

Beszámolók ◦ Szemlék ◦ Közlemények

TÁJÉKOZTATÁSI INTÉZMÉNYEK ÉS SZOLGÁLTATÁSOK

Kémiai információkereső rendszer az Egyesült Államokban

Az egyesült államokbeli *Illinois Institute of Technology Research Institute Computer Search Center* (IITRI—CSC = Illinoisi Műszaki Kutatóintézet, Számítógépesítési Kutatói Központ) egy, a kereskedelemben megvásárolható mágnesszalagos adatbázisokon alapuló kémiai információkereső rendszert hozott létre és üzemeltet 1969 óta.

A központ létrehozásának célja a vegyészek és vegyészmérnökök számára a *gépi adatbázisok olyan kézi és gépi feldolgozási rendszereinek kialakítása*, amelyek valóban alkalmasak az ilyen típusú információforrások zökkenőmentes és magas színvonalú használatára.

A számítógépi program *PL/I* nyelven készült el, az *IBM 360/40* típusú és annál magasabb sorozatszámú gépekre. Legkevesebb 2 mágnesszalagos egységgel, 1 mágneslemez egységgel és legfeljebb 200 db 15 szavas keresőképpel azaz 3000 keresési ismérvvel számolva 256 K byte központi memória szükséges a rendszer működtetéséhez.

A CSC szolgáltatásai

Az *SDI* szolgáltatás jelenleg

- a Chemical Abstracts Condensates (CAC)
- a Biological Abstracts (BA)
- a Bioresearch Index (BIORI) és
- a COMPENDEX (EI)

adatbázisokon alapul.

A szelektív információterjesztési rendszer a felhasználók igényeit messzemenően figyelembe veszi, így többek között a profilokat kifejező keresőképek összeállításakor a Boole összefüggések minden korlátozás nélkül használhatók, a keresőkép egyes elemei súlyozhatók, az

eredményt a számítógép különféle módokon adhatja ki stb.

Az *elemek*, amelyek szerint információkeresést végezni lehet a következők:

- a dokumentum címében előforduló kifejezések;
- index-kifejezések;
- a szerző neve;
- a vállalat neve;
- CODEN kód, amely az eredeti dokumentumot tartalmazó folyóiratnak a kódja;
- a kiadó ország;
- Chemical Abstracts szekció szám;
- Biological Abstracts CROSS kód;
- Biological Abstracts BIOSYSTEMATIC kód;
- Engineering Index Card-A-Lert kód.

A keresőkép elemei lehetnek egy vagy több szavasak, mondatok, vagy szavak részei.

A felhasználó — kívánság szerint — a találatokat a szerzők betűrendje, a súlyszám, vagy a referencia sorszáma szerinti sorrendben kapja kézhez, egyenként 12,7x20,2 cm méretű cédulán vagy sornymotatón kihozott leporello papíron, illetve további reprodukcióra alkalmas nyomdatechnikai nyersanyag formájában.

A felhasználónak küldött cédulák közül az első (*1. ábra*) tartalmazza a profil sorszámát, a vizsgált adatbázis nevét, az éppen vizsgált mágnesszalag kibocsátási dátumát, a mágnesszalagon lévő rekordok számát, a keresőképnek megfelelő és kinyomtatott hivatkozások számát, valamint a program futtatásának dátumát. A hivatkozásokat tartalmazó többi cédula (*2. ábra*) a hivatkozás bibliográfiai adatain kívül tartalmazza a profil sorszámát, a vonatkozó index-kifejezéseket és azokat az elemeket, amelyek a találatot eredményezték és a találat súlyát.

A *SDI* szolgáltatás céljára az információkeresést olyan gyakorisággal végzik, ahogy a mágnesszalagok beérkeznek.

Retrospektív információkeresést manuális vagy számítógépes formában is vállal a CSC. Esetenként döntik el,

OCTOBER 13, 1971

PROFILE B1X019011A

BA PREVIEWS VOL. 52, NO. 19 WAS SEARCHED

ISSUE CONTAINED 5836 CITATIONS

HITS FOR THIS ISSUE: 4

NUMBER OF HITS PRINTED: 4

COMPUTER SEARCH CENTER
IIT RESEARCH INSTITUTE

COMPUTER SEAR

1. ábra Vezetőkártya

ABSTRACT NO. 106935 BA VOL. 52, NO. 19 PROFILE B1X019011A

STODORN JW, BROWN ES.

AUGMENTED TITLE: A ROBINSON LIGHT TRAP MODIFIED FOR SEGREGATING
SAMPLES AT PREDETERMINED TIME INTERVALS WITH NOTES ON THE EFFECT OF
MOON LIGHT ON THE PERIODICITY OF CATCHES OF INSECTS.

J APPL ECOL, VOL. 8, NO. 1, PP. 69-75, 1971, (ASTM CODEN: JAPEA)

CROSS INDEX: 01010-07003 07200-07504-07508*10604-64072-

BIOSYSTEMATIC INDEX: 07508 75300

PROFILE TERMS CAUSING HIT: INSECT LIGHT

WEIGHT FOR THIS CITATION: 0

COMPUTER SEARCH CENTER IIT RESEARCH INSTITUTE • 10 W. 35th ST. CHICAGO, ILL. 60616 • 312/225-9630

2. ábra Hivatkozásokat tartalmazó kártya

hogy egy-egy feladatot géppel, vagy kézzel – a nyomtatott referálólapokat felhasználva – érdemesebb elvégezni. A kézi visszakeresést bonyolultabb profilok esetében természetesen nem alkalmazzák.

A *Private Libraries System* elnevezésű programrendszer segítségével nyújtott szolgáltatás eredményeképpen bármilyen dokumentumgyűjteményből a megrendelő „testére szabott” géppel olvasható adattár áll elő. Ilyen dokumentumgyűjtemény része lehet például az SDI szolgáltatásban kapott hivatkozások halmaza is. A kívánásra létrehozott adattárakból természetesen csak a megrendelő kereshet vissza információkat. Számos segédprogramot dolgoztak ki az adattárak kiegészítésére, javítására is. Ily módon lehetőség van arra is, hogy az

SDI szolgáltatásnál kapott valamely hivatkozáshoz tartozó dokumentum elolvasása után a saját adattárba egyéni megjegyzések is bekerülhessenek.

A CSC megrendelésre vállalja hasonló *rendszerek működésének megszervezését*, így a software-t és a know-how-t, bárkinek rendelkezésére bocsátja. E szolgáltatás keretében

átadja a számítógépes programok teljes dokumentációját;

a megrendelő számítógépén megvalósítja a próbafuttatást;

betanítja a személyzetet, különös tekintettel a lehetséges hibaforrásokra;

kiképezi a profilszerkesztőket;

átadja a rendszer működtetésével szerzett statisztikai ismereteket, valamint a saját készítésű indexeket, listákat;

olyamatosan karbantartja és fejleszti a programrendszert, és

állandó konzultációt nyújt.

A keresőkép

A keresőkép a kérdést (a profilt) a vizsgálni kívánt adatbázis terminológiájának és a keresőrendszer konvencióinak felhasználásával fogalmazza meg. A CSC-ben a keresőképet egy 3 részből álló adatlapon kell megadni. A kifejezések között logikai kapcsolatokat is lehet definiálni. Ilyen összekapcsolt kifejezések lehetnek pl. a szinonimák, rokonértelmű szavak, illetve hierarchikus (szűkebb, ill. tágabb értelmű) szavak. A kifejezéseket súlyozva (0–9 közötti értékű súlyszámmal) a felhasználó közölheti, hogy melyiknek milyen jelentőséget tulajdonít. Ez finomabb logikai megfogalmazást és nagyobb visszakeresési lehetőséget eredményez. Arra is mód van, hogy a hivatkozások kinyomtatása a súlyozás szerinti csökkenő sorrendben történjék és az alsó súlyhatárérték megadásával a kis relevanciaérték hivatkozások kinyomtatását el lehessen hagyni.

A felhasznált adatbázisokban szerepel a dokumentumnak a szerző által megfogalmazott – tehát az információs központ által nem módosított – címe. A keresőképben ezért egy bizonyos fogalom minden lehetséges változatának szerepelnie kell ahhoz, hogy visszakereshető legyen. A CSC rendszerében ennek megkönnyítésére a szavak szógyökökre (kifejezés-törédekekre) való bontásának valamennyi formáját alkalmazzák. A felbontást „*”-gal jelölik. Pl. a *MYCIN feltüntetése a keresőképben azt eredményezi, hogy valamennyi (kb. 20–30) –MYCIN végződésű szó releváns lesz. Így pl.: ACTOMYCIN, ANTIMYCIN, BIOMYCIN, NEOMYCIN, STREPTOMYCIN stb.

A *MYCIN* feltüntetésével azok a kifejezések is relevánsak, amelyekben a „MYCIN” töredék a szó belsejében fordul elő.

A szavak célszerű felbontása azonban nem egyszerű dolog. Ezért, a felbontás elősegítésére készítik az ugynevezett KLIC (*Key-Letter-in-Context*) indexet. A KLIC index hasonló a permutált KWIC indexhez, de ebben csak a kifejezések valamennyi betűje szerinti permutációk szerepelnek. A KLIC index segítségével mutatható ki pl., hogy az *YBD töredék csupán a MOLYBDENUM szóban szerepel, ezért ennek megadása alkalmas a molibdén visszakeresésére.

Más a helyzet azonban pl. az RNA betűsorozattal, amely a ribonukleinsav világszerte ismert angol rövidítése, mert az *RNA* töredékre több mint 200 irreleváns válasz is lehetséges, például:

ALTERNATE (alternál)

BARNACLE (apró tengeri kagyló)

CARNATION (szegfű)

DIURNAL (mindennapos)

FINGERNAIL (köröm)

MATERNAL (anyai)

A CSC szakemberei bizonyos egyezményes felbontások alkalmazását javasolják.

Igy például a kémiai gyakorlatban igen gyakori „előállítás” kifejezés megadására (angolul: preparation) a PREPAR*, PREPARAT*, PREPARATION, PREPARATION* felbontások közül egyedül a PREPAR* felbontást célszerű alkalmazni.

Ugyanígy a „tisztítás” témakörben a PURIFIC* használatát javasolják a PURIFICAT*, PURIFICATION, PURIFY helyett.

A „szintézis” témakörben a SYNTH* felbontás a javasolt a SYNTH*, SYNTHES*, SYNTHESI*, SYNTHESIS helyett.

Nyilvánvalóvá válik az ilyen felbontások előnye, ha a 3. ábrát nézzük. Ezen egy angol nyelvű szótár részlete látható, amely az *analízis* (analysis) szócikk környezetét mutatja be. Az ábráról látható, hogy a különböző felbontások mely szavakat mutatják relevánsnak. A szavak értelmének vizsgálatából az angol nyelven tudó olvasó, tehát a keresőkép szerkesztője megállapítja, hogy a valóban releváns szavak a 11–26-ig terjedők, ezért az ANALY* felbontás a helyes.

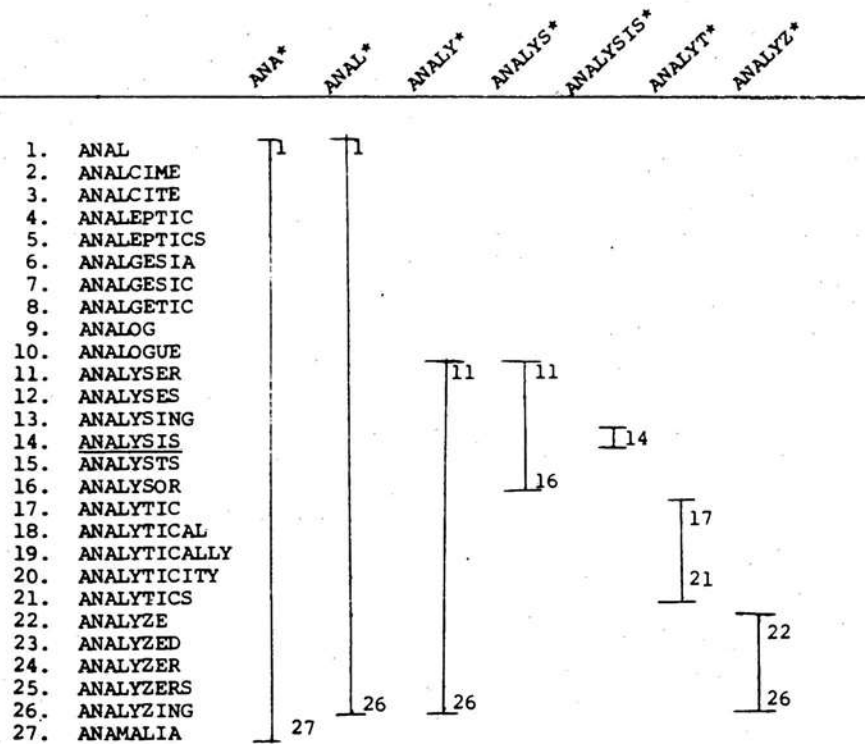
A keresőképek szisztematikus vizsgálata alapján arra következtetésre jutottak, hogy 10%-ukban jó eredménnyel lehet ezeket az egyezményes felbontásokat alkalmazni.

Bizonyos esetekben azonban az adatbázisok olyan rövidítéseket tartalmaznak, amelyek nem kereshetők vissza a szavak fenti típusú felbontásával. Pl. az ANALYSIS témakörben a *Chemical Abstracts* a 3. ábrán látható releváns szavaknak egy részét a mágnesszalagon ANAL. rövidítéssel adja meg, tehát a visszakereséshez az ANALY* és ANAL. kifejezések szerepeltetése egyaránt szükséges; logikai „vagy” relációval összekapcsolva.

A profilszerkesztésnek hatásos eszköze az adatbázisokban előforduló kifejezések betűrendes és az előfordulásuk gyakorisága szerinti indexe. Ezeket az adatbázisok 1–1 heti anyagából készítették el.

Az adatbázisban nagyon gyakran előforduló szavak túl sok találatot eredményeznek, hacsak nincsenek logikai „és” kapcsolatban más szavakkal összekapcsolva, vagy pedig kis értékekkel súlyozva. Ezért a keresőprofilokat futtatás előtt ellenőrzik, hogy nem tartalmaznak-e olyan szavakat, amelyek az 50 leggyakoribb közé tartoznak.

A betűrendes index segítségével a kifejezések gyakoriságát figyelembe véve nagyjából meg lehet becsülni, hogy a futtatni kívánt keresőprofil kb. hány találatot fog eredményezni.



3. ábra Az ANALYSIS kifejezés különböző felbontási lehetőségei

A keresés

A géppel olvasható adattárak visszakeresésének egyik leglényegesebb mozzanata a keresőkép és az adatbázis alkotó rekordokban található kifejezések betűhű összehasonlítása. A keresést egyidejűleg végzik az összes keresőkép szerint.

Az egyik eljárás szerint valamennyi keresőkép valamennyi kifejezését összehasonlítják valamennyi rekord valamennyi szavával, kifejezésével.

Egy másik módszer szerint a keresőkép kifejezéseit invertálják és a rekordok valamennyi szavát összehasonlítják az invertált kifejezésekkel. Az invertált listát ebben az esetben annyiszor kell ismételtlen összehasonlítani ahány rekordot megvizsgálunk.

A harmadik változatban a keresőképekben található kifejezéseket invertálva, egyenként hasonlítják össze a rekordokkal. Így a rekordok halmazát kell annyiszor végigvizsgálni, ahány különböző szó, kifejezés szerepel a listában.

A negyedik eljárás a rekordok tartalmának invertálásán alapul. Az így előállított listát azután összehasonlíthatják a keresőképekben található kifejezésekkel egyenként vagy esetleg a keresőképeket is invertálva.

A fenti módszerek fő problémáját a rekordok szövegek felbontása jelenti. Kézenfekvő lenne azt mondani, hogy a kifejezés bármely olyan betűcsoport, amelyet szóközök vesznek közre. Elválasztójelnek azonban nemcsak a szóköz tekinthető, hanem például az összes un-

nem alfanumerikus karakter (vessző, pont, kötőjel stb.) is.

Azonban a két véglet egyike sem fogadható el, mert csupán a szóközöket tekintve elválasztójelnek, a szavakat közvetlenül követő írásjelek is a szavak részeivé válnának; a második felfogás szerint pedig bizonyos szavakra jellemző speciális karakterek (kötőjelek, görög betűk, zárójelek stb.) így figyelmen kívül maradnának.

Az ilyen konvenciókon alapuló keresési rendszer gyakorlatilag lehetetlenné tenné megfelelő profilok szerkesztését. A nehézségek elkerülésére a keresőkép szavait olyan bármilyen jelekből összeállítható karaktorsorozat-ként lehet meghatározni, amelyet össze kell hasonlítani a rekord bizonyos részével vagy részeivel. Ilyen definíció mellett azonban a rekordokat nem szabad szavakra felbontani, invertálni, hanem egyetlen jelsorozatnak kell tekinteni.

Az ilyen feltételek mellett, betűjelek szerinti összehasonlítás esetében bármely változatban a rekordokat egyformán, egyenként hasonlítják össze az összes profillal.

Ez azt jelenti, hogy az adathordozó mágnesszalagot egyetlenegyszer, sorosan vizsgálják végig. A különböző metodikák összehasonlítására 5000 (egyenként 200 jelből álló) rekordot tartalmazó adatbázis 200, egyenként átlagosan 15 kifejezésből álló keresőképpel történő összehasonlítását végezték el. Ezek a mennyiségek fordulnak elő általában a CSC SDI célú műveleteinél.

Tanulságos a különböző algoritmusoknak a hatékonyság sorrendjében való áttekintése. Ez a sorrend egyébként megegyezik a kifejlesztés és általában az üzemszerű használat sorrendjével is.

1. A legegyszerűbb módszer szerint a keresőkép valamennyi szavát *betűről betűre összehasonlítják* valamennyi rekorddal. Természetesen a keresőképben található kifejezések típus szerint a rekordok más és más részeivel kerültek összehasonlításra, azaz a szerzők neve csak a szerzők nevét tartalmazó mezőkkel.

Igy a rekord címrészének vizsgálata során a keresőképnek a címre vonatkozó szavait a címek egészéhez hasonlítják. A program először az első szót „fekteti” rá a címre, annak első betűjétől kezdve, majd az egyezés ellenőrzése után a keresőszót egy betűvel arrébb „csúsztatja” és ismét megvizsgálja az egyezést. A műveletet a rekord címrészének utolsó betűjéig el kell végezni, majd újratekinteni a következő címtípusú keresőszóval. Ezt az egyszerű és direkt módszert könnyű programozni, azonban igen drága. Az összehasonlítási alapul vett esetben $3000 \times 5000 \times 200$, azaz 3×10^9 összehasonlítást kell elvégezni egy keresésnél. Ezt az algoritmust hátrányai miatt üzemszerűen sohasem alkalmazták.

2. A módszer hatékonysága jelentősen javítható, ha a program a rekord címen *betűnként lépeget és minden betűhöz hozzáilleszti azokat a keresőszavakat, amelyek az adott betűvel kezdődnek*. A keresőszavakat tehát kezdőbetűjük szerint csoportokba kell osztani és így a cím egy betűjénél csak egy csoportnyi keresőszó vizsgálatára van szükség. Mivel az adatbázisokban 50 jel fordul elő, a példának választott esetben a 3000 keresőszó 50 részre osztva 60-as csoportokat alkotna. Így a vizsgálat-hoz körülbelül $60 \times 5000 \times 200$, azaz 6×10^7 összehasonlítást kellene elvégezni. Mivel azonban a különböző jelek előfordulása a szavak elején nem egyforma valószínűségű, a csoportok átlagos szószáma a tapasztalat szerint mintegy 100 lesz, ami az előbbi számot 10^8 -ra emeli. Ez azonban még mindig csak 3,33%-a az 1. algoritmus szerint végzett keresésnél szükséges összehasonlítások számának. Programozási szempontból ez a módszer az előzőnél csak kissé bonyolultabb.

3. A keresőszavak csoportjai számának növelésével *tovább lehet csökkenteni az egy csoportba eső szavak, azaz a szükséges összehasonlító vizsgálatok számát*. Ha például a keresőszavakat első két betűjük szerint osztjuk csoportokba, elvileg 2500, egyenként átlagosan 1, 2 szót tartalmazó csoportot kellene kapnunk. Az eloszlás azonban itt még kevésbé egyenletes, mint az előbb.

A létrehozott csoportok közül 60-ba jutott az összes keresőszó fele és az átlagos csoportnagyság 20 volt. A kereséshez tehát 2×10^7 ($20 \times 5000 \times 200$) összehasonlítás szükséges, ami az 1. módszerhez képest már csupán 0,67%. Még ennél a változatnál is elfogadható a megnövekedett hatékonyságért cserébe elvégzendő programozási többlet.

4. *Ha a csoportok számát tovább növeljük és a keresőszavak első három betűje szerint alakítjuk ki azokat, akkor a szükséges összehasonlítások további jelentős csökkenése ($1,5 \times 10^6$ -ra) mellett a programozási munka már drágává és munkaigényessé válik*. Ez a számítások szerint csak akkor lenne gazdaságos, ha 50–100 ezer keresőszó egyidejű vizsgálatáról kellene gondoskodni. Ezért ezt a változatot a CSC-ben nem alkalmazzák.

5. A csoportok számának növelésén kívül más lehetőség is van a szükséges összehasonlítások számának csökkentésére. Ha ugyanis eltekintenek attól, hogy a keresőszavaknak csupán a két *első* betűjét vegyék figyelembe és *a szavak belsejében lévő kétbetűs részleteket is számításba veszik* akkor bebizonyítható, hogy az 50 csoportban sokkal jobb az eloszlás. Ez a kétbetűs részlet az ugynevezett legkisebb közös kétbetűs részlet (angolul: *Least Common Bigram*), az LCB, amely a teljes adatbázisban a legritkábban fordul elő a vizsgált szó kétbetűs részletei között. A kétbetűs részletek frekvenciáját a korábban ismertetett KLIC indexből lehet meghatározni. A MOLYBDENUM keresőszó kétbetűs részletei közül például a BD a legritkább, nem pedig a kezdeti MO. Ez a szó tehát az új algoritmus szerint a BD jelű csoportba sorolódik be, majd az összehasonlításokat a már ismertett módon lehet elvégezni. A módszer előnyét mutatja pl. a BOOKKEEPING keresőszó is. „KK”-val egyetlen angol szó sem kezdődik, tehát az LCB technika alkalmazása nélkül a „KK”-jelű csoportba nem jutna szó. Így viszont ez a szó oda kerül, csökkentve a BO csoportban lévő szavak számát. Az LCB alkalmazásával a csoportonkénti átlagos szószám 5-re csökken, ami a példának vett esetben $5,10^6$ összehasonlítást tesz szükségessé. Programozási szempontból a módszer alig jelent pluszmunkát a 3. módszerhez képest.

Ezzel a módszerrel dolgozva a keresőszavakhoz egy számpárost is kell rendelni. Az első szám megadja azt, hogy az LCB-t megelőzően hány jelet tartalmaz a szó, a második pedig a szó teljes hosszát jelenti. Ez a két kiegészítő információ elengedhetetlen az összehasonlításához.

Relevancia, az eljárás jósága

A SDI felhasználók saját maguk végzik a kapott referenciák relevancia szerinti értékelését. *Relevánsnak ugyanis csak az a relevancia tekinthető, ami a felhasználó számára az*. Az a tény, hogy egy hivatkozás releváns-e, vagy nem, az idő függvénye is, ezért az értékelést minden esetben a lehető legrövidebb időn belül kell elkészíteni. A visszajelzéseket a CSC felhasználónként visszamenőleg is nyilvántartja és amennyiben érdemesnek látszik, javaslatot tesz a keresőkép megváltoztatására. A keresés pontosságát (a releváns hivatkozások számának az összes

találathoz viszonyított százalékos arányát) tekintve mind az alacsony mind pedig a magas érték általában problémát jelez. 90% felett valószínű, hogy a túlságosan bonyolult logikai feltételek, illetve a túlságosan specifikus keresőszavak használata miatt sok, egyébként a profil témakörébe eső hivatkozást nem sikerült megtalálni. 25% alatt pedig túl sok fölösleges hivatkozást kap a felhasználó, valószínűleg a gyakori nagy valószínűségű, nem deskriptív keresőszavak alkalmazása miatt. Meg kell jegyezni, azonban, hogy a 0%-os eredmény nem jelent feltétlenül rosszat, csupán esetleg annyit, hogy a keresett témakörben nincs publikáció az adatbázisban.

Előfizetési díj

Az alap előfizetési díj alatt egyetlen maximálisan 25 keresőszóból álló keresőkép egy évig való futtatása értendő, amelynek során futtatásonként átlagosan 50

hivatkozás (évente átlagolva) nyomtatható ki külön térítés nélkül. Minden további 1–10 keresőszó alkalmazásáért vagy minden további 1–50 kinyomtatott hivatkozásért pótdíjat kell fizetni. Ezek ára a következő:

	Előfizetési díj dollárban	
	Alapdíj	Pótdíj
CAC csak páros vagy csak páratlan héten	165	60
CAC mindkét sorozat	250	100
BIORI	130	45
BA	200	75
BIORI • BA	250	100
EI	200	25–75

Four-year summary, educational and commercial utilization of a chemical information center, Part I.–II. The Larc Reports, 7. köt. 1–2. sz. 1974./

(Lázár József)

Chemical Abstracts Service információs szolgálata

Information Tools 1976. 1. Descriptions, Products and Services. Chemical Abstracts Service, 1975. Ohio State University, 24 p.

Information Tools 1976. 2. Specifications, Price and Ordering. Chemical Abstracts Service, 1975. Ohio State University, 40 p.

1876. április 6-án 35 kémikus a New York College gyógyszerészeti előadótermében megalapította az *Amerikai Kémiai Társaságot* (*American Chemical Society*). A tudományos társaságnak ma már több mint 110 000 tagja van. Évente 1800 helyi és nemzetközi rendezvényt szerveznek, s a kémiai kutatásokat évi 3,5 millió dollárral támogatják. A társaság megalakulása kezdetétől megkülönböztetett figyelmet fordított a kémiai szakirodalom publikálására és később feltárására is. A „publications program” keretében még 1907-ben évi 12 000 referátum közlésével megindult a *Chemical Abstracts (CA)* nyomdai előállítású kiadása, s ennek az „irodalom-centrikus” hagyománynak köszönhető az is, hogy a *Chemical Abstracts Service Source Index (CASSI)* megszervezésekor a kémiai irodalomról retrospektív mutató készült 1830-ig visszamenően, természetesen a másolat szolgáltatásának lehetőségével.

A *Chemical Abstracts Service (CAS)* jelenlegi (1975. évi) rendszerét a 176. oldalon lévő ábra mutatja.

A CAS adatbázisában az alábbi dokumentumokat dolgozzák fel:

14 000 féle tudományos és műszaki folyóiratot, 150 országból;

a kémia és vegyipar területét érintő szabadalmi leírásokat 26 országból; ezek közül Ausztria, Dél-Afrika,

Franciaország, Japán, Kanada, NSZK és az Egyesült Államok szabadalmi leírásait teljességgel, a többi 18 ország – köztük Magyarország – szabadalmi leírásait válogatva dolgozzák fel;

konferenciák, kongresszusok anyagait; kutatási jelentéseket; disszertációkat; jelentősebb kémiai könyveket bármely országból.

A CA teljes kiadása az 1976. évi tervek szerint összesen 454 000 dokumentumot fog feldolgozni.

A teljes adatbázis alapján jelenik meg a *Chemical Titles (CT)*, valamint a *Chemical Abstracts Service Source Index (CASSI)*, nyomtatott formában és mágnesszalag adathordozón is.

A kémia tárgyköreit, a közel 2 millió adatot – fogalom, molekuláris név és képlet, kémiai szerkezeteket, formákat stb. – tartalmazó gyűjtemény a *Registry Handbook*-ban, illetve annak folyamatos kiegészítéseiben jelenik meg.

A teljes adatbázis átvizsgálásával, de csak válogatott információkkal, hetente jelenik meg a tartalmi referátumokat közlő *Chemical Industry Notes*, vezetők, piackutatók, beruházó szakemberek részére.

Kétségtelen, a legnagyobb jelentőségű szolgáltatás a nyomtatott mágnesszalagos, valamint mikrofilmes formában is megjelenő *Chemical Abstracts*. A feldolgozott dokumentumok mennyiségét az alábbi összeállítás szemlélteti:

1907–1916 között	190 ezer dokumentum
1917–1926 között	220 ezer dokumentum
1927–1936 között	540 ezer dokumentum
1937–1946 között	510 ezer dokumentum
1947–1956 között	660 ezer dokumentum
1957–1961 között	630 ezer dokumentum
1962–1966 között	1 020 ezer dokumentum
1967–1971 között	1 470 ezer dokumentum