

Az Amerikai Kémiai Társaság folyóiratainak teljes szövegét tartalmazó adatbázis online keresése

1980–1982 között az Amerikai Kémiai Társaság (American Chemical Society = ACS) könyv- és folyóiratrésze a BRS (Bibliographic Retrieval Services, Inc.) online szolgáltató központon keresztül kísérleti céllal online hozzáférést biztosított több száz önkéntes jelentkezőnek az ACS folyóiratainak teljes szövegét tartalmazó adatbázishoz.

A kísérlet célja az adatbázis alkalmazhatóságának és életképességének a megállapítása volt. A kedvező fogadtatás alapján 1983 júniusában a BRS-nél nyilvánosan is hozzáférhetővé tették az ACS 18 folyóiratának teljes szövegét 1980-tól kezdődően tartalmazó adatbázist. Jóllehet az elmúlt néhány évben egyéb kiadványok teljes szövegét is elérhetővé tették már online (pl. a Mead Data Central *Lexis* és *Nexis* adatbázisai és a John Wiley and Sons: *Encyclopedia of Chemical Technology* című kiadványa), az ACS az első tudományos társaság, amelyik ebben a formában teszi hozzáférhetővé kiadványait.

A kísérleti adatbázis jellemzői

Az ACS folyóiratok számítógépes szerkesztése során rutinszerűen képződő, a folyóiratok valamennyi szöveges elemét (cím, szerző, bekezdések, lábjegyzetek stb.) tartalmazó mágnesszalagok képezik az adatbázis alapját. A szöveg egyes elemeit valamennyi folyóirat esetén azonos adatmezőkben tárolják függetlenül attól, hogy milyen elrendezésben jelennek meg végül is a nyomtatott kiadványban. Ez az egységes kezelés megkönnyítette a BRS munkáját, amikor az adatbázis létrehozásakor kialakította a kereshető és kinyomtatható adatmezőket. Az adatokat már a folyóirat-szerkesztés befejezését követő héten beviszik az adatbázisba.

A szövegszerkesztés során létrejövő adatállományban információ található a később kézzel behelyezésre kerülő ábrák méretére és helyére a folyóiratban. Ez az információ az online adatbázisból kimarad, viszont az ábraalírások benne maradnak, kereshetők és kiírathatók.

A kísérlet utolsó szakaszában (1982. augusztus–1982. október) az adatbázisban található dokumentumok számát és a lefedett éveket az 1. táblázat mutatja. Az adatbázis adatmezőit a 2. táblázat foglalja össze.

1. táblázat

Primer folyóiratok kísérleti adatbázisa

Folyóirat neve	Dokumentumok száma	Évek ^a
<i>Journal of the American Chemical Society</i>	5017	1980–1982 ^b
<i>Journal of Organic Chemistry</i>	3774	1980–1982
<i>Journal of Medicinal Chemistry</i>	2981	1976–1982
<i>Biochemistry</i>	2851	1980–1982
<i>Inorganic Chemistry</i>	2643	1980–1982
<i>Journal of Physical Chemistry</i>	2489	1980–1982
<i>Analytical Chemistry</i>	1960	1980–1982
<i>Macromolecules</i>	986	1980–1982
<i>Journal of Agricultural and Food Chemistry</i>	967	1980–1982
<i>Journal of Chemical and Engineering Data</i>	413	1980–1982
<i>Industrial & Engineering Chemistry Process Design and Development</i>	357	1980–1982
<i>Industrial & Engineering Chemistry Product Research and Development</i>	345	1980–1982
<i>Organometallics</i>	398	1982
<i>Industrial & Engineering Chemistry Fundamentals</i>	256	1980–1982
<i>Journal of Chemical Information and Computer Science</i>	188	1980–1982
<i>Accounts of Chemical Research</i>	175	1980–1982
<i>Environmental Science and Technology</i>	148	1982
<i>Chemical Reviews</i>	53	1980–1982
Összesen	25901	

a) 1982 november 19-ig.

b) A *Journal of the American Chemical Society* 1980 júliusával kezdődően szerepel az adatbázisban.

2. táblázat

Online kereshető és kiíratható adatmezők

Az adatmező jele	Az adatmező tartalma
AN	Az egyes dokumentumokat azonosító szám (kéziratszám)
CD	A folyóiratot azonosító kód (CODEN)
SO	Bibliográfiai adatok
TI	Cím
AU	Szerző(k)
AF	A szerző(k) munkahelye
AB	Referátum
TX	A cikk teljes szövege bekezdésként; az egyes bekezdések külön TX mezőben található; a TX(1) tartalmazza az első bekezdést, a TX(2) a másodikat stb.
FN	Számozott lábjegyzetek; az egyes lábjegyzetek külön FN mezőben található: FN(1), FN(2) stb.
RF	Nem számozott hivatkozások; minden egyes hivatkozás külön RF mezőben található: RF(1), RF(2) stb.
CP	Ábrafeliratok; minden egyes ábrafelirat külön CP mezőben található: CP(1), CP(2) stb.
RN	CAS Registry-számok; az egyes Registry-számok külön RN mezőben található: RN(1), RN(2) stb.; minden egyes Registry-szám mellett szerepel a neki megfelelő vegyület neve vagy a szövegben található a vegyületre utaló szám.

A teljes szövegű adatbázis keresése

A bemutatott példák a BRS rendszernek a teljes szövegű adatbázisok keresésénél különösen fontos keresési sajátosságait szemléltetik.

Boole-operátorok. A teljes szövegű adatbázisok keresésénél különösen fontos szerepük van azoknak az operátoroknak, amelyek a keresőszavak egymáshoz viszonyított viszonylag közeli előfordulását (ugyanabban a bekezdésben vagy még közelebb legyenek egymáshoz) kívánják meg. A 3. táblázat mutatja be a BRS rendszerben használható Boole-operátorokat és funkciójukat.

3. táblázat

Boole-operátorok*

Boole-operátor	Funkció (a keresőszavak egymáshoz viszonyított helyzete)
AND	Ugyanabban a dokumentumban legyenek
OR	Ugyanabban a dokumentumban legyenek
NOT	Ne legyenek ugyanabban a dokumentumban
SAME	Ugyanabban a bekezdésben vagy mezőben legyenek
WITH	Ugyanabban a mondatban legyenek
ADJ	Egymás mellett legyenek az adott sorrendben

Minél inkább korlátozó jellegű operátort alkalmazunk, annál nagyobb a valószínűsége, hogy az összekapcsolt keresőszavak tartalmi kapcsolatban állnak egymással. Kevesebb találatot kapunk, de ezek nagy valószínűséggel relevánsak lesznek. Az 1. ábra a különböző operátorok alkalmazását mutatja be egy példán.

A releváns információ megtalálása az előhívott dokumentumon belül. A BRS rendszer két lehetőséget is biztosít az előhívott teljes dokumentum (a cikk teljes szövege) azon részleteinek a kiíratására, amelyek a találatot behozó keresőszavakat tartalmazzák. Ezek: az "Occurrence Field" (előfordulási mező) és a "HITS" (találatok) nyomtatási paraméter.

Az Occurrence Field (OC) mező. Ez a mező a keresés során képződik és a találatokért felelős keresőszavak pontos helyét rögzíti a teljes dokumentumon belül. Az OC mező segítségével kiíratathatók a dokumentum azon részei, amelyek a kívánt keresőszavakat tartalmazzák. Ennek alapján eldönthető, hogy a cikk érdekes-e a kereső számára, és hogy

érdemes-e egyéb részleteit vagy a teljes dokumentumot kiíratni. Az OC mező hiányában a kereső kénytelen lenne a teljes dokumentum kiíratásának és átvezetésének időigényes és költséges útját választani, hogy megtalálja a releváns információt.

1-:	nucleophilic and displacement and fluorine	RESULT	92
2-:	nucleophilic same displacement same fluorine	RESULT	14
3-:	nucleophilic with displacement with fluorine	RESULT	5
4-:	nucleophilic adj displacement adj fluorine	RESULT	3

1. ábra Különböző szintű operátorokkal való keresés

A 2. ábra mutat egy tipikus OC mezőt, amely minden egyes, a találatot behozó keresőszóra megadja, hogy melyik adatmező hányadik mondatának hányadik szavaként fordul elő. Az OC mező megjelenítése után, az ott jelzett bármelyik mező vagy mezők kombinációja kiíratható.

OC Adatmező	Mondat	Nem tiltott szó*
TI (1)	1	11
AB (1)	1	25
TX (1)	2	7
TX (2)	2	2
TX (8)	6	23
TX (18)	3	5
TX (44)	4	16
FN (2)	1	23
FN (8)	1	10

2. ábra Előfordulási mező

A következő keresési példa illusztrálja az OC mező használatát.

3- : (thermodynamic adj propert\$) with argon
RESULT 3

A keresőkérdés, amely azt követeli meg, hogy a "thermodynamic" és a csonkolt "propert\$" szavak egymás mellett és ugyanakkor az "argon" szóval egy mondaton belül legyenek, három találatot eredményezett.

* Általában csak az AND, OR és NOT operátorokat tekintik igazi Boole-operátoroknak, míg a SAME, WITH és ADJ operátorokat helyzeti vagy kapcsolati operátoroknak hívják. Ez a cikk valamennyit Boole-operátornak tekinti.

* A nem tiltott szavak a szöveg azon szavai, amelyek nem szerepelnek a BRS tiltott szavak (stop words) listáján (triviális szavak, mint az of, the, and stb.).

A következő parancs hatására az első dokumentum SO, TI és OC mezője kerül kiírásra

4- : ..browse 3 so,ti,oc/doc = 1

1
SO ACCOUNTS OF CHEMICAL RESEARCH, VOL. 013, NO. 8, 1980, P 290-296.
TI A VAN DER WAALS PICTURE OF THE ISOTROPIC-NEMATIC LIQUID CRYSTAL PHASE
TRANSITION.

OC PARAGRAPH	SENTENCE	NS-WORD
TX (9)	3	13
TX (9)	3	33

END OF DOCUMENT

A találatokért felelős mezők a 9. bekezdésben található.

Egy újabb paranccsal kérhető ezután ennek a bekezdésnek a kiírása:

..browse 3 tx(9)/doc = 1

A 9. bekezdést tartalmazó kiírásban aláhúzásra kerültek a találatot eredményező szavak.

1
TX PARAGRAPH 9 OF 36. THE VAN DER WAALS PICTURE. FOR EXAMPLE, WHEN PHS IS TAKEN FROM THE COMPUTER STUDIES OF ALDER AND WAINWRIGHT, 17

CALCULATIONS OF SEVERAL DIMENSIONLESS THERMODYNAMIC PROPERTIES, E.G. THE RATIO OF LIQUID TO SOLID VOLUME OR THE MOLECULAR ENTROPY OF FUSION IN UNITS OF BOLTZMANN'S CONSTANT, HAVE BEEN SHOWN TO AGREE VERY CLOSELY WITH EXPERIMENTAL DATA ON ARGON NEAR ITS TRIPLE POINT. FURTHERMORE, EQ 3 HAS BEEN DERIVED FROM "FIRST PRINCIPLES" BY KAC, UHLENBECK, AND HEMMER, 18 WHO STUDIED THE STATISTICAL MECHANICS OF SYSTEMS INTERACTING VIA PAIR POTENTIALS WHICH CAN BE DECOMPOSED INTO A HARD-CORE REPULSION PLUS AN ATTRACTION HAVING MAGNITUDE +APPRX+GAMMA+ AND RANGE +APPRX+1/+GAMMA+. THEY SHOWED THAT EQ 3 IS EXACT IN THE LIMIT +GAMMA+ +FDARW+ 0.

A "HITS" nyomtatási paraméter. A nyomtatási parancsban a "HITS" paraméter megadásával egy lépésben, az OC mező külön megjelenítése nélkül kiírathatók a dokumentum azon részei (egyéb mezőkkel kombinálva), amelyek a találatért felelős keresőszavakat tartalmazzák.

Az OC mező ismertetésénél bemutatott keresési példát véve alapul, a következő paranccsal került sor a SO, TI, AU, AF és a találatokért felelős mezők kiírására az 1-es és a 2-es dokumentumból:

4- : ..browse 3 so,ti,au,af,hits/doc = 1-2

A kiírás eredménye:

SO ACCOUNTS OF CHEMICAL RESEARCH, VOL. 013, NO. 8, 1980, P 290-296.
TI A VAN DER WAALS PICTURE OF THE ISOTROPIC-NEMATIC LIQUID CRYSTAL PHASE TRANSITION.

AU (1) GELBART, W. M. (2) BARBOY, B.
AF (1,2) DEPARTMENT OF CHEMISTRY, UNIVERSITY OF CALIFORNIA, LOS ANGELES, LOS ANGELES, CALIFORNIA 90024.

TX PARAGRAPH 9 OF 36. THE VAN DER WAALS PICTURE. FOR EXAMPLE, WHEN PHS IS TAKEN FROM THE COMPUTER STUDIES OF ALDER AND WAINWRIGHT, 17
CALCULATIONS OF SEVERAL DIMENSIONLESS THERMODYNAMIC PROPERTIES, E.G. THE RATIO OF LIQUID TO SOLID VOLUME OR THE MOLECULAR ENTROPY OF FUSION IN UNITS OF BOLTZMANN'S CONSTANT, HAVE BEEN SHOWN TO AGREE VERY CLOSELY WITH EXPERIMENTAL DATA ON ARGON NEAR ITS TRIPLE POINT. FURTHERMORE, EQ 3 HAS BEEN DERIVED FROM "FIRST PRINCIPLES" BY KAC, UHLENBECK, AND HEMMER, 18 WHO STUDIED THE STATISTICAL MECHANICS OF SYSTEMS INTERACTING VIA PAIR POTENTIALS WHICH CAN BE DECOMPOSED INTO A HARD-CORE REPULSION PLUS AN ATTRACTION HAVING MAGNITUDE +APPRX+GAMMA+ AND RANGE +APPRX+1/+GAMMA+. THEY SHOWED THAT EQ 3 IS EXACT IN THE LIMIT +GAMMA+ +FDARW+ 0.

2
SO THE JOURNAL OF PHYSICAL CHEMISTRY, VOL. 086, NO. 9, 1982, P 1722-1729.
TI THERMODYNAMIC PROPERTIES OF LIQUID MIXTURES OF ARGON + KRYPTON.
AU (1) BARREIROS, S. F. (2) CALADO, J. C. (3) CLANCY, P. (4) PONTE, M. N. (5) STREETT, W. B.
AF (1,2,3,4,5) CENTRO DE QUIMICA ESTRUTURAL, COMPLEXO I, 1096 LISBOA, PORTUGAL.

Online tallózás a dokumentumban

Egy dokumentum valamely részének kiírása után a BROWSE parancs egy egyszerűsített formájával megjeleníthetők a dokumentum egyéb, tetszőleges részletei, vagyis online "tallózhatunk" a dokumentumon belül. Ennek illusztrálására nézzük a következő keresési példát:

1- : (caffeine or coffee or tea) with (toxic\$ or damage or adverse)

RESULT 3

A

2- : ..browse 1 ti,so,oc/doc = 1-3

paranccsal kiírásra került mindhárom dokumentum TI, SO és OC mezője:

1
SO JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY, VOL. 028, NO. 1, 1980, P 22-25.

TI A COMPOSITIONAL STUDY OF PODS OF TWO VARIETIES OF MESQUITE (PROSOPIS GLANDULOSA, P. VELUTINA).

OC PARAGRAPH	SENTENCE	NS-WORD
TX (6)	3	12
TX (6)	3	28

2
SO JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY, VOL. 028, NO. 5, 1980, P 960-963.

TI MYCOTOXIN PRODUCTION BY ALTERNARIA SPECIES GROWN ON APPLES, TOMATOES, AND BLUEBERRIES.

OC PARAGRAPH	SENTENCE	NS-WORD
TX (4)	2	5
TX (4)	2	12
TX (4)	2	33
TX (18)	3	1
TX (18)	3	5
TX (18)	3	14
TX (18)	3	15
TX (18)	5	4
TX (18)	5	8
TX (18)	5	10

3
SO BIOCHEMISTRY, VOL. 021, NO. 9, 1982, P 2036-2048.
TI ANALYSIS OF AN ALLOSTERIC BINDING SITE: THE NUCLEOSIDE INHIBITOR SITE OF PHOSPHORYLASE A.

OC PARAGRAPH	SENTENCE	NS-WORD
TX (7)	5	3
TX (7)	5	10

A második dokumentum OC mezője szerint a dokumentum 4. és 18. bekezdésében is előfordulnak a találatot eredményező keresőszavak. A 4. bekezdés a következő paranccsal írható ki:

3- : ..browse 1 tx(4)/doc = 2

2
TX PARAGRAPH 4 OF 23. THE ALTERNARIA METABOLITES WITH DEMONSTRATED MAMMALIAN TOXICITY BELONG TO THREE CLASSES OF COMPOUNDS: TENUAZONIC ACID (TEA), A TETRAMIC ACID; DIBENZO-ALPHA-PYRONE INCLUDING ALTERNARIOL (AOH), ALTERNARIOL MONOMETHYL ETHER (AME), AND ALTENUENE (ALT); ALTERNARIOL I AND II (ATX-I AND -II), WHICH ARE TOXIC SUBSTANCES OF UNKNOWN STRUCTURE. ATX-II IS THOUGHT TO BE A DEHYDRO FORM OF ATX-I. TOXICITY TO MICE FROM SPECIFIC AMOUNTS OF THESE COMPOUNDS HAS BEEN ESTABLISHED (PERO ET AL. 1973). ONYALAI, A COMMON HEMATOLOGIC DISORDER AMONG AFRICAN BLACKS, IS CAUSED BY SALTS OF TEA FROM MOLD CONTAMINATION OF GRAIN (STEYN AND RABIE, 1976).

Ezt követően az adott dokumentum egyéb adatmezői az egyszerűsített BROWSE paranccsal is kiírathatók. Például:

a *TX(18) parancs hatására megjeleníthető a 18. bekezdés:

PARAGRAPH 18 OF 23. RESULTS AND DISCUSSION. TEA IS USUALLY CONSIDERED THE MOST IMPORTANT TOXIC MATERIAL PRODUCED BY ALTERNARIA BASED ON THE POSITIVE CORRELATION BETWEEN IN VITRO PRODUCTION OF TEA AND THE TOXICITY OF CULTURES TO LABORATORY ANIMALS (HEPUNUCK ET AL. 1972). ANOTHER STUDY OF THE RELATIVE TOXICITIES OF THE ALTERNARIA TOXINS INDICATED THAT TEA WAS SOMEWHAT MORE TOXIC THAN THE OTHER COMPOUNDS (PERO ET AL. 1973). THE IMPORTANCE OF TEA PRODUCED UNDER NATURAL CONDITIONS HAS NEVER BEEN ESTABLISHED, SINCE THE ONLY REPORTED NATURAL OCCURRENCES OF TEA HAVE BEEN IN RICE PLANT LEAVES AND TOBACCO AND TRACE AMOUNTS IN TOMATO PASTE (SCOTT AND KANHERE, 1979).

