

## STO LET CHEMICAL ABSTRACTS – STO LET HISTORIE CHEMIE, 1907–2007

JAROSLAV ŠILHÁNEK

*Ústav organické technologie, Vysoká škola chemicko-technologická v Praze, Technická 5, 166 28 Praha 6  
silhanek@vscht.cz*

Klíčová slova: Chemical Abstracts, chemická informatika, chemické báze dat, historie chemie, chemická literatura

**Obsah**

1. Úvod
2. Jak to začalo aneb stručná historie „firmy“
3. Několik zastavení na stoleté cestě
4. Nástup počítačových technologií
5. Registrační systém CAS
6. Chemical Abstracts v našich zemích
7. SciFinder®
8. Jaké jsou současné možnosti práce s materiály Chemical Abstracts
9. A co dál?

**1. Úvod**

Kdo navštíví Ústřední knihovnu VŠCHT v Praze, najde v jedné, nepříliš velké místnosti kompletní soubor referátového časopisu Chemical Abstracts od 1. svazku z r. 1907. A všem návštěvníkům je zdůrazňováno, že zde jsou na regálech velmi stručnou formou sumarizovány prakticky všechny znalosti a poznatky ze všech chemických disciplín za celé 20. století. Návštěvník je pak buď zklamán, že toho zase až tak moc není nebo je šokován, jak to máme pěkně pohromadě a ještě přehledně seřazené. Žádná jiná vědecká disciplína nemá podobný nástroj, který jí umožňuje snadno a hlavně s vysokou spolehlivostí zjistit, co je známo a co ještě ne, a to hluboko do minulosti. A dnes můžeme ještě dodat, že v plně digitalizované verzi tohoto zdroje také neuvěřitelně rychle a spolehlivě. Jestliže tento nástroj už funguje plných sto let, určitě to stojí za připomenutí a alespoň stručnou rekapitulaci i vysvětlení, jak to funguje v současném světě digitálních informací.

**2. Jak to začalo aneb stručná historie „firmy“**

Sumarizovat výsledky vědeckého bádání formou krátkých abstraktů má podstatně delší tradici než 100 let. Dokonce předchůdce a určitě vzor pro Chemical Abstracts, německý referátový časopis *Chemisches Zentralblatt*, se může chlubit delší dobou své existence od r. 1830 až do r. 1969, tedy 139 let, nicméně zanikl a je dnes jen historickou reminiscencí. Bezprostředním předchůdcem Chemical Abstracts byl referátový časopis *Review of American Chemical Research*, který založil W. A. Noyes v r. 1895 a tentýž chemik poté, co se stal vedoucím pracovníkem National Bureau of Standards, zahájil na této instituci vydávání Chemical Abstracts v r. 1907. Samozřejmě ve velmi skromných poměrech v podobě 3 chemiků na poloviční úvazek. Po krátkém působení na University of Illinois se v r. 1909 tyto první abstraktoři přestěhovali na Department of Chemistry Ohio State University, v jejímž kampusu sídlila redakce Chemical Abstracts až do r. 1965. A i potom, když univerzitní kampus potřebám nestačil a bylo vybudováno nové sídlo, zůstává adresa CAS v blízkosti a je stále Columbus, Ohio. Dnes ovšem jako komplex budov s 1200 zaměstnanci, což už nepochybně velká „firma“ je, nebo i v pravém slova smyslu továrna na zpracovávání chemických informací.

Referátový časopis Chemical Abstracts byl jen dalším časopisem vydávaným Americkou chemickou společností a tato vědecká společnost je v podstatě vydavatelem do dneška. Ovšem narůstající objem činnosti záhy mnohonásobně přerostl problematiku vydávání ostatních časopisů, a tak v r. 1956 byla založena Chemical Abstracts Service (CAS)\*, jako více méně samostatná součást Americké chemické společnosti (ACS). Tím se vydávání CA oddělilo od produkce ostatních časopisů ACS, stalo se finančně soběstačné, současně ale podle amerických, hlavně daňových zákonů s určitým neziskovým statutem, přesněji s omezenou výší zisku. Tato skutečnost se projevuje na ceně jak předplatného, tak i v dnešní době přístupu k elektronickým verzím. Doplňme tento ekonomický aspekt ještě nostalgickou vzpomínkou, že až do konce 1. světové války bylo předplatné pro členy ACS 6 dolarů, v roce 1956 pak 20 dolarů, ale pro univerzity se předplatné zvýšilo ze 60 dolarů v r. 1955 na 350 dolarů v r. 1956. To byly ovšem stále idyllické ceny ve srovnání s dnešním předplatným ve výši téměř 30.000 dolarů za rok.

\* Chemical Abstracts jsou běžně zkracovány jako CA a pod touto zkratkou je chápán vlastní časopis nebo v širším slova smyslu báze dat, tedy data. CAS je naproti tomu producent, tedy Chemical Abstracts Service.

### 3. Několik zastavení na stoleté cestě

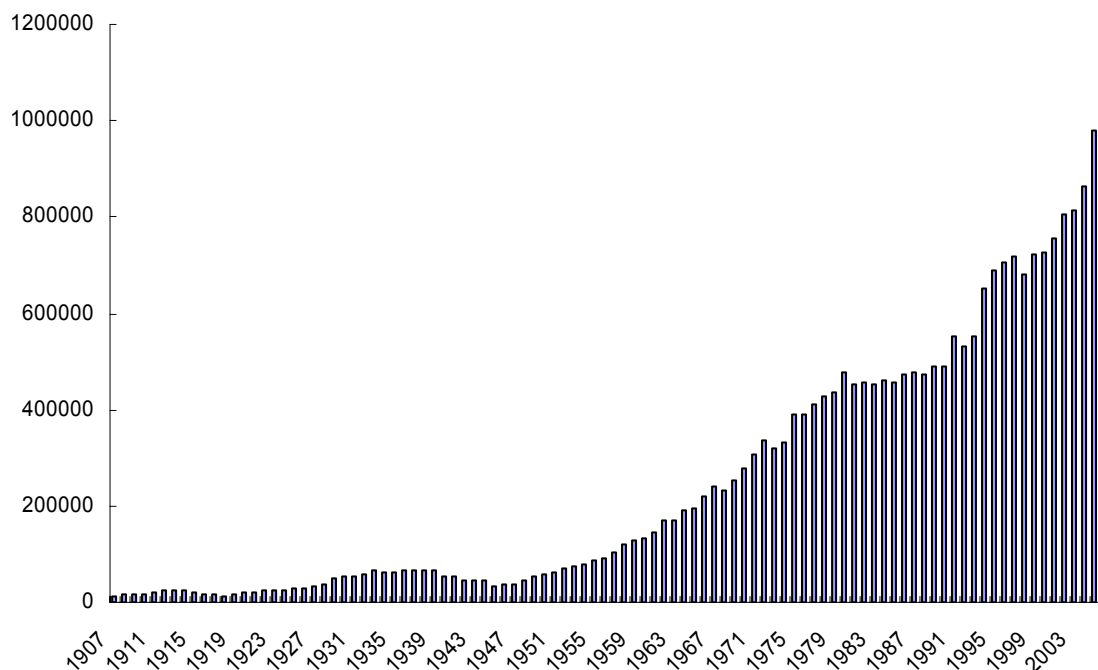
Historie Chemical Abstracts je za sto let pochopitelně velmi bohatá a byla vícekrát popsána prakticky ve všech monografiích pojednávajících o chemických informacích, resp. dříve o chemické literatuře. Velmi podrobně popisuje historii CA např. Maizel ve 3. vydání své knihy<sup>1</sup>. Určitě je zajímavé, kdo oněch sto let produkci CA řídil. Jednoduchý výčet je trochu komplikován existencí „Editors“ na samém počátku a „Directors“ zhruba od poloviny období, celkem ale najdeme jen 12 jmen, výhradně mužů, v prvním období s velkou vědeckou reputací, později s reputací expertů v oblasti chemických informací, ale až v současnosti stojí v čele CAS právník a úspěšný manažer v oblasti informačního průmyslu, Robert J. Massie. Ilustruje to jednak šťastnou ruku při výběru „správných“ lidí a pak skutečnost, že v současnosti jsou zřejmě důležitější manažerské schopnosti před vědeckou zkušeností.

To, že práce začínala v r. 1907 se třemi abstrakty na částečný úvazek a dnes má CAS kolem 1200 zaměstnanců, je jedním z velmi názorných ukazatelů růstu. Zajímavá je skutečnost, že prakticky až do skončení 2. světové války byl poměr mezi zaměstnanci na plný a částečný úvazek nejdříve spíše na straně částečných úvazků, později plné úvazky sice převažovaly, ale stále byly srovnatelné. Teprve až v 50. letech začal počet zaměstnanců na plný úvazek narůstat a v r. 1966 už statistiky částečné úvazky neuvádí. Ještě podstatnější je skutečnost, že od samého počát-

ku bylo vytváření abstraktů naprosto převážně zajišťováno externími abstrakty, kterých bylo už v r. 1907 129 a jejichž počet dosáhl maxima v r. 1966, a to 3292. Téměř polovina působila mimo USA, takže lze konstatovat, že po velkou část své existence byly CA v pravém slova smyslu produktem celosvětové chemické komunity. Maizel<sup>1</sup> uvádí, že na počátku šedesátých let více než 3/4 všech abstraktů bylo připraveno externisty. V sedmdesátých letech, hlavně jako důsledek přechodu na elektronické postupy zpracování primárních dat, počet externích abstraktorů postupně klesal a v současné době jsou externisté spíše výjimkou. Větší skupina působí v Japonsku. Neboli téměř vše je „vyráběno“ v sídle CAS.

Už v prvním roce bylo zpracováno 11 847 abstraktů na 3437 stránkách a tato čísla pochopitelně pomaleji i rychleji narůstala, opravdový strmý nárůst začal až po 2. světové válce (obr. 1). Velmi podstatná je skutečnost, že už v prvním roce bylo z výše uvedených 11 tisíc dokumentů 3853 patentů, což je plná třetina a od 2. ročníku se začaly uvádět i odkazy na knihy. Zpracovávání patentů je až do současnosti jednou z nejdůležitějších součástí CA, protože kromě plně komerčních a značně drahých patentovýchází firmy DERWENT, jsou dnes CA stále hlavním zdrojem patentových informací z chemických disciplín.

Stále se zrychlující tempo růstu objemu chemických informací také ilustruje skutečnost, že celkově první milión abstraktů byl dosažen za 30 let práce. Pro další milión bylo třeba o něco víc než polovina (18 let) a shodou okol-



Obr. 1. Růst počtu zpracovaných dokumentů v období 1907–2006 (cit.<sup>10</sup>)

ností právě v loňském roce 2006 stačil právě jen jeden jediný rok na zpracování 1 016 669 dokumentů, čímž celkový počet odkazů za 100 let dosáhl téměř 25 milionů.

Zpracovávání dokumentů, článků a patentů, do podoby stručného bibliografického záznamu s abstraktem, samozřejmě vyžaduje systém, v dnešní terminologii systém databázový. A s tím opět souvisí zásady a hloubka indexování, výběr předmětových hesel a jejich systém, citace zdrojů a řada dalších aspektů. Ve většině těchto aspektů ovlivnila politika a zásady CAS nejenom problematiku chemických informací jako takových, ale i vývoj v řadě odvětví chemie. Nejvýrazněji asi v oblasti nomenklatury, stále se řeší otázka, zda používat nomenklaturu podle CAS nebo IUPAC. Bez ohledu na individuální postoj k této otázce je skutečností, že CA a příslušné rejstříky představovaly a stále představují nejrychlejší řešení problému, jak nazvat tu či onu sloučeninu. Obdobnou roli sehrála CAS také v unifikaci bibliografických citací, řada světových chemických periodik, včetně našich, požadovala a požaduje od autorů citace podle CA. A samozřejmě soupis, dnes elektronický v podobě báze dat CASSI, všech zdrojů, ze kterých byl kdy převzat nějaký dokument do CA, představuje neocenitelný zdroj informací o odborných a vědeckých časopisech za 100 let.

Samostatnou kapitolou jsou rejstříky CA, jejich organizace, systém a zpracování, které představují monumentální soubor nástrojů pro jakoukoliv práci s chemickými informacemi. Vzorcové rejstříky a rejstříky chemických názvů (dříve součást předmětových rejstříků) představovaly *de facto* soupis všech do té doby známých chemických sloučenin. Už od počátku si editoři uvědomovali důležitost rejstříků a zahájili vydávání víceletých rejstříků, dříve desetiletých a od r. 1957 pětiletých, prakticky vůbec nejdůležitějších částí CA před současnou elektronickou verzí. Zatím poslední z těchto tzv. Collective Indexes CA za léta 1997–2001 zabere na regálu více než 10 m. Určitě není nadsázkou, řekneme-li, že rejstříky CA představovaly klíčovou bránu do světa chemických informací.

#### 4. Nástup počítačových technologií

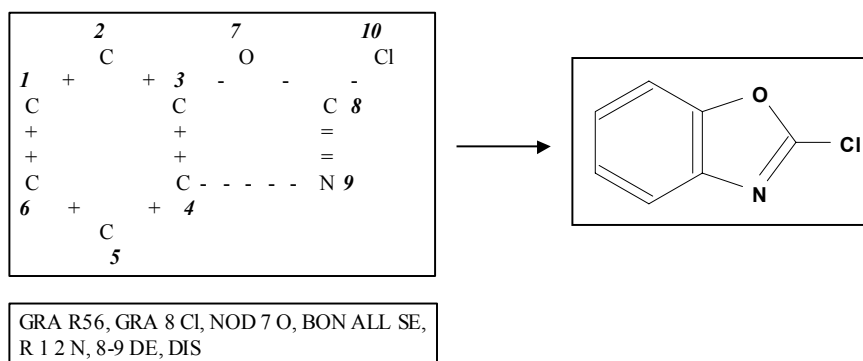
Během stoleté historie bylo nutno učinit řadu více či méně zásadních rozhodnutí, ale včasná orientace na počítačové technologie bylo asi vůbec to nejdůležitější. Především díky jemu CA „přežily“ stále se prodlužující interval mezi vyjitím práce v primárním časopise a jeho abstraktem, protože klasické ruční zpracovávání informací nemohlo rychle narůstajícímu objemu stačit. To byl také důvod, proč *Chemisches Zentralblatt* v r. 1969 přestal vycházet, protože zpoždění dosahovalo téměř 2 roky. I z dnešního hlediska těžkopádné počítačové technologie té doby představovaly obrovský pokrok při zpracovávání abstraktů, protože místo excerpování dat z primárních zdrojů na kartotéční lístky a následné pracné vytváření rejstříků jejich tříděním nabízely nesrovnatelně efektivnější možnosti. Přímou učeňnicovou ukázkou počítačové historie je produkce nepochybně prvního „počítačového“ bibli-

ografického časopisu *Chemical Titles*<sup>2</sup>, který CAS začala vydávat už v r. 1961. Znamená to, že počítačová éra trvá pro CA už více než polovinu celé jejich historie. Pro mnohé současníky, kteří žijí v představě, že počítače jsou produktem posledních let, to může být šokující zjištění.

Šedesátá léta je možné označit v pravém slova smyslu za pionýrskou dobu zavádění počítačových technologií do produkce jak CA, tak chemických informací vůbec. Klíčovou otázkou byla pochopitelně identifikace chemických sloučenin, do té doby realizovaná jedinečně prostřednictvím nomenklatury s přispěním sumárních vzorců. Určitě patří k nejprogresivnějším rozhodnutím tehdejšího vedení CAS volba spojovacích tabulek vycházející z teorie grafů jako základní princip reprezentace struktur<sup>3</sup>. Je nutné si přitom uvědomit, že grafické terminály a grafická rozhraní byly hudbou daleké budoucnosti a tato cesta znamenala mnohem větší nároky na paměťová média tehdejších počítačů. Důsledkem tohoto rozhodnutí je pak dnešní registrační systém chemických sloučenin mající důsledky nejenom v samotné chemii, ale v celé řadě dalších oblastí vědy i veřejného života.

Zásadní orientace na počítačové technologie byla završena počátkem sedmdesátých let stanovením celé řady pevných zásad a pravidel, včetně nomenklaturních. Rozhodným rokem byl rok 1972, jinak začátek dalšího pětiletého rejstříkového období, od kterého si řada, v té době již zkušených uživatelů CA z řad chemiků, začala stěžovat, „... že se teď v abstraktech nedá nic najít“. Důvod spočíval v tom, že nomenklatura se stala důsledně systematickou (CA dovoluje používat jen 8 triviálních názvů), byl vybudován hierarchický systém závazných předmětových hesel a celá řada dalších systémových opatření jako logický požadavek počítačového databázového systému. Uživatelům byl poskytnut jako nástroj Index Guide, doslova průvodce rejstříky, který tyto zásady shrnoval a umožňoval např. zjistit vzájemné vazby a odkazy mezi hesly apod. Naprostá většina tehdy stanovených zásad platí stále.

Po nasazení počítačových technologií do vytváření CA bylo postupně umožňováno využívání těchto technologií i uživateli tak, jak to jejich vývoj dovoľoval. Nejdříve byla poskytována digitalizovaná data na právě dostupných paměťových médiích největším chemickým a farmaceutickým koncernům pro využívání na jejich velkých počítačích. Dalším krokem pak byl produkt nazývaný CA Condensates distribuovaný na magnetických páskách a provozovaný buď na lokálních počítačích velkých chemických firem, nebo v tzv. databázových střediscích. Condensates proto, že se na tehdejší médium nevešly celé záznamy abstraktů, ale jen zkrácené formy. Celé záznamy včetně textů abstraktů a soupisu hesel byly k dispozici obdobnou formou od r. 1978 jako CA Search a strukturální vzorce byly doplněny ve verzi CAS ONLINE kolem r. 1980. Stále ovšem jen na vzdálených databázových centrech. Teprve až ve druhé polovině devadesátých let, kdy médium CD-ROM dosáhlo maxima svých možností, bylo možné nabídnout místo tištěného sešitu každý týden jeden CD disk za měsíc se 4 individuálními sešity, jako elektronickou analogii tištěné verze pod označením CA on CD.



Obr. 2. Strukturální dotaz v osmdesátých letech minulého století a dnes

Postupný vývoj využívání digitální formy CA prošel mnoha stále se zdokonalujícími stadii tak, jak se vyvíjely síťové technologie, paměťová média i koncové stanice, v dnešním chápání tedy osobní počítače. Jen jako malá ilustrace je na obr. 2 naznačena formulace dotazu na strukturu celkem jednoduché sloučeniny 2-chlorbenzoxazolu v osmdesátých letech minulého století a dnes. Jen pro vysvětlení, struktura byla „kreslena“ souborem příkazů GRA, NOD, BON atd., + je dohodnutý znak pro aromatickou vazbu<sup>3</sup>.

Vývoj využívání počítačových technologií byl završen úplnou digitalizací veškerého materiálu od r. 1907, takže v současné době je možné pracovat pomocí klávesnice a obrazovky s kompletním souborem dat získaných z primárních zdrojů jako s jedním souborem, tedy se všemi informacemi o chemických disciplínách za 100 let. A je nutné dodat, že nástroj, se kterým je dnes možné s tímto souborem pracovat, aplikace označovaná chráněným názvem SciFinder<sup>®</sup>, resp. ve verzi pro univerzity jako SciFinder<sup>®</sup> Scholar, je považována za zatím vůbec neefektivnější nástroj pro práci s nejenom chemickými informacemi, ale s vědeckými informacemi vůbec.

## 5. Registrační systém CAS

Výše zmíněné klíčové rozhodnutí o způsobu grafické reprezentace struktur chemických sloučenin hrálo na počátku hlavní roli při vytváření CA, především při excerpci chemických sloučenin a jejich zařazování do rejstříků. Abstraktoři tak dostali do rukou nástroj, s jehož pomocí mohli rychle zjistit, zda daná struktura už byla zaznamenána nebo je nová, a to bez ohledu na názvoslovné problémy. Při vkládání údajů do počítačové báze dat se tak vytvářelo propojení mezi sumárními vzorci, systematickým názvem nebo i více názvy a graficky vyjádřenou strukturou a vznikl tak záznam (record) shromažďující všechny identifikační data o dané konkrétní sloučenině. Identifikace tako-

vého záznamu, např. jeho automaticky přidělovaná numerická adresa, je tak souhrnnou identifikací dané sloučeniny. Tato forma se po dalším vývoji standardizovala do podoby tzv. Registračního čísla (CAS Registry Number<sup>®</sup>, CAS RN)<sup>4</sup>, které se postupně vžilo jako nejuniverzálnější nástroj pro identifikaci chemických látek a sloučenin. Tato čísla jsou vytvářena (přidělována) právě v okamžiku, kdy je do systému zařazována dosud neznámá sloučenina (nebo její i velmi malá modifikace), z čehož plyne, že se jedná o čísla v zásadě pořadová bez jakékoliv vazby na skutečnou strukturu dané sloučeniny. Registrační číslo je tak chemickou obdobou rodného čísla a je jen na CAS, aby zajistila nezaměnitelnost a jedinečnost této identifikace.

Registrační čísla se ukázala být velmi vhodnou formou nejenom pro vyhledávání informací o té které chemické sloučenině v elektronických zdrojích, ale stala se dnes celosvětově užívanou identifikací při prakticky jakýchkoliv manipulacích s chemikáliemi, včetně legislativních předpisů a zákonů. To je nepochybně významný vklad CAS do celosvětové problematiky dohod a jednání o způsobech regulace výroby a užívání chemických látek. Jen z čistě chemického hlediska to přineslo jisté problémy vyplývající právě z kategorického požadavku na identifikaci každé, v praxi se jakkoliv vyskytující, chemické látky a sloučeniny registračním číslem. Řada chemických látek a sloučenin je nabízena na komerční bázi a je tudíž registrovaná a má přiděleno registrační číslo, ale nepatří k ní žádný abstrakt primárního zdroje. To je pak důvod, proč se dnes setkáváme s řadou sloučenin opatřených registračním číslem, o kterých ale nezískáme žádné informace v CA.

Nárůst počtu registračních čísel, tedy počtu známých chemických látek a sloučenin, je ještě více impresivní, než grafem na obr. 1 demonstrováný nárůst počtu abstraktů. Zatímco soubor v r. 1968 měl něco přes 1 milión záznamů, v r. 1990 překročil 10 miliónů, za dalších 10 let se blížil 30 miliónům a na konci r. 2006 to bylo 88,758.285 miliónů. Tato čísla jsou ale poněkud zkršlována postupným rozšiřováním záběru, hlavně zařazováním tzv.

„Sequences“, neboli nejobecněji fragmentů biologických makromolekul. Chemických sloučenin v užším slova smyslu je více než 31 miliónů. Tak, jako jsme dříve hledali informace o chemických sloučeninách ve vzorcovém rejstříku nebo rejstříku chemických názvů, tak máme dnes k témuž účelu bázi dat CAS REGISTRY<sup>SM</sup>, která je sice součástí souboru bází dat CA, ale identifikuje se samostatně.

## 6. Chemical Abstracts v našich zemích

Je samozřejmé, že i v českých a slovenských institucích byly CA běžně využívány, ještě v šedesátých letech minulého století existovalo kolem 50 předplatných. Pak začal tento počet rychle klesat, jednak pro nedostatek deviz, jednak omezováním cenově výhodných tzv. členských předplatných. U nás existují pravděpodobně jen dva kompletní soubory tištěných CA od 1. svazku z r. 1907, jeden na VŠCHT v Praze a druhý na Ústavu organické chemie a biochemie AV ČR. Jsou ovšem kompletní do ukončení odběru tištěných sešitů, k čemuž docházelo koncem 90. let. Předplatné dále pokračuje pro verzi CA on CD a pravděpodobně i v podobě tzv. CA Selects<sup>TM</sup>, neboli jako úzce oborově zaměřené tištěné sešity.

Dnes ale již jen klesající počet pamětníků ví, že přes nepřízeň politické a ekonomické situace, jsme měli možnost se na nástupu počítačových technologií a celém postupném vývoji zpracovávání a zpřístupňování CA podílet ve velmi širokém měřítku. Počátkem sedmdesátých let se totiž vytvořila příznivá konstelace zájmu o počítačové informační zdroje na straně jedné a skupinou poučených chemiků na straně druhé, jehož výsledkem byl nákup tehdy právě uváděné na trh báze dat CA Condensates na magnetických páskách. Takže už od r. 1972 bylo možné na počítačích tehdejší Ústřední technické základny UVTEIN (dnešní Ministerstvo informatiky) pracovat s touto verzí CA, ovšem jen vsádkově na základě předem formulovaných dotazů. Přípravu dotazů zajišťovala Skupina strojových řešerů Ústřední informační služby chemie, součást Výzkumného ústavu technických a ekonomických informací chemického průmyslu, která v době své největší aktivity zpracovávala a rozesílala až 2800 řešerů týdně. Počátkem osmdesátých let se pak nakupovaly pásky s verzí CA Search, tedy už plná verze záznamů CA a tato praxe trvala až do r. 1992, kdy byl nákup pásek financovaný ze státního rozpočtu ukončen. Dokonce byl již v r. 1979 provozován „online“ přístup, ovšem jen na několika pracovištích, a to do báze Chemistry Industry Notes (CIN), což je stále existující přidružená báze k CA<sup>5</sup>. Koncem osmdesátých let pak už bylo možné pracovat prostřednictvím telefonních linek a modemů i s bází CA Search v širším měřítku. Takže jsme měli možnost plných 20 let využívat prakticky stejné technologie přístupu k informacím CAS jako vespole země. Zda a jak tato úroveň přístupu ovlivnila náš výzkum, může být pochopitelně předmětem diskuse, v každém případě jsme měli možnost získat praktické zkušenosti s využíváním informací v digitální podobě, což se nepo-

chybně v následné, už politicky neovlivněné etapě určitě zúročilo.

## 7. SciFinder<sup>®</sup>

Bylo by určitě velmi zajímavé zjistit, kolika chemikům je tento pojem v současnosti všeobecně známý a kolika vůbec nic neříká. Ti šťastnější, kteří už jeho prostřednictvím mohou s CA pracovat, už dnes tvrdí, že si bez něj nedovedou práci vůbec představit. V každém případě je to další ilustrace, jak si CAS dokáže stále držet vedoucí postavení v oblasti vědeckých informací ve zcela obecném slova smyslu. Je naprosto logické, že tak zásadní změna šíření vědeckých poznatků, jakou nabízejí počítačové a síťové technologie, nemůže zůstat na úrovni převádění stávajících tištěných zdrojů do digitální podoby, ale musí se orientovat na využití možností, které tyto technologie nabízejí.

SciFinder<sup>®</sup> je počítačová aplikace, můžeme říci klient, který umožňuje vysoce efektivním způsobem pracovat s excerpovaným materiálem CA, tedy s bibliografickými citacemi, abstrakty, názvy a strukturami sloučenin, ale i s jejich reakcemi a celou řadou dalších informací, které CAS při své činnosti nasbírala<sup>6</sup>. V jedné aplikaci tak pracujeme se všemi individuálními a navzájem propojenými rejstříky, včetně takových, které v tištěné podobě neexistují.

Tato aplikace je koncipována jako maximálně intuitivní, neboli umožňující pracovat prakticky bez jakýchkoliv školení a manuálů. Budiž konstatováno, že tak úplně to není pravda, ale to už není ani tak chyba této aplikace, ale jen důsledek toho, že chemie není jednoduchá věda a tudíž ani vyhledávání vědeckých informací v tomto vědním oboru nemůže být zcela primitivní. V každém případě není na uživateli vyžadováno, aby nějak formuloval svůj dotaz pomocí logických operátorů a už vůbec nemusí znát označování databázových polí nebo jiných detailů databázové struktury. Koncepční dotaz, tedy dotaz, který je formulován pomocí textových hesel a výrazů, může být zadán formou věty v lingvistickém slova smyslu a je ponecháno na této aplikaci, aby z ní vytvořila formálně správný dotaz. Nejedná se přitom o žádný postup v oblasti umělé inteligence, ale jen o důslednou aplikaci současných možností počítačových technologií, které mohou ze zapsané věty (myšlenky) vybrat významově důležité termíny, porovnat je se svým slovníkem a zkombinovat všechny logické alternativy současně s algoritmicke zabudovaným požadavkem na blízkost výrazů. Již zmíněný Index Guide, který umožňoval nalezení jedině použitých závazných termínů a odkazoval na synonyma nebo příbuzné termíny, je jednoduše zabudován do této aplikace evidentně spolu s řadou dalších nástrojů<sup>7</sup>. Obdobně jsou sloučeny vzorcové a názvoslovné rejstříky a propojení s bází CAS REGISTRY<sup>SM</sup> pak nabídne všechny struktury, které vyhovují jak vzorci, tak i názvu a pochopitelně nakreslené strukturně, včetně možnosti struktury modifikovat a vyhledávat tzv. substruktury, neboli struktury vnořené do složitějších struktur.

## 8. Jaké jsou současné možnosti práce s Chemical Abstracts ?

Počítačové technologie nacházely uplatnění nejdříve při vlastním zpracovávání primárních informací a teprve v poslední době došlo k jejich aplikaci také na straně uživatelů, chemiků v laboratořích i průmyslu. Paradoxně ale nástup těchto aplikací do určité míry zkomplikoval práci s CA v tom smyslu, že tradiční možnost, tj. navštívit nejbližší odbornou knihovnu, totiž už bohužel nemusí vést k cíli. Jaké jsou tedy vlastně současné možnosti pracovat s CA?

Především je nutné si uvědomit, že i když CAS zpracovává primární zdroje v zásadě koncepčně stále stejným způsobem, tj. primární dokument nejenom zaznamená jako bibliografickou citaci a indexuje jej z hlediska obsahu a tematiky, ale také z něho excerpuje studované chemické sloučeniny i jejich vzájemné vztahy a v posledních letech excerpuje i fyzikálně-chemická data, má dnes daleko větší možnosti, jak takto získaný materiál nabídnout chemikům a také je široce využívá. Pro přehlednost je současná nabídka sumarizována v tabulce I.

Z uvedeného přehledu vyplývá nepříjemný závěr, že chemik, který nepracuje na instituci, která některou z uvedených forem předplácí (přesněji řečeno, má licenci na její využívání) nebo i zájemce z jiných vědních oborů, nemá moc možností na vybranou a může dospět k závěru, že se k tomuto zdroji jednoduše nedostane. Předplatné tištěných verzí bylo většinou ukončeno kolem r. 1997 až 1998, takže v současné době už je v našich knihovnách téměř desetiletá mezera. Forma CA on CD je formálně pokračováním tištěné verze<sup>8</sup> a pokud je realizováno ukládání postupně zasílaných disků na lokální server zpřístupňující data v místní síti včetně knihovny, představuje to logické pokračování a přesun od regálů k počítači v knihovně nebo na pracovišti. Bohužel, producent od r. 2002 striktně upozorňuje na vstupní obrazovce jednotlivých ročníků, že přístup je dovolen jen pracovníkům předplácející instituce.

Elektronické báze dat CA, CAPLUS<sup>®</sup> a CAS REGISTRY<sup>SM</sup> jsou v úplné verzi od r. 1907 přístupné v databázovém středisku STN International bez jakýchkoliv podobných omezení, ovšem podmínkou je otevření uživatelského účtu na tomto středisku a přistoupení na formu úhrady např. fakturou na základě počtu zobrazených odkazů, vzorců i zadaných dotazů a doby připojení k těmto bázím. Tato forma představuje mimořádně silný nástroj, který umožňuje maximálně využít potenciál excerpaných dat včetně i velmi detailních požadavků. Na druhé straně je ale nutné přijmout existenci dotazovacího jazyka, který sice není mimořádně složitý, ale přece jenom předpokládá určitou praxi a rutinu. Proto je tato forma využívána převážně největšími chemickými koncerny, které zaměstnávají profesionální rešeršéry s uplatněním především při patentových rešerších. Existuje ale i jednodušší nástroj, tzv. STNEasy, nicméně i zde je jedině možná forma úhrady platba podle objemu převzatých informací.

Konečně již zmíněný SciFinder<sup>®</sup> představuje mimo veškerou pochybnost téměř ideální nástroj pro práci s CA i s chemickými informacemi vůbec. Jeho jedinou nevýhodou a problémem je, že je striktně přístupný jen pracovníkům průmyslového výzkumu nebo univerzity, které si předplácejí licenční vstup a licenční poplatky jsou mimořádně vysoké. Ceny jsou sdělovány zájemcům na požádání, pro orientaci je možné jen uvést, že je značný rozdíl v ceně, kterou platí průmyslový subjekt a univerzita, ale i pro tu se základní poplatek pohybuje kolem 1 miliónu Kč, a to za jeden současný vstup do systému. Je pravdou, že CAS nabízí značné slevy pro ty univerzity, které poskytují jen bakalářskou úroveň (50 %), nejvyšší cena platí pro univerzity s PhD programem v chemii a i zde je možnost využít tzv. „packages“ zvýhodňující několikanásobné přístupy. V každém případě je práce s programem SciFinder<sup>®</sup> velmi nákladná a hlavně pro chemické katedry na víceoborových univerzitách nebo pro menší chemické průmyslové podniky jen těžko dostupná.

Tento přehled současných možností využívání materiálů CA tak není pro ty chemiky, kteří nemají to štěstí, že

Tabulka I

Přehled současných možností využívání informačních produktů CAS

Skupiny informačních produktů CAS	Možnost využívání
Tištěné sešity, půlroční a pětileté rejstříky, skupiny sekcí a specializované sešity CA Selects	Tradičně volný přístup v knihovnách, v současné době žádná knihovna v České republice plnou verzi tištěných CA nepředplácí
Verze na CD-ROM, CA on CD, rejstřík periodik CASSI a báze dat regulovaných látek	Principiální možnost ukládání dat na lokální server a zpřístupnění v místní síti, od r. 2002 je přístup omezen jen na pracovníky předplácející instituce
Elektronické báze CA, CAPLUS <sup>®</sup> a báze dat REGISTRY pokrývající celé období od 1907	Vzdálený přístup na databázové středisko STN International na základě účtu a hesla
Přístup k celému objemu dat prostřednictvím programu SciFinder, resp. SciFinder Scholar	Institucionální přístup pro průmyslové podniky nebo univerzity

pracují na instituci, která do přístupu k elektronickým zdrojům investuje, příliš povzbudivý, což je bohužel skutečnost a také předmět stížností jak z průmyslového, tak i univerzitního prostředí<sup>9</sup>. Uvážíme-li na druhé straně současný objem primárních informací, který je v CAS zpracováván včetně stále stoupající úzké specializace vyžadující stejný rozsah znalostí od excerpторů a také stále stoupající náklady na produkci bází dat, není se zase až tak moc co divit.

## 9. A co dál ?

Přes nepochybně zcela dominantní postavení v oblasti chemických informací je nutné se nad další budoucností Chemical Abstracts zamyslet. Hlavní starost představuje stále rychlejší nárůst objemu produkovaných vědeckých informací a z toho vyplývající stále stoupající náklady na jejich zpracovávání a v konečném výsledku pak cena jejich využívání. Méně viditelná, ale stejně závažná je stoupající komplexita zveřejňování výsledků výzkumu jdoucí do detailů daného problému bezprostředně srozumitelných jen úzké skupině přímo zainteresovaných specialistů. To pak klade odpovídající nároky na šíři znalostí excerpторů a opět zvyšuje náklady na zpracování primárních zdrojů, protože velmi důležitým aspektem produktu, jako je CA, je spolehlivost a přesnost zpracovaných materiálů. To jsou obecné problémy většiny sekundárních zdrojů, ale především takových, které jsou od počátku založeny na intelektuálním, tedy „ručním“ zpracovávání primárních zdrojů. A k těm CA určitě patří.

Zatím se zdá, že jakékoliv automatizované zpracovávání tak komplexních vědeckých informací, jaké jsou obsaženy v typické publikaci z chemických disciplín, je hrdbou vzdálené budoucnosti, i když i CAS už určitě využívá a bude využívat všech postupně dosahovaných úspěchů na tomto poli jako více či méně užitečné nástroje usnadňující vlastní lidskou excerpční práci. I když se zdá, že CA nemá v současné době prakticky žádnou alespoň částečně srovnatelnou konkurenci, budoucnost závisí především od toho, zda chemický výzkum bude ochoten investovat do kvalitních informačních zdrojů stále vyšší finanční prostředky. Takže se na závěr nabízí otázka pro věstce: Budou Chemical Abstracts existovat za dalších sto let?

## LITERATURA

1. Maizel R. E., v knize: *How to Find Chemical Information*, 3. vyd., str. 60. John Wiley & Sons, New York 1998.
2. Šilhánek J.: *Chemická informatika*. Vydavatelství VŠCHT, Praha 2002.
3. Šilhánek J.: Chem. Listy 91, 237 (1997).
4. Weisgerber D. W.: J. Am. Soc. Inf. Sci. 48, 349 (1997).
5. Kalousek J., Vlasák R., Koniček J.: Čs. Informatika 21, 340 (1979).
6. Ridley D. D.: *Information Retrieval: SciFinder<sup>®</sup> and SciFinder<sup>®</sup> Scholar*. J.Wiley, New York 2002.
7. Wagner A. B.: J. Chem. Inf. Model. 46, 767 (2006).
8. Šilhánek J., Zetková L.: *Chemical Abstracts na CD-ROM*. Vydavatelství VŠCHT, Praha 1997.
9. Flaxbart D.: Issues in Science & Technology Librarianship (ISTL), *Winter 2007*, staženo z <http://www.istl.org/dne16.2.2007>.
10. CAS<sup>®</sup>, Statistical Summary 1907–2006; staženo z <http://www.cas.org/EO/casstats.pdf> dne 10.4.2007.

**J. Šilhánek** (*Department of Organic Technology, Institute of Chemical Technology, Prague*): **Hundred Years of Chemical Abstracts – Hundred Years of History of Chemistry, 1907–2007**

The world most important source of chemical information, Chemical Abstracts (CA), is celebrating hundred years of its existence. This is undoubtedly a good reason for summarizing at least the most important steps and to look both back and ahead. To abstract and index most of the published results of chemical research has been definitely the main and crucial role of this service and the material collected represents actually compressed history of chemical research in the last 100 years. The CA played also an extremely important role in listing and identification of all known chemical compounds through a registry system resulting in indispensable CAS REGISTRY database, which now approaches 100 million items. Among many and sometimes quite difficult decisions, the very early orientation on computer technology was the most important and this is the basis for the present unbeatable position of CA on the chemical information market. The most prestigious product of computer methodology is the SciFinder, which is probably the most sophisticated and efficient tool for searching scientific information. The only complaint is the price of CA, which is unfortunately closely related to ever-increasing production of new chemical information and to the commitment of CA to keep a very high standard of quality of the information products.