

On-line szakirodalmi információkeresés az ORBIT rendszerrel

Az Egyesült Államokban a hatvanas évek elején a legnagyobb szakirodalmi információs cégek felismerték a számítógép nyújtotta lehetőségeket és az egyre hatalmasabb mennyiségű információt mágnesszalagon tárolták, kurrens és retrospektív információkeresés céljaira.

A hetvenes évek elejére elért technikai fejlődés, amelynek eredményeképpen elterjedté vált az időosztásos (time-sharing) működési mód és a mágneslemezes tárolás, nagyságrendekkel lerövidítette a háttértárolókon lévő adatok hozzáférési idejét. Ez a fejlődés csakhamar tükröződött az információkeresés technikájában is: megszülettek az on-line szakirodalmi figyelő rendszerek. Egy ilyen on-line kereső rendszer, az ORBIT (*System Development Corporation, USA*) működését mutatjuk be a *Chemical Abstracts Condensates* adatbázisán.

Ez az adatbázis a *Chemical Abstracts*-ban az utóbbi öt évben megjelent referátumok bibliográfiai adatait és kulcsszavait tartalmazza. A *kulcsszavas index* (Keyword Index) azonban nem olyan merev szabályok szerint készül, mint a nyomtatott kiadvány hagyományos *tárgy-indexe* (Subject Index). Az utóbbi csak korlátozott (ellenőrzött) szókinccsel és az IUPAC (*International Union of Pure and Applied Chemistry = Elméleti és Alkalmazott Kémia Nemzetközi Szövetsége*) kémiai nevezérendszernek megfelelően írja le a dokumentumokat, míg a kulcsszavas leírásban a közlemények eredeti kifejezései és vegyületek pozíciószám nélküli elnevezései találhatók. Ez sokkal kényelmesebb a kutató vegyész számára, de nagyobb feladat elé állította a számítógépi programrendszer kidolgozóit.

Tegyük fel, hogy a kutató a rovarirtó szerek (pesticide) vékonyréteg kromatográfiás (thin layer chromatography) vizsgálatával foglalkozik, és kíváncsi a tárgykör irodalmára. Hogyan használja az ORBIT on-line kereső rendszert? A laboratóriumban vagy az egyetem (intézmény) könyvtárában elhelyezett terminálon, amely a cég hálózatába (központi számítógépéhez) kapcsolt telegép (esetleg képernyővel felszerelve), a felhasználó beüti az ORBIT rendszer hívójelét. Válaszul az írógépen a következő üzenet jelenik meg:

SS 1/C?

USER:

ami azt jelenti, hogy a rendszer a felhasználó (USER) 1. utasítását várja. A felhasználó, aki már korábban elsajátította a rendszer sajátos „programnyelvét”, erre a következőket billentyűzi:

THIN AND LAYER AND ALL CHROMATOGRAPHY:

OR TLC

(keresd ki a vékony és réteg és kromatog. . . , vagy röviden a TLC betűkombinációt tartalmazó referátumokat).

A hamarosan érkező válasz:

PROG:

SS 1 PSTG (2586)

(azaz, az 1. kereső utasítást végrehajtottam, 2586 db megfelelő rekordot találtam). Ezután megjelenik a

SS 2/C?

USER:

üzenet, tehát a rendszer várja a 2. utasítást. A felhasználó a rovarirtó szerekről szóló közleményekre kíváncsi, ezért a következő utasítást adja:

1 AND ALL PESTICIDE#

amire a

PROG:

SS 2 PSTG (93)

válasz érkezik, azaz az ORBIT 93 ilyen dokumentumot talált a 2586-ban. További kulcsszavakkal a kör egyre szűkíthető, végül a felhasználó kiirattathatja az őt érdeklő közlemények adatait.

Mielőtt azonban erre példát látnánk, megmagyarázzuk az első utasításban szereplő kettőspont, a másodikban szereplő # jel, és a mindkettőben szereplő ALL szócscsa jelentőségét. A CHROMATOGRAPHY: hatására az ORBIT rendszer a rekordok szövegében – a kulcsszavas indexben – azokat a szavakat keresi, amelyek ezzel a kilenc betűvel kezdődnek. Az ennek megfelelő rekordokból a rendszerbe beépített korlátozás szerint azonban csak az első 400-at venné figyelembe, ezért tette a felhasználó az ALL (összes) szócskát a megfelelő helyre.

Hasonló szerepe van a második utasításban a # jelnek, ugyanis a PESTICIDE# utasításban a # jel bármilyen karaktert jelöl. Így a program nemcsak a „pesticide” hanem például a „pesticides” (az előző szó többes száma) szót is figyeli.

Már a fenti egyszerű példa is szemlélteti a rendszer rendkívüli rugalmasságát.

Természetesen nemcsak a kulcsszavak segítségével lehet kerestetni, hanem a szerző, rövidített folyóiratcím, a publikálás éve, szabadalom száma és tulajdonosa, a közlemény nyelve, társszerző stb. szerint is.

A következő példán bemutatjuk, *hogyan határolhatja körül a felhasználó a keresett közleményeket.*

SS 3/C?

USER:

ALL ISOTHIAZOL:

(A felhasználó az izothiazol. . . szórészt tartalmazó rekordokra kíváncsi)

PROG:

SS 3 PSTG (130)

USER:

3 AND NOT P (UC)

A felhasználó tehát a program által a SS 3-al jelölt, azaz a 3. utasítás szerint talált 130 db dokumentumból azokat válogattatja ki, amelyek nem szabadalmi leírások (P jelentése: szabadalom).

Az ORBIT válasza:

PROG:

SS 4 PSTG (80)

Nyolcvan illet talált.

SS 5/C?

USER:

3 AND GER (LA)

(Válassz ki a 3. utasításban találtak közül a német – GER – nyelven írottakat!)

Erre a válasz:

SS 5 PSTG (7)

hét illet talált a program. Végül a felhasználó a 3. utasítás szerint találtakból az 1973 után megjelent közleményeket kéri:

SS 6/C?

USER:

3 AND GREATER THAN 73

PROG:

SS 6 PSTG (30)

azaz 30 ilyen van. A keresés befejezése után a felhasználó a megtalált közlemények bibliográfiai adatait kéri. Erre szolgál a kinyomatást kérő utasítás. A

„PRT 5, TI, AU, SO, PY”

jelsorozat beütésére például az ORBIT kiírja az 5. utasításban talált 7 db német nyelvű dokumentum címét (TI), szerzőjét (AK), megjelenési helyét (SO) és idejét (PY).

Az ORBIT rendszer nyújtotta lehetőségek közül még a NEIGHBOR (szomszéd) utasítást mutatjuk be. A felhasználó által beütött

„NB ACRYL 4”

utasítás hatására az adatbázisban fellelhető, az ábécé sorrendben az acryl szóhoz legközelebb eső négy szomszéd íródik ki, azaz

PROG:

POSTINGS

TERMS

1

ACRYDANO/(IT)

2

ACRYIC/(IT)

29

ACRYL/(IT)

2

ACRYLACETATE/(IT)

8

ACRYLALDEHYDE/(IT)

UP N OR DOWN N ?

Ahol a baloldali számok az egyes szavak előfordulásainak számát, a zárójelben lévő rövidítés (IT) pedig a minőséget jelenti, azaz az acrydano szó tehát egyszer fordul elő az adatbázisban, mégpedig deskriptorként (Index Term). Az utolsó sorban az ORBIT azt kérdezi, hogy folytassa-e a további szomszédok kiírását felfelé (up) vagy lefelé (down) a betűrendben.

Nagyobb mennyiségű bibliográfiai adatot nem érdemes a terminálon közvetlenül kiírni, mert ez drága és esetleg kényelmetlen is. Ezért az ORBIT rendszer lehetővé teszi, hogy a felhasználó levélcímét is beütve, az eredményt számítógépi nyomtatásban postafordultával megkapja.

Több hasonló on-line keresőrendszer működik már az USA-ban, sokféle szakterület adatbázisán. Bár jelenleg az on-line irodalomkutatás még viszonylag drága, a prognózisok szerint ez a módszer az USA-ban egy évtizeden belül kiszorítja a hagyományos eljárásokat.

/BUCKLEY, J. S. jr.: *Planning for effective use of on-line systems*; PREWITT, B. G.: *Searching the Chemical Abstracts Condensates data base via two on-line systems = Journal of Chemical Information and Computer Sciences*, 15. köt. 3. sz. 1975. p. 161-164; p. 177-183./

(Valkó Péter)



Automatikus referátumkészítés

Sok kísérlet történt már eredeti dokumentumok automatikus referáltatására, de eddig még ezzel a módszerrel a hagyományos referálás színvonalát nem sikerült elérni. Nem is valószínű, hogy a közeljövőben erre sor kerülne. Ennek ellenére az ilyen irányú kutatások nagy erővel folynak, mert

a hagyományos referálás költséges és időigényes;

a közeljövőben egyre több publikáció lesz, már géppel olvasható formában is hozzáférhető, ezzel viszonylag olcsó és kényelmes adatbázis hozható létre a kísérletekhez;

a referálandó szövegek gépi olvashatósága folytán az automatikus referálás olcsóbb és gyorsabb lesz a hagyományos referálásnál;

az automatikus referátum gépi továbbfeldolgozásra alkalmas formában áll rendelkezésre;

bár az automatikusan készült referátumok esetleg soha nem érik el a hagyományos módon készülték színvonalát, gyakorlatban kielégíthetik az igényeket, különösen megfelelő utószerszeggel.

A most ismertető rendszer, amely a *Chemical Abstract Service (CAS)* számára készült, két szempontból különbözik az eddig használatos rendszerektől. Egyrészt egy viszonylag szűkebb terület, a kémia – ezen belül pedig a gyógyszerkémia – tárgykörére specializált, másrészt előírták, hogy a program által készített referátumoknak meg kell felelniük a CAS szabványoknak.

Az ADAM elnevezésű referálási módszer az eredeti dokumentum egyes kiragadott részeit illeszti össze úgy, hogy azok a dokumentum tartalmát reprezentálják, de nem egyszerűen mondatokat ragad ki az eredeti szövegből, hanem kissé át is alakítja azokat. Mindazonáltal az így készült referátumban nem lesznek teljesen új, illetve összevont mondatok, ellentétben a hagyományos referátummal.

A korábbi algoritmusok előfordulási statisztika szerint választották ki a szövegből a fontos mondatokat. Az