

zák. Utóbbiakat a szerző vagy a referáló jelöli ki a cím kiegészítése végett.

A kezdeti próbálkozások után a szerzők e szolgáltatásnak csak specifikus fejezeteit vették igénybe, mert miután kutatási profiljukat szerzők és címszavak különböző kombinációiból állították össze, módjukban volt ezek körének leszűkítése. (Kezdetben számos, érdeklődési területükön kívül eső tétel is bekerült a válogatásba.)

Ugyanazokat a témákat – párhuzamosan – megfigyelhették az ASCA szolgáltatásban is. Ez mintegy 2500 folyóiratot dolgoz fel, de nem terjed ki kutatási jelentésekre, könyvekre, szabadalmi leírásokra, konferenciaanyagokra. Témaköre átfogó, beletartozik pl. a jogtudomány, a szociológia, orvostudomány és pszichológia is. Tárolja az adott cikk minden bibliográfiai adatát, de nem tartalmaz kulcsszavakat.

Jóllehet a kétféle rendszerben a keresőprofil azonos érdeklődési kört fejezett ki, mégsem volt azonos, elsődlegesen az eltérő keresési módszer miatt: a CAC kulcsszavakkal, az ASCA a szerzők nevével dolgozik.

Ideális esetben a keresés eredményeképpen meg kellene kapni az összes releváns cikket, mégpedig nem-releváns tételektől teljesen mentesen. A felhasználók és feldolgozók véleménye szerint viszont a gyakorlatban 20–40%-os relevancia-arányra kell beállni. Ennél kisebb mérvű relevancia esetén a keresőprofil nem eléggé specifikus, ennél nagyobb relevancia esetén viszont növekszik annak a veszélye, hogy egyes cikkek kimaradnak.

A szerzőknek 1972 szeptembere és 1973 áprilisa közötti időszakra vonatkozó megítélése szerint az ASCA gyorsabb, mert a mindkét rendszerben feldolgozott cikkek az UKCIS szolgáltatásában későbbi időpontban szerepeltek. A hét hónap alatt kapott 477 releváns cikkből 341 (71%) származott az ASCA rendszerből és 239 (50%) az UKCIS rendszerből; 103 (21%) volt az olyan cikkek száma, amelyeket mindkét rendszer megadott.

Az eredmények táblázatos egybevetésekor kitűnt, hogy bár a CAC tématerülete nyilvánvalóan szűkebb volt, mint az ASCA-é, mégis az UKCIS rendszer által kiválogatott anyagnak mintegy 61%-a nem szerepelt az ASCA anyagában és e 61%-nak a fele kutatási jelentés és konferencia-anyag volt. Kitűnt, hogy az ASCA-rendszerben a hivatkozások szerinti keresési lehetőség jól használható, míg a CAC-2 rendszerben a kulcsszavas keresési lehetőségnek jelentős korlátai vannak.

Jóllehet az ASCA rendszerben sikeresebben meg lehetett fogalmazni a profilokat, mint az UKCIS numerikus kódjelzeteivel, mégis az utóbbi révén sok olyan cikkhez jutottak hozzá, melyeket az ASCA segítségével nem találtak volna meg. Ez okból kifolyólag mindkét rendszert továbbra is használják.

Általánosan levonható következtetések:

viszonylagos teljességhez nem mellőzhető teljesen a kézi visszakeresés és a referátumok nyomán követése;

rendszeresen ellenőrizni kell a keresőprofilokat, az eredményeket és az irodalmi stílusváltozásokat a profilok naprakész állapotba hozatala és a hibák kiküszöbölése érdekében;

új keresőprofilokat nagy gonddal kell megírni és ehhez minden bizonnyal némileg ismerni kell a keresett terület szakirodalmát.

/PARRY, A.A. – LINFORD, R.G. – RICH, J.I.:

Computer literature searches – a comparison of the performance of two commercial systems in an interdisciplinary subject – The Information Scientist, 8. köt. 4. sz. 1974. p. 179–187./

(Keveyné Bundy Mária)

Dokumentumtároló és -kereső rendszerek jövője

Az Institute for Graphic Communication 1975 májusában Ipswichben (Mass.) rendezett konferenciája a szakirodalmi információ kezelésének leghatékonyabb módszereit vizsgálta. E tevékenységi körben 1973-ban az Egyesült Államokban 5,7 millió személy dolgozott. (81%-uk ipari magánvállalkozásokban), az összes ráfordítás 27 milliárd dollár körül volt. És bár az ipari tevékenység csaknem minden területen jelentősen korszerűsödött – elsősorban az automatizálás révén – az információs terület máig is alapvetően manuálisan tevékenykedik. Az információs igények rohamosan nőnek, s ha az új technika alkalmazása nem tör be, 1980-ban 19 millióra becsülhető az itt foglalkoztatandók száma. Pedig a fejlett technológia már rendelkezésre áll, s az ezt alkalmazó fejlett rendszerek is néhány éven belül hozzáférhetőek lesznek. Ezek figyelmen kívül hagyása igen súlyos hiba volna.

/Journal of Library Automation, 8. köt. 3. sz. 1975. p. 266./

(Sárdy Péter)

Válaszadó mondatok és válaszadó táblázatok visszakeresése teljes szövegből

A szakember a munkája során felmerülő problémákra választ a szakirodalomban keres. Ha a lehetséges forrásokat nem ismeri, információkereső szolgáltatást vehet igénybe, s így egy (esetleg referátumokkal kiegészített) irodalomjegyzéket kap. Ennek két nagy hátránya van: a kikeresett tételek közül sok nem releváns, de a dokumentumokról csak kézbevitellel állapítható meg, tartalmazzák-e a választ, és mi a válasz.

Az eredeti szövegekből történő „szövegrészlet vissza-keresés” során a bibliográfiai leírás mellett szerepel az a részlet (mondat, bekezdés stb.), ami miatt a dokumentumot kikerestük. Így az irreleváns tételek gyorsan kiejthetők, sőt az irodalomkutatás a választ is tartalmazza (ami természetesen nem feltétlenül teszi feleslegessé a teljes szöveget tartalmazó dokumentum kézbevitelét). Ezt az eljárást jogi dokumentumokra a 60-as évek közepe óta alkalmazzák.

A cikkben ismertetett kísérletben *válaszadó-szövegrész* alatt olyan szövegrész értendő, amely a választ vagy tartalmazza, vagy indikálja (azaz: valószínűsíti, hogy a dokumentum tartalmazza a választ). *Válaszadó-dokumentum az, amelyik tartalmaz válaszadó-szövegrészt.* A cél a válaszadó-dokumentumok lehető teljes körű kikeresése s ugyanakkor a tévesen kikeresett tételek minimalizálása volt. Az eljárás lényege a dokumentum nem indexelt, teljes szövegének keresése, folyamatos ember-gép párbeszéddel. Ez nélkülözi az indexeléssel járó hátrányokat és költségeket, növeli viszont a felmerülő gépi költségeket és az adatrögzítési munkát.

A kísérlet céljára mesterségesen alakították ki a gyógyszereknek a szív működésre gyakorolt hatásaira vonatkozó szövegyűjteményt, konkrét kérdésekből kiindulva, de lehetővé téve a téves visszakeresés lehetőségét is. A kísérleti időszakban 90%-os átlagos teljességet sikerült elérni, a téves kikeresések száma kérdésenként átlagosan 3 volt.

A kereső eljárás első és alapvető lépése manuális: a feltett kérdés alapvető szavaihoz *szinonimákat is tartalmazó kereső szójegyzék készül.* Így az azonos tartalmú, de eltérő formájú szavakat tartalmazó válaszadó-szövegrészek is kikereshetők. A keresésben szavak mellett kifejezések, ill. szótövek is szerepelhetnek. A kereső szójegyzék készítéséhez a legkülönbözőbb kézikönyveket vették igénybe. A teljességi hiányosságok fele a szójegyzék tökéletlenségére volt visszavezethető.

A gépi keresés eredményeként minden olyan szövegrész (mondat, cím táblázat, képaláírás stb.) kiírásra kerül, amelyben *legalább egy szó megegyezett a kereső szójegyzék valamelyik szavával.* A szövegrészek bizonyos rangsorban kerülnek kiírásra (cím, összefoglalásból származó mondat, tárgyszó, táblázat, a szöveg más részéből származó mondat). A keresés teljessége a fenti szövegrész-típusok egymás utáni bevonásával: csak a cím: 5%; összefoglalás is: 35%; tárgyszó a kísérletben nem szerepelt; táblázatban is: 80%; más szövegrész is: 90%. A teljességi hiányosságok másik felét az okozta, hogy a keresőszavak a cikkben több, egymástól sokszor messze eső mondatban helyezkedtek el.

Az igen jónak mondható eredményt több tényező indokolja: a biomedikai szaknyelv erősen normalizált; a teljes szöveg került keresésre (egy MEDLARS-kísérlet szerint a teljességi hiányok közel fele a hiányos indexelésnek tulajdonítható); bibliográfiai hivatkozások helyett szövegrészek kerültek kikeresésre és kiírásra.

A keresési kísérlet költségei három fő részből tevődtek össze: a szöveg számítógéppel olvasható formájú leírása, tárolása és keresése. E költségek jövőbeni csökkenése várható, ugyanis a szakfolyóiratok közül egyre több alkalmaz számítógépet, s így a teljes szövegek melléktermékként mágnesszalagon is előállnak. A szükséges nagykapacitású és gyors hozzáférésű memóriák fajlagos költségei is csökkennek. Pillanatnyilag olyan keresőrendszer látszik gazdaságosnak, amely címekre, referátumokra vagy összefoglalásokra és táblázatokra, képaláírásokra terjed ki.

Egészében megállapítható, hogy magas szintű szöveg-kereső rendszer ma már technikailag is, gazdaságilag is lehetséges.

/O'CONNOR, J.: Retrieval of answer-sentences and answer-figures from papers by text searching = Information Processing and Management, 11. köt. 5/6. sz. 1975. p. 155–164./

(Sárdy Péter)

Rank Xerox-9200-as másológép

Az 1975. évi Hannoveri Vásáron mutatták be a Xerox-9200 jelű nagy teljesítményű másológépet a reprográfiai kiállítás keretében. A berendezés azóta az NSZK öt nagy városában működik.

A Rank Xerox 9200-as készülék teljesítménye 7200 másolat óránként, azaz másodpercenként két másolat. Ez a teljesítmény figyelemre méltó, mert a legtöbb irodai ofset géppel is csak ritkán érhető ez el. *A készülék xerox rendszerű másolást végez, de a korábbi típusoknál sokkal gyorsabban, tökéletesebben és jobb minőségben.* Minden másolat az eredetiről készül, ezért a minőség mindig változatlan marad az első példánytól az utolsóig, függetlenül a példányszámtól.

A készüléket olyan integrált szortírozógéppel építették össze, melynek 50 rekesze két szintre osztott. Ha a 25 rekeszes alsó szint megtelik, automatikusan átkapcsol a felső szintre. Közben kiüríthető a megtelt alsó rész, így folyamatos, gyors munka végezhető, még 50 oldalnál nagyobb terjedelmű dokumentumok másolása esetén is.

A másolandó dokumentumokból egyszerre legfeljebb 50 db adagolható a készülékbe. *A gép A4 formátumot másol, de lehetőség van kisebbítésre is, így A/3-ról A/4-re vagy A/4-ről A/5 méretre.*

A papírtartóban 2500 db papír tárolható. Egy kisegítő tartóba 400 db papír helyezhető, egyrészt azért, hogy ha az első tartóból kifogy a papír, a folyamatos munka tovább biztosítható legyen. Másrészt, ha másolás közben más formájú vagy színű papírt kívánnak közbeiktatni, pl. fedőlap céljára, akkor egy gomb megnyomásával bekapcsolható a második adagoló.