

Következtetés

Az ESA-IRS új árpólitikájának legfőbb előnye, hogy felszabadít bennünket az idő szorítása alól, sokkal interaktívabb keresésre ösztönöz. Minden gyakorlott kereső tudja, hogy a bonyolult kérdések megválaszolása csak iterációs módszerrel oldható meg. A gyakorlatlan végfelhasználók számára is előnyös az ESA-IRS árstruktúrája. Ők ugyanis még az egyszerűbb keresésben is lassúak, s az alacsony kapcsolati díj nem bünteti őket lassúságukért.

A használók számára az az árrendszer előnyös, amelyben előre könnyen ki lehet számítani a keresés teljes költségét. Mivel az ESA-IRS-nél a "terminálulási" és a találati díj az alapvető összetevő, ez könnyen becsülhető. Hagyományos árstruktúra esetén, mint amilyen a Dialogé, a használónak mind a kapcsolati időt, mind a találatok számát figyelembe kell venni a költségbecsléshez. A Data-Starnál még nehezebb a számítás, mert a távközlési díjakat az átvitt adatcsomagok száma is befolyásolja, míg az STN-nél néhány adatbázisban (pl. a CA-ban) még speciális díjak is vannak.

A hagyományos árrendszerű hostok különféle megoldásokat ajánlanak az online keresés árának optimalizálására (nagyobb kapcsolati sebesség, rövidebb output formátumok, offline nyomtatás, szelektív output). Az ár akkor optimális, ha minimalizálják a

kapcsolati időt, s szelektív outputtal csökkentjük a találati díjat. Az ESA-IRS-nél csak ez utóbbi árcsökkentő tényező, itt az alacsonyabb ár sokkal inkább a keresési technikákon és stratégiákon múlik, mint az output formátumokon. Az alacsony kapcsolati díjú rendszerekben gazdaságosabb többlet keresési idővel finomítani a keresést, s csak a releváns tételeket kinyomtatni (erre ad módot az ESA-IRS ZOOM lehetősége).

A nemzetközi adatbázisok gyakorlott használói akkor oktatták az online keresésre, amikor a legfőbb szempont a keresési idő csökkentése volt. Ettől megszabadulva egy sokkal interaktívabb – és sokszor jobb – használati mód alakítható ki. Ahogyan D. T. Hawkins az információ-visszakeresés árstruktúrájáról megállapítja (In search of ideal information pricing. = Online, 13. köt. 2. sz. 1989. p. 15–30.), az árak kialakításának arra kell törekedniük, hogy ne időegységeket és számítógép-teljesítményt adjanak el, hanem információk biteket és bájtokat.

/ESKOLA, P. – SORMUNEN, E.: Cost comparison of online searching in four hosts: Data-Star, Dialog, ESA-IRS and STN. = Online Review, 14. köt. 5. sz. 1990. p. 303–316./

(Murányi Péter)

Vegyületekkel kapcsolatos információk számítógéppel segített keresésének stratégiája

A vegyületek pontos kémiai adatai alapvető fontosságúak, ha fel akarjuk becsülni a környezetre gyakorolt veszélyes hatásukat. Az adatgyűjtés nehéz feladat. Azonkívül, hogy a kémia alapos ismerete szükséges, a keresőnek tisztában kell lennie a számítógépes információforrásokkal, a bibliográfiai, numerikus vagy faktografikus és a teljes szövegű adatbázisokkal.

A probléma definiálása

Először is tisztázni kell, vajon közvetlenül visszakereshető-e az információ, vagy több lépés szükséges. A kémia területén a legtöbb számítógépes adatbázisban a CAS nyilvántartási száma (Chemical Abstracts Service Number) nem kereshető mező, a keresőnek a kémiai nevet kell használnia az adat visszakeresésére. Ez a keresést bonyolultabbá és nehezebbé teszi.

1988-as adatok szerint 3000 online adatbázis létezett, ezek közül körülbelül 200 foglalkozik vegyületekkel.

A probléma megoldására korábban több mint 100, vegyületekről információt tartalmazó adatbank és adatbázis elérésére alkalmas rendszert fejlesztettek ki. Ezek az adatbázisok hat központon keresztül

érhetőek el – CIS, Datacentralen, Data-Star, DIMDI, ESA, STN –, és mindegyiknek saját speciális keresőnyelve van.

Ezenkívül minden adatbázisnak megvan a maga struktúrája és több más sajátossága is. Ahogy a több adatbázisban való párhuzamos keresés mindinkább elterjed, az ehhez kapcsolódó számítógéppel segített stratégia kialakításának szükségessége is nyilvánvalóvá válik. Két megközelítést fontolhatunk meg. Vagy egy parancsnyelvfordítót fejlesztünk ki, amely a keresési parancsokat lefordítja a különböző központok parancsnyelvére, vagy előre definiált keresések halmazát állítjuk fel, majd megfelelő stratégia segítségével az optimális keresési utat azonosítjuk. Itt ez utóbbi megközelítés érvényesül.

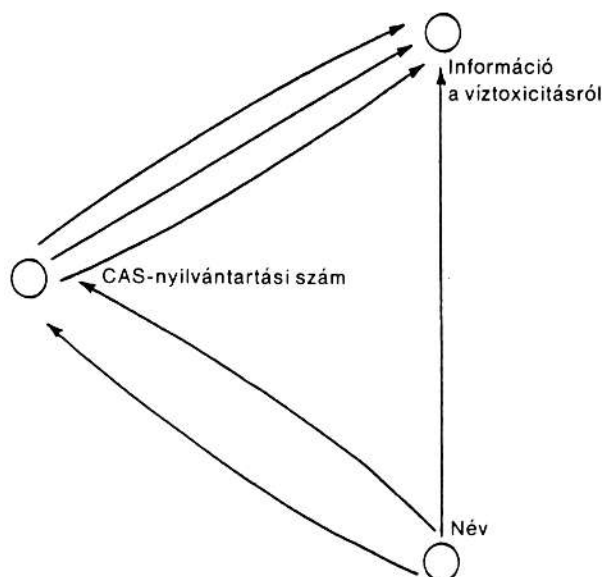
Nézzünk egy olyan példát, ahol meghatározott kémiai anyag vizitoxicitásáról kell információ, de csak a márkanev ismert. Az 1. ábra bemutat egy közvetlen keresést, amellyel kicsi a valószínűsége, hogy adatot találunk; mutat két alternatív keresési lépést, amellyel a CAS-nyilvántartási szám található meg; és három adatbázis kiválasztását, amelyek tartalmazzák az adatot.

Az ábra irányított gráfot mutat. A csúcsok azt reprezentálják, milyen jellegű információ szükséges

(név, CAS-nyilvántartási szám, információ a víztoxicitásról), az egyes élek pedig azt, milyen úton találunk információt (pl. online adatbázisokban). A gráfok kiindulópontjai a már meglevő információt mutatják, a nyílak vége a szükséges információt.

Főntos, hogy olyan rendszert alakítsunk ki, amely gyors, pénzkímélő, s a sikernek nagy a valószínűsége. Ennek a megközelítésnek a kialakítása két lépésből áll:

- ▶ a keresési gráf összeállítása,
- ▶ a költségek és a siker valószínűségének felbecsülése.



1. ábra A keresés sematikus vázlata

A keresési gráf kialakítása

A kereső lefolytathatja a keresést a kémiai név alapján, de vannak más elérhető kritériumok is, mint a nyilvántartási számok (pl. CAS-nyilvántartási szám, RTECS-szám, GUD-szám) és az összegképlet. Ezért szükséges az úgynevezett közbülső információ visszakeresésének a lehetőségét is meghatározni. Az

erre irányuló lépést addig kell folytatni, amíg új típusú információ már nem jelentkezik. A végső, irányított gráf több forrást fog tartalmazni, mint ami pillanatnyilag elérhető. A cikkben szereplő példában ezek az összegképlet és a vegyület neve. Mivel csak a név elérhető, ezért egy második lépésben redukálni kell a gráfot, így ez végül csak elérhető információforrásokat tartalmaz.

Optimális kereséshez olyan adatsorozatot kell összeállítani, amely tartalmazza az összes szükséges információt egy irányított gráf meghatározott éleihez. Egy optimális kereséshez kidolgozott ajánlásnak tartalmaznia kell az adatforrást, a szükséges információt és az adatbázis típusát is.

Az élek jellemzése

A gráf éleit úgy jellemezhetjük, hogy a siker valószínűségét és a költségeket kapcsoljuk hozzájuk.

A siker valószínűsége megbecsülhető kémiai anyagok mintája alapján. Egy, a vegyületek környezeti hatását vizsgáló bizottság 512 vegyületből álló listájából választották ki mintául azt a 68 anyagot, amely potenciális élelmiszer-szennyező.

A költséget külön kell vizsgálni sikeres és sikertelen keresés esetén. A teljes költség mindkét esetben a keresés és a talált információk megjelenítéséből tevődik össze. Nem minden adatbázisban lehet közvetlenül keresni egy adott fogalomra, néhol a magasabb rendű fogalmat használhatjuk csak, ilyenkor viszont nagyobb lesz az irreleváns információ aránya.

A várható költség meghatározásánál a következőkre kell figyelni:

- ▶ a találatok átlagos száma siker esetén,
- ▶ a találatok átlagos száma kudarc esetén,
- ▶ a keresés átlagköltsége,
- ▶ egy találat megjelenítésének átlagos költsége.

Ez alapján meghatározható, melyik adatbázisban érdemes közvetlenül a név alapján keresni, melyikben, ha a CAS-nyilvántartási számot akarjuk megtudni, s hol folytassuk a keresést, ha már megtudtuk ezt.

/BENZ, J.–VOIGT, K.–MÜCKE, W.: *Strategy for computer-aided searches for information about chemicals.* = *Online Review*, 13. köt. 5. sz. 1989. p. 383–394./

(Murányi Péter)

Mikrográfia, optikai lemez: nem VAGY, hanem ÉS

A dokumentumkép-feldolgozás szakemberei sokáig a mikrofilmzés, illetve az optikai lemezek híveire oszlottak, a nem szakmabeliek közül pedig sokan ingadoztak a két tábor között.

A felhasználók kétségei jogosak voltak. Mostanáig ugyanis a felvételező berendezés megvásárlásával tartósan lekötötték magukat az egyik vagy a másik technikához. Most azonban megvalósulóban van az a

kombinált berendezés, amely egyazon pozícióból párhuzamosan képes egyrészt letapogatni a dokumentumot mágneses vagy optikai lemezes tárolás céljára, másrészt mikrofilmfelvételt készíteni róla. A Kodak már kifejlesztett olyan letapogatót és mikrofilmfelvételt, amelyek azonos szállítóasztalról dolgoznak, és hamarosan elérhetők lesznek a kombinált berendezések is. Így megnyílik az út a kombinált tárolórendsze-