
 HANGES DERIVED DUE DURING
 EACH EFFECT EFFECTS EFFECT.
 EFFORTS EXTENT FACTOR FACTOR
 S FACTORY FOR FORMER FORMI
 NG FROM GAINED GENERAL HAVI
 NG IN INSIDE INTO ITS
 LIABLE M MADE ME
 ANS MEANS. NEW OF O
 FF ON ONE ONLY
 OR ORDER ORIGIN. OTHER
 OTHER. OTHER. OVER PAIRS
 PARTIAL PERIOD PERIODS PLACE
 D RAISE RAPID RAPIDLY READ
 ILY RECENT RESULTS ROLE SIM
 PLE SOME STATE STATES ST
 UDIES STUDY T TASKS T
 HE THEIR THROUGH THRUST
 TO TWO TYPE TYPES
 TYPES. TYPE. UNDER UNIFOR
 M UNIT UNITS UNIT. UNKNO
 WN USE USED USING WITH
 WITHOUT WORK WORKS WOR
 KS. / /FOR

1.ábra

Irreleváns szótár a ferritgyűrűs tárolóban

konyságára rámutat az is, hogy a releváns szavak 1,7 értékű ismétlődési tényezőjével szemben ez az érték a szótár irreleváns szavainál átlagosan 5,5.

Az irreleváns szavak szótárát szintén 8-csatornás lyukszalagra írtuk, s erről olvasta be a gép a KWIC-index készítésekor a belső tárolóba.

A VEZÉRSZAVAS INDEX SORTERJEDELMÉNEK MEGHATÁROZÁSA

A UNIVAC 1004 gép kiíró szerkezete 132 pozíciós sorok írását teszi lehetővé. A feldolgozandó címek információtartalmának vizs-

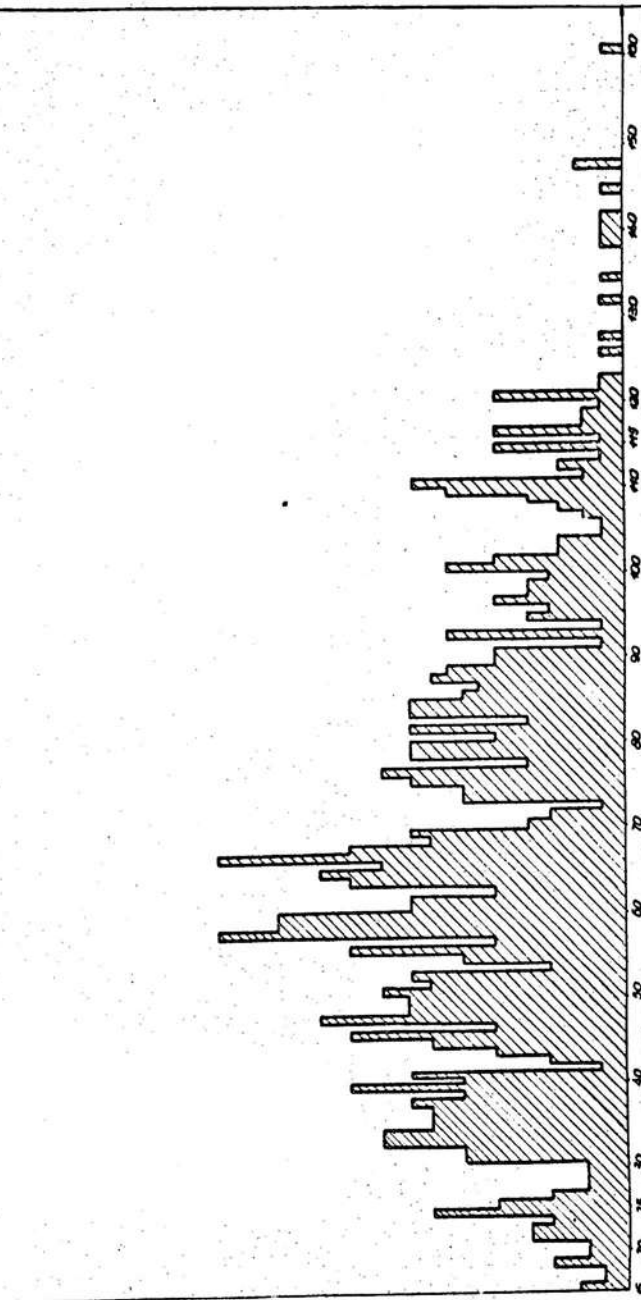
gálata alapján döntenünk kellett arra vonatkozólag, hogy egyrészt mennyiben állapítsuk meg a sorok hosszát a vezérszavas indexben, másrészt, hogy hová helyezzük a sorokon belül a vezérszavak oszlopát. A címek hosszúságának szóródását feltüntető diagram /2.sz. ábra/ azt mutatta, hogy a címek zömének hossza 19 és 123 pozíció közé esett. Az átlag 70 körül mozgott. Ezek az eredmények arra indítottak, hogy a kiírószerkezet által biztosított sorterjedelmet teljes egészében használjuk ki. Ez esetben ugyanis a címek többségét teljes egészében szerepeltethetjük, feltéve, hogy a permutálás adta sorokban un. gyűrűztetést alkalmazunk, azaz a szövegnek a sor végén tulfolyó részét a sor elején közöljük. A sorok szövegrészeinek maximális terjedelmét tehát 125 pozícióban állapítottuk meg. A fennmaradt további pozíciókba a címek azonossági jelzete került.

Vizsgálatokat folytattunk a KWIC-indexek sorterjedelme és információtartalma közötti összefüggés felderítésére. Elvként szögezhető le, hogy akkora terjedelmű sorok, melyekben a témát kifejező dokumentumcímeknek csak töredékei helyezhetők el a sor hosszától függően, csupán korlátozott mérvű vagy éppenséggel minimális információt szolgáltatnak. Az első KWIC-indexek még Amerikában 60 pozíciós szövegsorokkal készültek. Eme érték alatti szövegsorok informálóképessége a minimális igényt sem üti meg. A sorok informálóképességének fokozása érdekében a sorok hosszát később növelni kezdték. A magas információszolgáltató igénytel készült indexnél a sorterjedelem 100 pozíció fölé emelkedik. A 120 pozíció körüli érték a legfelső szintet képviseli.

A vezérszavak oszlopának elhelyezésénél egyrészt abból az elvből indultunk ki, hogy megfelelőbb, ha a sor információtartalmának tetemesebb része a vezérszót követi, és nem előtte áll, mert a sorok olvasásakor a figyelem önkéntelenül elsősorban a vezérszóra összpontosul. De ezt indokolta az a másik körülmény is, hogy a soroknak a vezérszót követő részei néha jelentős darabon betűről-betűre egyeztek, s a sorok szövegeit egymástól megkülönböztető szavak csak a sorok vége felé jelentkeznek. Mivel a vezérszavas index betűrendezésénél a vezérszó és az azt követő szövegrész a mérvadó, a betűrend minél pontosabbá tétele érdekében kívánatos volt a soroknak ezt a második felét minél hosszabbnak venni. E megfontolások alapján úgy döntöttünk, hogy a vezérszavak oszlopát a sor végétől számított 70. pozícióra helyezzük. Ebből viszont az következett, hogy az index sorainak tárolására két-két lyukkártyát kellett igénybe vennünk. Az első kártyába a sor szövegének 55, a másodikban pedig 70 pozíciónyi részlete kerül.

GEPI SZÓTÁRAZÁS, PERMUTÁLÁS, GYÜRÜZTETÉS

A szövegben előforduló szavak releváns jellegének megállapítását automatikus szótározással, majd ennek alapján a cím permutálását és gyűrűztetését a UNIVAC 1004 gép egyetlen műveletsorban végezte. Megkezdése előtt a gép kapcsolótáblájában létrehoztuk a műveletek programjának realizálásához szükséges kapcsolásokat. Azután a második 961 pozíciós tárolóba lyukszalagról beolvastuk az ir-



2. ábra
A címek hosszúsági eloszlása

releváns szavak szótárát. Ezt követőleg indult meg a szöveg feldolgozásának folyamata. A címet tartalmazó lyukszalagról a gép behívta az első feldolgozandó címet az első tárolóba. A cím első szavát összehasonlította a szótár szavaival. Ha egyezést talált, eltolta a címet a következő szóig, s ennek egyeztetése következett. Ha a cím szavát nem találta meg a szótárban, azt releváns szóként kezelte. A releváns szó figyelembevételével a címből két metszetet készített és helyezett el az első tárolónak egy további mezőjében. Az egyik metszet a releváns szót megelőző 55 szövegpozíciót tartalmazta, a másik metszetben pedig a releváns szó első betűjétől számított 70 pozíciónyi szövegrész szerepelt. Ha a metszetek készítésekor akár a sor elején, akár a végén a cím szövegének része már nem fért el a sorban, következett a gyűrűztetés. Ez abban állt, hogy a sor elején kifolyó szövegrészt a sor végén levő üres helyre tolta be a gép, illetve a sor végére már nem férő szövegmaradékot a sor üres elejére helyezte. Ezután a két összetartozó szövegmetszetet a cím azonossági jelzetével együtt egy kártyapárba lyukasztotta. A kártyapárok szövegmetszetei a KWIC-index egy-egy sorát szolgáltatották. A kártyák lyukasztása után a gép áttért a cím következő szavára, megkezdte annak egyeztetését a szótárral, amit a leírt további műveletek követtek. Amikor a gép a "cím vége" jelhez érkezett, törölte a tárolóból a már feldolgozott címet, és a lyukszalagról új címet hívott be, s annak feldolgozásához kezdett.

A címből minden vezérszó nyomán egy-egy kártyapár készült, mely a cím szövegének egy-egy permutált variánsát tartalmazta. A 3.sz. ábrán bemutatjuk az egy-egy címből készült szövegsorokat úgy, amint azokat a UNIVAC 1004-es gép a feldolgozás során kapott kártyapárokból közvetlenül a feldolgozás után kiírta.

A VEZÉRSZAVAS INDEX BETÜRENDJÉT ELŐÁLLÍTÓ MŰVELETEK LYUKKÁRTYAGÉPEKEN

A UNIVAC 1004 gép által a szövegfeldolgozás eredményeként készített kártyamennyiség a beolvasott címek egymásutánjában és azon belül a vezérszavak esetlegessége által adódó sorrendben került ki a gépből. A kártyagyűjteményből eme állapotában is kiirattunk egy jegyzéket, melyet a permutálások helyességének ellenőrzésére használtunk. E jegyzéknek részlete látható a 3.sz. ábrán.

Ahhoz, hogy a kártyagyűjteményből a vezérszavas indexet kiirathassuk, a kártyákat előbb a vezérszavak alfabetikus rendjébe kellett átrendezni. Ezt a munkafolyamatot a szokásos lyukkártyagépekkel végeztettük.

Első lépésünk az volt, hogy a kártyapárokból leválasztottuk a második kártyákat, amelyek a vezérszót és az azt követő szövegrészt tartalmazták. A második kártyákat ezután nemcsak a vezérszó, hanem az azt követő szövegrész összes pozícióján, tehát mind a 70 oszlopon rendezőgéppel szigorú alfabetikus egymásutánba rendeztettük. Ezáltal elértük, hogy a gyakran előforduló vezérszavak csoportja a vele egy kifejezéssé összefonódott második, harmadik stb. szó szerint

alcsoporthoz tagolódik, ami terjedelmes vezérszó csoportoknál nagyon meggyorsítja a keresést. Ha az alfabetikus rendezés csupán a vezérszavakra korlátozódott volna, a kapott sorrend a legminimálisabb igényt sem elégíthette volna ki. Még a manuálisan összeállított közönséges indexek alfabetikus rendje is kiterjed arra a további egy-két szóra, amely ezekben a vezérszót követni szokta. Nem lenne értelme a géppel való indexkészítésnek, ha az eredmény még a kézzel készített indexek nivelőjét sem érné el.

Ezt a sorrendet át kellett vinni az első kártyák rendezetlen mennyiségébe. E célra egy segédszámot vettünk igénybe, mellyel mint átvívó közeggel juttattuk át az alfabetikus sorrendet a második kártyák kvantumából az első kártyák kvantumába. Ezután az első kártyákat a segédszám sor segítségével az új rendbe, azaz a második kártyák szigorú alfabetikus rendjébe rendeztettük. Végül a kártyapárokat géppel újra egyesítettük.

E műveletek eredményeként a teljes kártyagyűjtemény a vezérszavak betűrendjébe került, s abban a kártyapárok egymás mellett állottak.

A VEZÉRSZAVAS INDEX KIIRATÁSA

A rendezett kártyagyűjteményből a UNIVAC 1004-es gépen irattuk ki a vezérszavas indexet, melynek részletét a 4.sz. ábra tartalmazza. A cím azonosítási jelzetét és a sorok első felét az első kártyából, második felét a második kártyából írta ki a gép. A "cím vége" jelet a kiírt szövegben a címek elejére helyeztettük. Ez megkönnyíti a sor elejének megtalálását s ennek folytán a szöveg elolvasását.

Az indexnek ama tételeit, amelyeknél a vezérszó egy numerikus érték, a kiírt szövegben a betűrend végére helyeztettük.

A BIBLIOGRÁFIAI JEGYZÉK KIIRATÁSA

A feldolgozott címeket azonosítási jelzeteik sorrendjében tartalmazó jegyzéket magáról a címeket tartalmazó lyukszalagról irattuk ki. A bibliográfiai jegyzékben azok az alcímek, amelyeket a lyukszalagon különálló címként lyukasztottunk, szintén különálló címként szerepelnek. Az azonosítási jelzetben természetesen numerikus jellel utaltunk a főcímre /5.sz. ábra/.

A címek részletesebb bibliográfiai adatait elhagytuk, mivel külön gépi munkát ezekkel a járulékos adatokkal nem kívántunk végezteni. Könnyen belátható, hogy egy teljesértékű bibliográfiai jegyzék előállítására nem jelentett volna különösebb gépi problémát.

Ugyanez volt az indoka annak is, hogy a szerzői mutató készítését szintén mellőztük.

- 0260 # A COMPARATIVE STUDY OF CARBON MONOXIDE
0265 # DETERMINATION OF THE EXTENT OF ADSORPTION WITH
0270 # PLASMA CUTTING IN
0275 A STUDY OF THE OXIDATIVE DECARBOXYLATION OF FURFURAL IN
0280 # COMBINED VARIABLE LIGHT TRANSMISSION FILTER
0285 # THE INFLUENCE OF
0290 CATION USING CODES CARRYING ADDRESSES AND CONTENT OR ONLY
0295 DETERMINATION OF THE EXTENT OF ADSORPTION WITH ACID BASE
0300 CONCENTRATION TO A PLATINUM SURFACE
0305 # INVESTIGATING THE MIGRATION OF THE MASS AND THE
0310 # INVESTIGATING THE AIR WASHERS OF
0315 OPERATION.
0320 # INVESTIGATING THE AIR WASHERS OF
0325 HANGE EQUILIBRIA OF A TERNARY CATIONIC SYSTEM IN AQUEOUS
0330 HE FATIGUE TESTING OF WELDED ALUMINUM ZINC MAGNESIUM
0335 # EFFECT OF HEAT TREATMENT OF TUNGSTEN
0340 S METALLURGY, 9. METALLOGRAPHY OF COMMERCIAL
0345 # THE EFFECT OF
0350 # THE EFFECT OF
0355 # THE EFFECT OF
0360 RATORY MEASUREMENTS.
0365 # DECOMPOSITION OF SODIUM
0370 # ON
0375 STATISTICAL EVALUATION OF THE FATIGUE TESTING OF WELDED
0380 SMALL AMOUNTS OF VANADIUM AND CHROMIUM IN PRODUCTS OF THE
0385 Y:
0390 # PHOTOMETRIC DETERMINATION OF SMALL
0395 # ITERATIVE
0400 # NEW METHODS FOR THE
0405 # STABILITY
0410 # SOLUTION OF INVOLVED TASKS IN THE FIELD OF STRESS
0415 UDY OF CARBON MONOXIDE ADSORBENTS USED IN CLASSICAL GAS
0420 # PROCESS FOR GRANULOMETRIC FLOW
0425 RACTICAL PROBLEMS OF USING ORGANIC REAGENTS IN CHEMICAL
0430 # ON ALUMINUM
0435 # A READILY REGENERABLE
0440 # A NEW METHOD AND
0445 EXCHANGE RESIN. THE TECHNICAL
0450 # PRACTICAL
0455 # BEJONG THEORY
0460 D ARC WELDING EQUIPMENT FOR USE WITH THIN WIRES AND ITS
0465 # AN
0470 ION EXCHANGE EQUILIBRIA OF A TERNARY CATIONIC SYSTEM IN
0475 # HUNGARIAN MADE CARBON DIOXIDE SHIELDED
0480 # DECOMPOSITION OF SODIUM ALUMINATE SOLUTIONS BY
0485 # STRENGTH OF CEMENTED JOINTS IN
0490 # ASYNCHRONOUS MOTORS WITH
0495 # DECAY CURVE OF EXCITED
0500 # STRUCTURE AND PROPERTIES OF
0505 # INVESTIGATION OF IRRADIATION INDUCED DAMAGE OF
0510

4. Ábra

Részlet a KWIC-indexből

- 0010/65 MECHANIZATION OF PROCESSING WORK AT THE HUNGARIAN CENTRAL TECHNICAL LIBRARY AND DOCUMENTATION CENTRE.
 0020/65 SOME EXAMPLES OF DATA STORAGE AT THE HUNGARIAN INSTITUTE FOR PETROLEUM AND NATURAL GAS RESEARCH.
 0030/65 OBSERVATIONS ON THE PLASMA ELECTROLYTIC EXPLOSION IN A FLOW SYSTEM.
 0040/65 ON THE PROXIMATE METHOD FOR CALCULATION OF HEATING.
 0050/65 ON THE KINETICALLY CHARACTERISTICS OF INDUSTRIAL HEATING.
 0060/65 THE MEDIUM POWER THERMIONIC GENERATOR.
 0070/65 A HEATLY REGENERABLE ANTIMONY GERTILLIUM NEUTRON SOURCE.
 0080/65 A NEW METHOD AND APPARATUS FOR DETERMINING LIQUID VAPOUR EQUILIBRIUM CURVES EVALUABLE BY MASS MEASUREMENT.
 0090/65 ION EXCHANGE EQUILIBRIA OF A TERNARY CATIONIC SYSTEM IN AQUEOUS ALCOHOL SOLUTIONS.
 0100/65 THE EFFECT OF SINUSOIDAL CURRENT ON ELECTRODE PROCESSES. XVII.
 0110/65 THE EFFECT OF ALTERNATING CURRENT ON THE CORROSION OF MERCURY.
 0120/65 DETERMINATION OF THE EXTENT OF ADSORPTION WITH ACID BASE ADSORPTION INDICATORS.
 0130/65 ADSORPTION OF IODINE, SILVER AND COPPER IONS FROM SOLUTIONS OF LOW CONCENTRATION TO A PLATINUM SURFACE.
 0140/65 THE CHEMISTRY OF FREE RADICALS. SYNTHESIS AND STUDY OF SOME STABLE FREE RADICALS OF THE BANFIELD TYPE.
 0150/65 SYNTHETIC LINEAR POLYMERS. XVI.
 0160/65 A CORRELATION BETWEEN THE SPECIFIC VOLUME AND MOLECULAR WEIGHT OF POLYMER HOMOLOGOUS OLIGOMERIC COMPOUNDS ABOVE THE SECONDARY TRANSITION POINT.
 0170/65 SYNTHETIC LINEAR POLYMERS. XVII.
 0180/65 CORRELATION BETWEEN THE MELTING POINT AND MOLECULAR WEIGHT OF POLYMER HOMOLOGOUS COMPOUNDS.
 0190/65 SYNTHETIC LINEAR POLYMERS. XIX. PROPERTIES OF SYNTHETIC LINEAR POLYMERS BY CALCULATION.
 0200/65 DETERMINATION OF SOME KINETICALLY INDUCED SOLID STATE POLYMERIZATION.
 0210/65 THE INFLUENCE OF ADDITIVES ON THE SOLID STATE POLYMERIZATION OF N-VINYLSUCCINIMIDE.
 0220/65 CATALYTIC PROCESSING OF ROCKS BY MEANS OF A CATION EXCHANGE RESIN. THE TECHNICAL APPLICATION OF CATALYTIC DISSOLUTION.
 0230/65 COMPLEXOMETRIC TITRATION OF ZINC IN THE PRESENCE OF COPPER.
 0240/65 A COMPARATIVE STUDY OF CARBON MONOXIDE ADSORBENTS USED IN CLASSICAL GAS ANALYSIS.
 0250/65 DETERMINING THE QUANTITY OF FREE RADICALS FORMED DURING THE RADIOLYSIS OF WATER BY MEANS OF STABLE FREE RADICAL.
 0260/65 DETERMINATION OF SMALL AMOUNTS OF PHENOL IN WATER.
 0270/65 SOME THEORETICAL AND PRACTICAL PROBLEMS OF USING ORGANIC REAGENTS IN CHEMICAL ANALYSIS. IV.
 0280/65 A STUDY OF THE DIMETHYLGLOXIME COMPLEXES OF TRANSITION METALS BY INFRARED AND ULTRAVIOLET SPECTROPHOTOMETRY.
 0290/65 KINETIC ANALYSIS IN ORGANIC CHEMISTRY. II.
 0300/65 NEW METHODS FOR THE ANALYSIS OF ROCKS AND MINERALS.
 0310/65 REACTIVITY TESTS ON URANIUM TRIOXIDES.
 0320/65 REACTIONS OF SILVER OXIDES WITH A SOLUTION CONTAINING SILVER IONS.
 0330/65 PHOTOMETRIC DETERMINATION OF SMALL AMOUNTS OF VANADIUM AND CHROMIUM IN PRODUCTS OF THE ALUMINIUM INDUSTRY.
 0340/65 A STUDY OF THE OXIDATIVE DECARBOXYLATION OF FURFURAL IN ACIDIC SOLUTIONS BY MEANS OF INDICATION WITH RADIOACTIVE CARBON.
 0350/65 PROCESS FOR GRANULOMETRIC FLOW ANALYSIS.
 0360/65 APPENDICE OF THE CATALYTIC PROPERTIES.
 0370/65 THE EFFECT OF THE SURFACE REACTION ON THE MECHANISMS AND THE STRUCTURES OF THE PERMEABILITY OF SOILS.
 0380/65 CYLINDER DEVICES TO BE PLACED IN THE GROUND FOR THE FIELD DETERMINATION OF THE PERMEABILITY OF SOILS.
 0390/65 EXPERIMENTAL DETERMINATION OF THE ION CONTENT OF RAUHAITES BY HYDROMETALLURGICAL MEANS.
 0400/65 THE KINETICS OF FLOTATION.
 0410/65 AFFECT OF SUDDEN PRESSURE DROP DUE TO BOILER TUBE FAILURE ON THE STRESS CONDITIONS OF A ROILER DRUM.
 0420/65 INVESTIGATING THE INORGANIC MINERAL INTERFERENCE OF HUNGARIAN COAL WASHINGS SUBJECTED TO FLOTATION.
 0430/65 THE EFFECT OF NITROGEN ON SOME PROPERTIES OF TYPE MTA 50 STEEL.
 0440/65 DETERMINING THE INITIAL TEMPERATURE OF MARTENSITIC TRANSFORMATION.
 0450/65 EXAMINATION OF NONMETALLIC INCLUSIONS IN STEEL BY MEANS OF RADIOISOTOPES.
 0460/65 EFFECT OF HEAT TREATMENT OF TUNGSTEN ALLOYED HIGH SPEED TOOL STEEL ON TOUGHNESS AND EDGE LIFE.
 0470/65 THE EFFECT OF ALLOYING WITH TITANIUM ON THE WELDABILITY OF TYPE MTA 50 STEEL.
 0480/65 DECOMPOSITION OF SODIUM ALUMINATE SOLUTIONS BY ARTIFICIAL COOLING.
 0490/65 ENCYCLOPEDIA OF FERROUS METALLURGY. 9. METALLOGRAPHY OF COMMERCIAL GRADE IRON ALLOYS. 2.
 0500/65 STATISTICAL EVALUATION OF THE FATIGUE TESTING OF WELDED ALUMINIUM ZINC MAGNESIUM ALLOY SHEETS.
 0510/65 THE EFFECT OF ALTERNATING CURRENT ON THE ELECTROLYTIC CORROSION OF STEEL. I. MODEL MEASUREMENTS.
 0520/65 THE EFFECT OF ALTERNATING CURRENT ON THE ELECTROLYTIC CORROSION OF STEEL. II. LABORATORY MEASUREMENTS.
 0530/65 STRESS CORROSION OF WELDED STRUCTURES MADE OF PLAIN STEEL.
 0540/65 KINEMATIC GEOMETRIC PRINCIPLES OF PRODUCING ROUNDED OFF POLYGONS.

5. ábra

Részlet a bibliográfiai jegyzékből

AZ ELJÁRÁS MENETE

A 6.sz. ábrán látható az eljárás menetének folyamatdiagramja. A folyamatára három részre tagolódik, melyek a manuális műveleteket, a UNIVAC 1004-es géppel, valamint a szokásos lyukkártyagépekkel végzett műveleteket tartalmazzák. A folyamatdiagram feltünteti, hogy az egyes műveletek milyen sorrendben követték egymást, illetve miként kapcsolódtak egymáshoz. A UNIVAC gép szektorában a szótárzás, a permutálás-gyűrűztetés és a kártyalyukasztás műveleteit egy vastagabb keretbe foglaltuk, jelezve azt, hogy ez a három művelet egyetlen összefüggő gépmenetben folyt le.

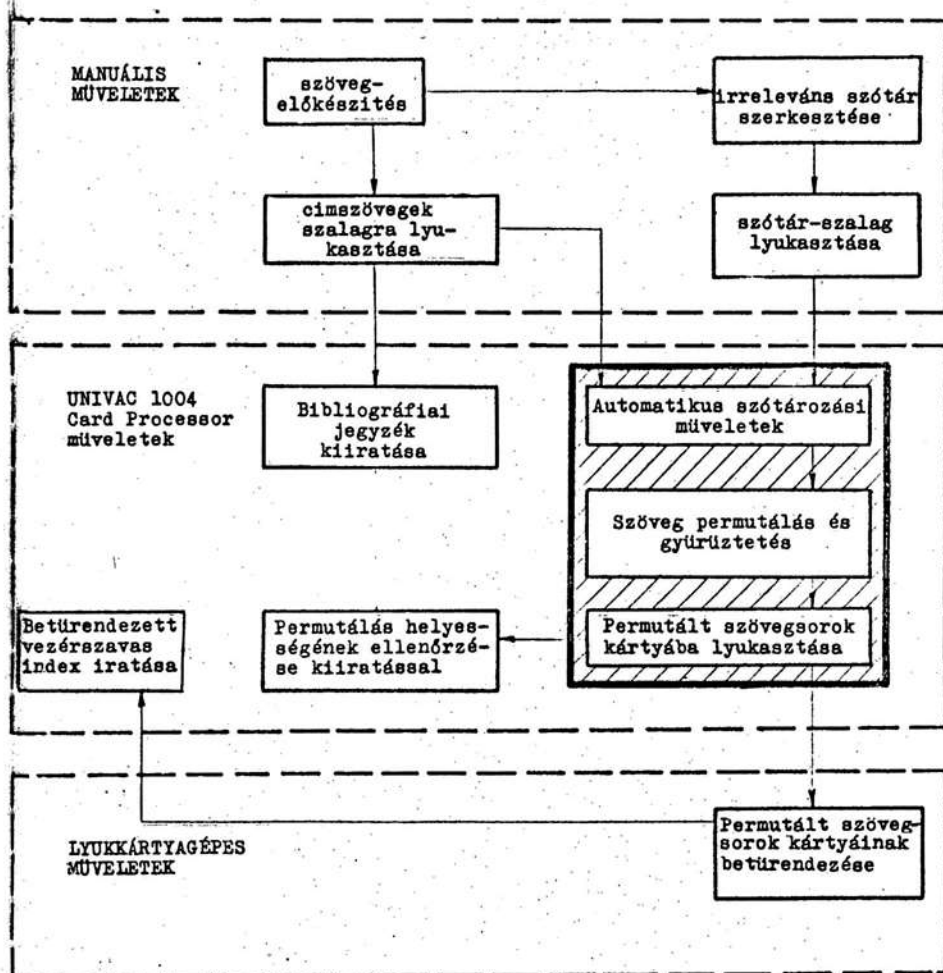
AZ ELJÁRÁS ÉRTÉKELÉSE

Az itt ismertetett eljárás feltételezi, hogy az igénybevett géppark a szokásos lyukkártyagépeken kívül még egy olyan modern, gyorsműködésű gépet is magába foglal, amely korszerű ferritgyűrűs belső tárolóval rendelkezik. Minél nagyobb a beépített tároló kapacitása, annál könnyebb egy jó minőségű index előállítás, illetve annál jobb minőségű index állítható elő a géppel. Kisebb kapacitású tároló nem teszi lehetővé az eljárás alkalmazását, csupán az index információfeltáró képessége lesz gyengébb. A kisebb tárolóban ugyanis korlátozottabb méretű irreleváns szótár helyezhető el. Viszont minél kisebb az irreleváns szótár, annál több lesz az indexben a tájékoztatás szempontjából érdektelen vezérszó, tehát annál terjedelmesebb lesz maga a vezérszavas index. Az irreleváns szótár terjedelmétől és szerkezetétől függ a vezérszavas indexnek, mint információs rendszernek a zajszintje.* Bizonyos mennyiségű zaj még az elektronikus adatfeldolgozó rendszerekkel előállított KWIC-indexekben is van, és ez az index használatánál el is viselhető. A nagyobb mérvű zaj sem teszi az indexet használhatatlanná, lassítja azonban a benne való keresést, növeli az index terjedelmét, s ezzel emeli a gépi feldolgozási költségeket. Indexünkben az irreleváns szavakat is tartalmazó teljes indexhez képest a zajszintet 38,6 % -kal sikerült csökkenteni.

A KWIC-index az irreleváns szavak szótárának igénybevétele nélkül, azaz automatikus szótározás nélkül is előállítható, mégpedig két módon: vagy mindennemű előkészítő művelet mellőzésével, vagy a releváns szavaknak manuális kijelölésével.

Az első esetben nem tesznek különbséget a szöveg releváns és irreleváns szavai között. Ennek következtében az index terjedelme nyilvánvalóan növekedni fog. Esetünkben ez valamivel több mint a fele lett volna az elkészült indexnek. Kisebb anyagoknál ilyen mér-

* Itt a zaj kifejezést információelméleti értelemben használjuk, és alatta az információ keresésénél zavaró, az indexbe feleslegesen felvett vezérszavak mennyiségét értjük.



6.ábra
Az eljárás elvi folyamatábrája

vü növekedés még elviselhető. A perfotechnikás /lyukkártyás és lyukszalagos/ megoldás még azt is lehetővé teszi, hogy a nagyon zavaró hatású indextételek kártyáit a permutálás után kézzel kiemelhessék a kártyagyűjteményből: vagyis az index javítására is van lehetőség. Különösen ezt érdemes ezt az egyszerűsített eljárást alkalmazni, ahol valamilyen oknál fogva az irreleváns szavak aránya a szövegben minimális.

A másik módszer abban áll, hogy a szöveg előzetes átnézése során, - ami amúgy sem mellőzhető, - a relevánsoknak minősülő szavak elé jelet helyeznek. E vezérlő jeleket a szövegnek gépi közegre /lyukszalagra, lyukkártyára/ való bejelölésekor szintén bejelölik. Az eljárásnak kettős előnye van. Egyrészt az index sokkal tömörebb lesz, mert abban csak azok a szavak fognak vezérszóként szerepelni, amelyek emberi elbírálás alapján minősülnek relevánsoknak. A mi esetünkben a releváns vezérszavas index terjedelmének csökkenése 46,6 % lett volna, tehát 8 %-kal több, mint az automatikus szótárazással készült KWIC-indexnél. A másik előnye pedig abban nyilvánul, hogy elmarad a szótárazás igényelte gépi idő, aminek költségkihatásai kétségtelenül nem elhanyagolhatók.

A UNIVAC 1004 gép használata esetén az indexelőállítás különféle módszerei a következők szerint alakulnak. A gép alap kivitelében csupán egyetlen 961 pozíciós belső tárolóval rendelkezik. Ez a tároló elegendő kapacitású ahhoz, hogy a gép az index előállításának vonatkozó műveleteit elvégezhesse: a lyukszalagról való beolvasást, a vezérszavak szerinti szövegeltolást, a szöveg gyűrűztetését, valamint a kártyába lyukasztást. Ebben az esetben nincs tárolóhely az irreleváns szavak szótára számára, tehát a géppel csak a fentebb említett két eljárás szerint készíthető KWIC-index, vagy úgy, hogy a szöveg minden szavára permutálunk, vagy pedig úgy, hogy a vezérszavakat előzetesen manuális uton jelöljük ki.

Fejlettebb kivitelében a gép további 961 pozíciós tárolókkal bír. Az általunk használt gépben egyetlen további tároló volt, ezért mi csak aránylag korlátozott terjedelmű szótárral dolgoztunk. A maximális felszereltségű gépekben az alaptárolót további három, egyenként 961 pozíciós ferritgyűrűs tárolóval egészítik ki. Ez már jelentősen nagyobb irreleváns szótár alkalmazását teszi lehetővé.

A UNIVAC 1004 Card Processor alkalmazása KWIC-index előállítására bebizonyította, hogy szöveggörnyezetes vezérszavas index tárolt programu elektronikus számítógép nélkül, kapcsolótáblás lyukkártyás berendezéssel is gazdaságosan előállítható. Lehetséges volt a teljes program dugaszolása a kapcsolótáblán, és ehhez a táblán rendelkezésre álló 62 programlépést nem is kellett egészen igénybevenni. A gép műveleteinek kapcsolótábláról történt vezérlése nemcsak leegyszerűsítette a programozást, de gyorsabbá is tette az index előállítását, mintha a programot belsőleg tárolták volna, amire a UNIVAC 1004 gépen is lett volna lehetőség.

A kísérletet a gépen alapos szellemi előkészítés után november elején folytattuk le. Mindössze kilenc nap állott rendelkezés-

re és 1966. november 10-én a kész KWIC-indexet sokszorosítás céljából nyomdába adhattuk. Ez volt a Magyarországon elkészült első ilyen típusú index.

""

PATAKY, E. - OROSZ, G. - BALÁZS, S.: KWIC-index produced with punched card system in Hungary

One well proved method of automatic indexing is the KWIC-index. For its production electronic data processing systems of medium or large capacity are used with stored programs. But the method, reviewed here, produces KWIC-index on a small but up-to-date punched card data processing equipment, without the use of stored program and magnetic tape units.

The equipment applied was a UNIVAC 1004 III Model-C Card Processor. The program compiling the index was controlled from a plugboard. The 62 program steps available were sufficient to produce an index with permuted titles. The index was processed by the machines on 80-column punched cards. The text to be processed was fed into the magnetic core memory by punched tapes. In addition to the basic magnetic core unit with 961 positions, the equipment had also a second built-in core memory of 961 positions. Thus vocabulary operations could be performed automatically. In this second magnetic core matrix there were stored 115 irrelevant words of the text, and with the help of the vocabulary the machine decreased the size of the alphabetical keyword index by 38,6 per cent. The printer, having 132 print positions per line /speed: 36 000 lines per hour/ enabled the processing and printing of index lines with 125 characters. These long lines were punched by the machine on two joint cards. Alphabetic ordering of the permuted index lines was done through a sorter. As an experiment, 600 titles of the English language publication: Hungarian Technical Abstracts, volume 1965, of the OMKEK /Hungarian Central Technical Library and Documentation Centre/ were processed.

The KWIC-index completed in November 1966 is the first machine-made index of this type in Hungary. The applied procedure is such economical, that preparation of further indexes is in progress.

""

ПАТАКИ, Э.-ОРОС, Г.-БАЛАЖ, Ш.: Составление указателя-КВИК при помощи перфотехники в Венгрии

В области автоматического составления указателей одним из оправдавшихся методов является указатель-КВИК. Для составления указателя-КВИК используются электронные системы обработки данных с большой программой хранения. Изложен метод по изготовлению указателя-КВИК при помощи устройства по обработке данных на малых, но современных перфокартах, при отсутствии программы хранения и памятьных единиц с магнитной лентой. Применено устройство UNIVAC 1004 III Model-C Card Processor. Управление программой по составлению указателя осуществлялось с коммутационного щита. Для составления указателя с переставленным контекстом оказались достаточными 62 программных строки, имевшиеся в распоряжении. Указатель составлен машиной на 80-ти столбчатых перфокартах. Подлежащие обработке заглавные тексты с перфокарты отсчитывались в операционное хранилище машины. Кроме операционного хранилища с магнитным стержнем на 961 позицию, представляющего собой основное приспособление, устройство имеет еще вторую встроенную память на 961 позицию. Благодаря этому предоставляется возможным выполнять операции по автоматической лексикографии. Другая ферритокольцевая матрица хранила 115 неважных слов обрабатываемого текста и машиной - при помощи словаря - размер указателя порядковых слов по алфавиту сократился на 38,6%-ов. Применением пишущего механизма, рассчитанного на скорость в 36 000 строк/час и ширину строк в 132 позиции, стало возможным обработать и выписать строки указателя в размере 125 букв. Эти длинные строки были расставлены машиной на двух присоединенных перфокартах. Расстановка пермьютированных строк указателя в алфавитном порядке осуществлялась применением сортировочной машины, снабженной коммутационным щитом. В пробном порядке было обработано 600 заглавий публикации Hungarian Technical Abstracts, вышедшей в свет в 1965 году в редакции Государственной Технической Библиотеки и Центра Научно-Технической Информации. 10 ноября 1966 года был составлен первый в Венгрии указатель-КВИК машинным путем. Данный метод оказался экономичным, вследствие чего составляются дальнейшие указатели.

i, i

PATAKY, E. - OROSZ, G. - BALÁZS, S.: Herstellung eines KWIC-Indexes mit Perfortechnik in Ungarn

Ein bewährtes Verfahren für die automatische Herstellung von Indexen ist das KWIC-Verfahren. Hierfür werden jedoch im allgemeinen grössere elektronische Datenverarbeitungssysteme mit gespeichertem Programm verwendet. Mit dem hier beschriebenen Verfahren

kann jedoch ein KWIC-Index ohne Programmspeicherung und ohne Magnetband-Speichereinheiten mit einer kleinen, aber modernen Lochkarten-Datenverarbeitungssystem hergestellt werden.

Die benützte Einrichtung ist ein UNIVAC 1004 III, Model-C Card Processor. Die Steuerung des den Index herstellenden Programmes erfolgt von einer Schalttafel. Die zur Verfügung stehenden 62 Programmlinien sind für die Verfertigung des permutierten Indexes ausreichend. Der Index wurde durch die Maschine auf 80-spaltigen Lochkarten hergestellt. Die Eingabe der zu verarbeitenden Texte erfolgte von Lochstreifen in den Kernspeicher. Die Einrichtung enthielt ausser dem zur Grundausrüstung gehörenden Magnetkern-Kernspeicher mit 961 Positionen, noch einen zweiten eingebauten Kernspeicher mit 961 Positionen. Dadurch war die Durchführung von automatischen Wörterbuch-Operationen ermöglicht. Durch diese zweite Magnetkern-Matrix wurden 115 irrelevante Wörter des zu verarbeitenden Textes gespeichert, und mit Hilfe dieses Wörterbuches konnte der Umfang des Ordnungsindex um 38,6 % vermindert werden. Der Ausgabedruker, der mit einer Geschwindigkeit von 36 000 Zeilen/Stunde hat, ermöglichte infolge seiner Zeilenlänge von 132 Positionen die Verarbeitung und die Ausgabe von Indexzeilen mit 125 Buchstaben. Diese langen Zeilen wurden von der Maschine auf zwei zusammengehörende Lochkarten gelocht. Das alphabetische Ordnen der permutierten Titel erfolgte durch eine Sortiermaschine. Probenweise wurden 600 Titel von der in englischer Sprache erscheinenden Publikation der OMKDK /Ungarische Technisch-Wissenschaftliche Zentralbibliothek und Dokumentationszentrum/ "Hungarian Technical Abstracts", Jahrgang 1965, verarbeitet.

Der in dieser Weise bis 10. November 1966 verfertigte KWIC-Index ist der erste in Ungarn fertiggestellte maschinelle Index dieser Art. Das geschilderte Verfahren ist wirtschaftlich, daher wird die Herstellung weiterer Indexe fortgesetzt.

:~: