

Osobní vzpomínky na (ne)dávné problémy s vyhledáváním chemických informací

Cílem chemické informatiky je zpracovávat a získávat informace o chemických látkách, chemických výrobcích, jejich použití nebo výrobě. Získávání těchto údajů je součástí jak základního, tak aplikovaného výzkumu, ale i obchodní činnosti či ochrany životního prostředí a ochrany zdraví. Chemie, jako obor, má vypracován mimořádně dokonalý informační systém s dlouhou tradicí. Po více než sto let fungoval v prostředí tištěných informačních zdrojů, jejichž typickým představitelem je referátový časopis CHEMICAL ABSTRACTS, který vycházel anglicky. Vedle toho vycházely analogické časopisy v němčině a ruštině. Uživatelská kvalita referátového časopisu závisí jednak na tom, jak rozsáhlou oblast oboru podchycuje, ale i na tom, jak je rozšířen, tj. v kolika knihovnách je dostupný. Při hledání je nutné mít danou práci fyzicky v ruce.

V celkem nedávné (nebo už dávné?) době jsem trávil skoro 2 měsíce za rok v knihovně. Nosil jsem jeden svazek Chemical Abstracts za druhým, ve vybraných sekcích jezdil prstem po textu, a vypisoval práce, které mě zajímaly. K získání textu žádané práce bylo nutné splnit tyto kroky:

- zaznamenat citaci (časopis, ročník, strana...),
- zjistit, která knihovna požadovaný časopis má (seznamy českých knihoven byly tehdy hodně děravé, často neúplné),
- dojet do knihovny, kopírovací služby byly nedokonalé,
- práci prostudovat a vrátit.

Získávání informací bylo jistým druhem dobrodružství, ale také hlavně časově náročné, pečlivé a trpělivé práce.

Po druhé světové válce došlo k mimořádnému rozvoji mnoha odvětví vědy a techniky, vývoj probíhal v oblasti elektronických systémů a rozvíjel se kosmický výzkum, ve kterém počítače hrály klíčovou roli. Kromě nepochybně klíčové role právě v oblasti kosmického výzkumu a navazujících civilních i vojenských aplikací a obecně představě, že počítače hlavně „počítají“, se zapomíná na to, že počítače slouží jako vynikající nástroj pro ukládání a následnou, často velmi jednoduchou, manipulaci s velkými objemy dat. To je ideální nástroj právě pro vytváření tzv. referátových časopisů, protože odpadá nutnost zaznamenávat pro každého autora, název článku, předmětové heslo, bibliografický odkaz i další údaje samostatný kartotéční lístek formátu A4 pro možnost následného vytváření rejstříků. Byla to právě americká Chemical Abstracts Service, která také díky spolupráci s tamějším počítačovým průmyslem měla přístup k nejvýkonnějším počítačům své doby. Tuto možnost využila a už od konce šedesátých let ukládala data excerpovaná z primárních zdrojů do počítačových pamětí. S následným využitím opět elektronických sázecích strojů byla tak tištěná čísla časopisu a hlavně rejstříky, „vyráběny“ pomocí počítačů. Pro chemiky to mělo ale

zatím jedinou výhodu, že se výrazně zkrátila doba mezi otištěním článku v časopise a jeho abstraktem. Výše popsaná pracovní metodika práce v knihovnách se moc nezměnila, chemici museli počkat na další vývoj až do devadesátých let, kdy se tato rešeršní činnost přesunula z knihoven k počítačům. Dodejme ale, že princip práce je stále stejný, jak je popsán výše, jen je nesrovnatelně efektivnější a pohodlnější.

Vedle počítačově celkem jednoduché práce s velkými soubory dat a systémů jejich rychlého prohledání, mají ještě větší význam už skutečně počítačové aplikace využívající teorie grafů, což odstranilo snad neotravnější práci s tištěnými rejstříky, a to vyhledávání informací o chemických sloučeninách. Až do zhruba poloviny devadesátých let bylo jedinou cestou dopracovat se nějak k systematickému názvu (budiž, za pomoci molekulových vzorců) a pak s maximální pozorností výše zmíněnou „prstovou metodou“ prohlížet tištěné rejstříky názvů za současného ad hoc zvládnutí záludností abecedního řazení zahrnující numerické lokanty, pořadí substituentů, několik různých závorek a další nealfanumerických znaků. Se závěrem, zda hledaná sloučenina existuje či nikoliv. Během cca 15 let se stalo samozřejmým, že hledanou strukturu nakreslí chemik na obrazovku a systematický název je vlastně vedlejším produktem takové rešerše. Jak struktura, tak i název i jakékoliv další excerpované informace jsou uloženy na jedné počítačové adrese, která se tak stala velmi účinným nástrojem chemické informatiky pod označením CAS RN, neboli Chemical Abstracts Registry Number. Je to údaj bez fyzikálního významu, který vlastně nahrazuje a současně upřesňuje jednoznačnou identifikaci chemické sloučeniny. Na základě řady mezinárodních dohod je toto číslo používáno v řadě dokumentů, např. pro údaje o legislativě chemických látek se mi na obrazovce objevila rozsáhlá tabulka popsaná písmeny, která pro mne byla naprosto nesrozumitelná, ale pokud tam byl sloupec CASRN, s typickými čísly, která jsem znal a byl jsem je schopen interpretovat. A samozřejmě není problém nechat si zobrazit i nakreslený vzorec látky.

Počítače změnily přístupovou cestu k informacím. Chemik již nemusí cestovat do knihovny, v níž je uložen časopis tištěný na papíru, protože většina časopisů, zpráv z konferencí a jiných odborných setkání je přístupná i v elektronické formě. Některé časopisy, např. Chemical Abstracts, přestaly vycházet v tištěné verzi, a jsou přístupné jen ve verzi elektronické, kterou je možné zobrazit na počítači. Tento stav se promítne i do vztahu chemiků k odborné literatuře. Je pravděpodobné, že informace, které nejsou dostupné přes počítač, budou opomíjeny.

Jozef Horák