



ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ  
FF UK V PRAZE

Lucie Šerá

# Vyhledávání informací v databázích zaměřených na chemii

Verze 1.0

Praha

Listopad 2008

## 1 CHEMICKÁ INFORMATIKA

Chemie je obor, který má velmi dlouhou a úspěšnou tradici práce s informacemi. Jedním z prvních referátových časopisů vůbec byl právě časopis zaměřený na chemii a ze systematické práce nadšených chemiků **19. století** čerpají uživatelé databází jako je Gmelin či Beilstein dodnes.

Oblast vyhledávání chemických informací je součástí oboru **chemické informatiky** („chemoinformatics“ nebo „chemiinformatics“). Zajímavý je fakt, že na rozdíl od teoretiků informační vědy, kteří se od pojmu informatika distancují [Vlasák, 1999], neboť mají za to, že tento pojem si pro sebe přivlastnila počítačová věda, chemikové se pojmu informatika drží, ačkoliv podle definice je to disciplína zabývající se také především informacemi. Definice z jednoho ze sjezdů Americké chemické společnosti zní: *Chemická informatika je obecný název pro disciplínu, která v sobě zahrnuje návrh, vytvoření, organizaci, ukládání, správu, vyhledávání, analýzu, šíření, zviditelňování a využívání chemických informací, nejenom jako takových, ale také jako nástroj nebo index, který je zastupuje pro získávání jiných dat, informací a znalostí* [Warr, 1999 v překladu J. Šilhánka, 2002].

Nicméně např. Peter Willet, odborník vzdělaný jak v chemii, tak v knihovnictví a informační vědě, řadí do chemické informatiky i svoji oblast zájmu, tj. vyhledávání podobnosti reakcí a molekul, reprezentace 2D a 3D struktur atd. a sám přiznává, že se v průběhu své kariéry soustředil hlavně na příslušné algoritmy, datové struktury a výpočetní efektivitu [Willet, 2005].

Chemická informatika tedy mnohem více využívá pro své potřeby **počítačové technologie** a pak je tedy termín informatika zřejmě na místě. Oba aspekty této disciplíny již patrně není možno oddělit, protože se vyvíjely zároveň. Dnes už je standardem přistupovat do chemických databází pomocí strukturního vzorce, což by nebylo možné bez efektivní reprezentace této struktury v počítačové podobě. Nejlépe vystihuje proměnu obsahu chemické informatiky vývoj v pojmenování jejího hlavního literárního zdroje, časopisu, který od roku 1961 vycházel pod názvem **Journal of Chemical Documentation**, od roku 1975 pod názvem **Journal of Chemical Information and Computer Sciences** a od roku 2005 vychází jako **Journal of Chemical Information and Modeling** [JCHIM].

## 2 CHEMICKÉ INFORMACE

Podle Fugmanna [Fugmann, 1985] jsou chemické informace vyhledávány častěji a intenzivněji než informace z jiných oborů. Chemik, zejména v průmyslu, mění poměrně často předmět svého zájmu a potřebuje se rychle seznámit s oblastí nového úkolu. Poukazuje také na to, že i když je chemie rychle se rozvíjející obor, mnoho chemických informací zůstává **platných i po velmi dlouhé době**. Za dostatečně přesně popsané a reprodukovatelné považuje experimenty již z osmdesátých let 19. stol. Navíc, pokud se má opravdu potvrdit platnost určité teorie, je nutné provést velké množství experimentů, které nejsou obvykle v silách jedné pracovní skupiny. Je tedy nutno, aby další vědci navazovali na práce předcházející a vývoj tak mohl jít dopředu.

Typické pro chemické informace je jejich **různorodost**. Databáze jsou nuceny zpracovávat poměrně velké množství různých typů datových formátů, ve kterých jsou informace uloženy. Základem jsou **textová a numerická data** a dále se jedná o různé formy obrázků pro **reprezentaci chemických struktur** (dvou i trojrozměrné) a **reakcí, spektrální informace, grafy** atd.

Hlavní specifita tkví ale v samé podstatě oboru a tou jsou chemické látky, které je potřeba ve vhodném formátu uložit, aby byly zpětně vyhledatelné a aby se uživatel dostal k informacím o jejich vlastnostech, které již jednu byly pozorovány nebo změřeny. O způsobech vyhledávání zejména faktografických chemických informací se bude podrobně hovořit dále.

Následující výčet je souhrnem základních údajů, které je možno v chemických databázích vyhledávat.

### 2.1 BIBLIOGRAFICKÉ INFORMACE

Bibliografické informace jsou informace o primárním dokumentu, konkrétní fakt zjistíme až ze studia primárního dokumentu. Povaha bibliografických informací se výrazně neliší od jiných oborů, jmenujme alespoň:

### Název zdroje

V základních zdrojích chemických informací se názvy časopisů vyskytují ve zkratkovitě podobě (vychází to z dlouholeté praxe referátových časopisů, kterou převzaly i na ně navazující bibliografické databáze).

### CODEN

Alfanumerický šestimístný kód, který přiděluje Chemical Abstract Service seriálovým i neseriálovým publikacím. Tento systém vznikl v roce 1963 v American Society for Testing and Materials pro označování vědecké a technické literatury, v roce 1975 převzala zodpovědnost instituce CAS [CODEN].

## 2.2 FAKTOGRAFICKÉ INFORMACE

Faktografické informace sdělují konkrétní fakt, není třeba dále hledat v literatuře. Jsou typické vysokou náročností na intelektuální zpracování – pracovník je musí v primárním zdroji vyhledat a správně interpretovat v kontextu, to znamená, že i náklady jsou mnohem vyšší, než při zpracování pouze bibliografických informací.

### 2.2.1 Identifikátory

Už od počátků informačního zpracování chemických informací se vědci snaží o co nejjednodušší a nejméně výstižné vystižení povahy chemické látky. Jako samozřejmé se jeví chemický název a vzorec, ale k látkám se často přiřazuje množství nejrozličnějších kódů – identifikátorů. Motivace jsou různé, ale základem je snaha jednoznačně určit látku a tak lépe organizovat a komunikovat dosažené poznání. Ať už šlo v minulosti o tištěné katalogy, abstraktové časopisy nebo o současné databázové systémy a identifikaci látek v nich. Odhlédneme-li od vědeckých databází, významným iniciátorem pro vznik nových a využití starých identifikátorů byl obchod s chemickými látkami a následná legislativní regulace.

#### 2.2.1.1 Identifikátory nevypovídající o struktuře látky

##### Chemical Abstracts Service Registry Number (CAS RN)

Používání registračních čísel Chemical Abstracts Service (CAS) bylo původně zavedeno pro účely **indexace sloučenin**, které se vyskytovaly v literatuře zpracovávané do báze abstraktů. Tuto funkci samozřejmě plní dodnes, ale navíc se registrační čísla stala mezinárodně uznávaným způsobem identifikace sloučenin (a nejen jich). Nalezneme je proto v mnoha obchodních katalozích, v právních normách atd.

Kromě klasických sloučenin, tedy chemicky čistých látek, které obsahují jeden druh molekul a v nich více různých atomů, se registrační čísla přidělují i **jiným typům látek** (ionty, radikály, sloučeniny lišící se uspořádáním atomů v prostoru atd.). Např. registrační číslo pro 1,2-dichlorbenzen je 95-50-1.

##### Beilstein Registration Number (BRN)

Jednoznačný identifikátor užívaný v rámci databáze Beilstein. Z důvodů konkurenčního boje, mezi producenty databázemi Chemical Abstracts a Beilstein nejsou od poloviny 90. let minulého století [UTA, 200?] přiřazovány k novým záznamům v databázi Beilstein registrační čísla CAS. Např. 606078 je BRN pro 1,2-dichlorbenzen.

##### Některé další identifikátory omezené na určitý typ sloučenin

**PDB ID** – čtyřmístný alfanumerický kód, který se uděluje v rámci Protein Data Bank [4] (spravuje Research Collaboratory for Structural Bioinformatics, USA). Zajímavostí této databáze obsahující téměř **40 000 struktur proteinů a nukleových kyselin** je, že je přístupná zdarma všem uživatelům a že struktury ukládají do databáze sami výzkumníci. Prestižní časopisy zabývající se tematikou biomolekul uložení struktur a uvedení PDB ID od autorů přímo vyžadují (viz např. Journal of Molecular Biology [JMB]). Např. 4HHB je identifikátor PDB pro hemoglobin.

#### 2.2.1.2 Identifikátory vypovídající o struktuře látky

##### Simplified Molecular Input Line Entry Specification (SMILES)

Identifikátor patří do rodiny lineárních notací (zápis struktury látky alfanumerickou formou, např. Wiswesserova lineární notace). Jeho autor David Weininger na něm začal pracovat v roce 1983. Kód SMILES může vytvořit člověk, pokud zná zásady, ale primárně je určen pro **počítačové zpracování** - jako reprezentace struktury látky uložená v databázích (jak v komerčních i zdarma dostupných databázích chemických látek, tak i v podnikových informačních

systémech). Majitelé licence (firma **Daylight**) tvrdí, že je dokonce nejpoužívanějším kódem pro tyto účely [Rovner, 2005]. Není ovšem vhodný pro určité typy látek, obecně pro ty s nekovalentními vazbami (van der Waalovy interakce atd.). Problémem je ovšem fakt, že pro jednu molekulu lze vygenerovat více správných kódů. SMILES, které dodržuje určité standardy, aby bylo zajištěno, že ze stejného kódu se vygeneruje vždy ta samá struktura a naopak, se nazývá **kanonické**. Např. pro 1,2-dichlorbenzen je kód SMILES: c1c(cccc1)Cl.

Zajímavostí je, že právě SMILES bývá také vstupním formátem pro **metody QSAR** (Quantitative structure-activity relationship), což jsou metody počítačové simulace, které odhadují vlastnosti látek na základě jejich strukturní podobnosti (např. toxické vlastnosti, aby bylo možné obejít testy na zvířatech atd.)

#### IUPAC International Chemical Identifier (InChi)

Jednoznačný identifikátor chemických sloučenin, vyvinutý v letech 2000-2004 v rámci projektu instituce IUPAC (International Union of Pure and Applied Chemistry). Jde o řetězec znaků, který je odvozen ze struktury látky [IUPAC, UC].

Možnost automatického generování jednoznačného identifikátoru ze struktury jej předurčuje k tomu, aby se v příštích letech stal **základním identifikátorem** pro přesné definování chemické entity. Výhodou je i jeho větší otevřenost vzhledem k licenčním podmínkám. Např. InChI=1/C6H4Cl2/c7-5-3-1-2-4-6(5)8/h1-4H opět pro 1,2-dichlorbenzen.

#### **2.2.1.3 Informace o složení a struktuře**

Poskytují různou mírou jednoznačnosti identifikace (nejvyšší u strukturního vzorce a systematického názvu).

##### Sumární vzorec

Uvádí informace pouze o **složení sloučenin** (např.  $C_6H_6$ ). Někdy je třeba při vyhledávání upravit pořadí zápisu jednotlivých prvků. Např. v rámci Chemical Abstracts se sumární vzorec zadává v tzv. **Hillově pořadí**: uhlík první, vodík druhý, další prvky v abecedním pořadí. Pokud není přítomen uhlík, všechny prvky se uvádí podle abecedy. Počet atomů se uvádí ihned za značkou prvku. Většina systémů nevidí rozdíl mezi velkými a malými písmeny [ACS].

##### Název

Celosvětovým trendem je co největší **přiblížení systematickému názvosloví** podle IUPAC. Některé sloučeniny však mají i několik správných názvů, v databázích je tudíž nutno zjistit, jakým způsobem jsou názvy zaznamenány, i když mnoho z nich vyhledává i mezi uloženými synonymy.

- **Triviální**

Ačkoliv jde o názvy, který nenesou informaci ani o složení, ani o struktuře, triviální názvy jsou hojně rozšířeny a to i pro látky, jejichž systematický název není nijak složitý (např. systematicky „ethen“ v porovnání s triviálním „ethylenem.“ Většinou je ovšem užívání názvu triviálního opodstatněno vzhledem k délce systematického názvu (pro příklad uveďme sloučeninu (1alfa, 4alfa, 4abeta, 5alfa, 8alfa, abeta) - 1, 2, 3, 4, 10, 10 – hexachlor - 1, 4 : 5, 8 – dimetano - 1, 4, 4a, 5, 8, 8a hexahydronaftalen, až na výjimky všude uváděnou pod názvem aldrin).

V mnoha případech byly triviální názvy odvozeny od zdroje, ze kterého se sloučenina získává (kyselina mravenčí), vlastností (modrá skalice), jmen objevitelů (Grignardova činidla) atd.

Speciálním druhem triviálních názvů jsou **názvy obchodní**, pod kterými jsou sloučeniny prodávány. Ty se vytváří většinou kvůli složitosti a délce systematického názvu účinné látky, ale i z marketingových důvodů (např. Paralen).

- **Systematický**

Název, ze kterého je možno odvodit složení i strukturu sloučeniny

- **Strukturní vzorec**

Podává nám úplný přehled o složení i struktuře látky pomocí grafického znázornění

#### **2.2.2 Fyzikální**

To jsou vlastnosti typu bodu tání, varu, hustota, rozpustnost atp. Např bod tání není důležitý jen pro zjištění samotného faktu, kdy se látka stává kapalnou, ale chemik z ní může usuzovat i např. nakolik je látka čistá, čili bez příměsí. U některých vlastností (např. bod varu) je nutné uvést i tlak, za kterého byla veličina měřena.

### 2.2.3 Reakční

Reakce je možné klasifikovat mnoha různými způsoby:

- podle typu: oxidace, redukce, adice, eliminace, substituce, polymerizace atd.
- podle „záměru“: reakce, příprava, výroba, mechanismus
- mohou být pojmenovávány buď pomocí reagujících látek (hydroborace, lipoperoxidace, esterifikace) nebo pomocí jmen jejich objevitelů (Diels-Alderova cykloadice, Wittigova reakce, Friedel-Craftsova acylace).

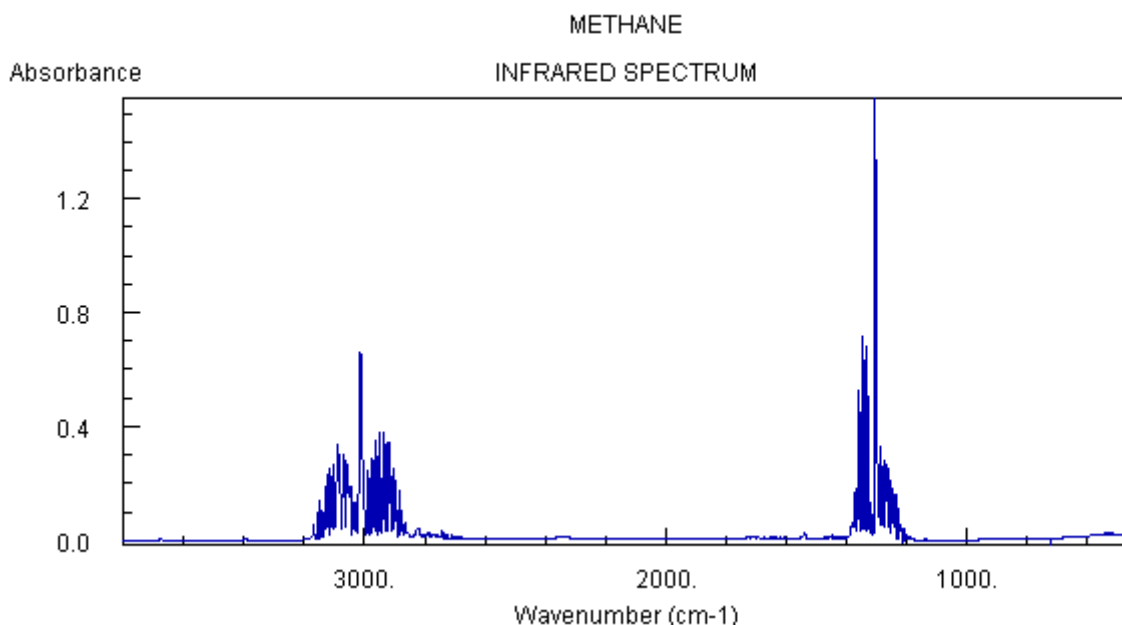
Informace o reakci se často objevují v názvu dokumentu, případně v abstraktu. Někdy bohužel pouze v samotném textu, takže jsou to indexátoři reakčních databází, kteří primární dokument podrobně procházejí a zpřístupňují veškeré informace o reakcích.

### 2.2.4 Spektrální

Velmi často jsou spektrální data omezena na výčet významných píků, ačkoliv v obrazových detailech se skrývá mnoho důležitých informací. Za všechny uveďme jediný příklad:

#### IČ spektra

Výstup analytické metody infračervené spektrofotometrie, založené na schopnosti látek absorbovat elektromagnetické vlnění v rozsahu odpovídajícímu infračervenému záření. IČ spektrum je pro každou látku specifické, slouží jako nástroj identifikace látky a určení složení.



**Obr. č. 1:** Příklad spektra pro methan. Převzato z databáze NIST Chemistry Webbook.

### 2.2.5 Toxikologické

V některých databázích lze vyhledávat informace o experimentech zkoumajících toxické vlastnosti látek. Tyto informace nabývají významu zejména pod tlakem legislativy týkající se nebezpečných látek. Významnými údaji popisující toxikologický experiment je testovaný živočišný druh, dávka a způsob aplikace, případně doba působení a samozřejmě efekt (nejčastěji tzv. střední smrtelné dávka nebo koncentrace).

### 2.2.6 Bezpečnostní

Existují specializované databáze bezpečnostních listů (MSDS – Material Safety Data Sheets), které poskytují tento druh informací. Mnohé z nich jsou dostupné i zdarma. Typickou vlastností týkající se bezpečnosti je bod samovznícení aj.

### 3 VYUŽÍVÁNÍ CHEMICKÝCH INFORMACÍ

Pokud jde o **profesní skupiny**, které využívají chemické informace, každého jako první patrně napadne skupina **vědeckých a výzkumných pracovníků**. V duchu pokroku vědy, navazování na předchozí výzkumy, snahy o minimalizaci prostředků na výzkum (zejména aby nedocházelo k jeho zbytečnému překrývání) a vzájemné komunikace mezi vědci se vytvořila síť odborných informací. Předstupněm této skupiny jsou **vysokoškolští studenti**, kteří se již od začátku studia seznamují s informačními zdroji a učí se je efektivně využívat.

Ale jsou i další, méně zjevné skupiny uživatelů, kteří chemické informace ve svém zaměstnání potřebují.

Na práci výzkumníků navazují **odborníci v podnicích**, kteří zavádějí výrobu do provozních podmínek, jež se v mnoha ohledech liší od podmínek laboratorních, proto i oni jsou hojnými uživateli chemických informací.

Hotové produkty chemické výroby se pak velmi často **převážejí** a vzhledem k hrozbě **dopravních nehod** a následného rizika poškození zdraví a životního prostředí funguje v Evropě systém dohod, které upravují povinnosti při přepravě nebezpečných chemických látek. Zaměstnanci chemických firem - **přepravníci**, kteří za plnění povinností vyplývajících z výše uvedených dohod zodpovídají, proto využívají určité typy chemických informací, stejně jako **dopravci**, jejichž dopravními prostředky se nebezpečný náklad vozí.

Během výroby a přepravy, ale i následně při manipulaci spotřebitele s chemikálií dochází k **nehodám**, při kterých zasahují záchranné složky – **hasiči a lékaři**. Pro obě tyto skupiny jsou zásadní rychlé informace o vlastnostech látky, která ohrožuje okolí, aby mohly adekvátně reagovat. Každou z nich při tom zajímají trochu jiná data – hasiče např. hořlavé, výbušné, žíravé vlastnosti. Lékaře údaje o toxicitě, zásaditosti, resp. kyselosti látky, nebezpečných metabolitech atd. Část těchto informací získávají z tzv. **bezpečnostního listu látky**. Ten by se měl udržovat aktuální, tudíž zpracovatelé těchto listů musí využívat jednak informace o látce vytvořené přímo v podniku (bod tání, hustota...) a jednak informace zvenjšku (zejména legislativní úpravy – řazení do skupin nebezpečnosti, symboly atp.)

Chceme-li pokračovat ve sledování životního cyklu produktu, na výrobu navazuje **prodej a marketing prodeje**. Zde se mohou zdroje chemických informací stát podporou ve smyslu Competitive Intelligence (získávání a analýza informací o firmách za účelem získání konkurenční výhody) a strategického plánování.

Specifickou skupinou uživatelů chemických informací (i když na hraně s farmakologickými a lékařskými informacemi) jsou **zaměstnanci farmakovigilančních pracovišť farmaceutických firem**. Těm zákon přímo nařizuje sledovat odborné informace, konkrétně nežádoucí účinky jejich léčiv. Stávají se tak velmi častými zákazníky informačních služeb typu **monitoring tisku a průběžné rešerše (SDI)**.

S rozmachem chemické výroby a zejména po několika bolestných zjištěních o závadnosti do té doby hojně využívaných látek (škodlivý vliv freonů na ozónovou vrstvu atmosféry, vysoké bioakumulační vlastnosti insekticidu DDT, toxické vlastnosti a perzistence polychlorovaných bifenylů atd.) vyvstala **nutnost regulace obchodu s chemickými látkami**. To znamená, že chemickými informacemi se zabývají dnes (alespoň zprostředkovaně) i **zákonodárci** a zejména různé **registrační instituce** (v ČR Státní ústav pro kontrolu léčiv, Ministerstvo zdravotnictví, Ministerstvo životního prostředí, v rámci Evropské unie od července 2008 nově Evropská chemická agentura se sídlem v Helsinkách).

A v neposlední řadě jsou uživateli chemických informací i sami **spotřebitelé**. S tím jak roste zájem veřejnosti o otázky životního prostředí, o otázku výživy atp., vznikají další a další zdroje, které se snaží jejich informační potřeby naplnit. Bohužel ne vždy jde o snahu objektivně informovat, což je dáno jistou nedůvěrou ve všechno „chemické,“ která v současné společnosti panuje. Ovšem zdaleka ne všechno „chemické“ je špatné a zdraví ohrožující a rozhodně ne všechno „přírodní“ je zdravé a přínosné.

Na příkladech databází z centra STN International ukážeme, které skupiny uživatelů využívají jaké typy informací.

### Všeobecně zaměřené databáze:

- **Beilstein Handbook of Organic Chemistry** – poskytuje široké spektrum informací téměř pro všechny odborné skupiny uživatelů, kteří se zabývají organickou chemií.
- **Gmelin** – taktéž pro všechny odborné skupiny zaměřené na anorganickou chemii.
- **Chemical Abstracts** – základní informační zdroj pro všechny obory chemie, využívají ho jak pracovníci výzkumných institucí, tak odborníci z praxe a studenti.
- **Chemical Abstracts Reaction Search Service (CASREACT)** – faktografická databáze s informacemi o reakcích, uživatelská základna bude spíše na akademické půdě.

### Oborové databáze:

- **Aluminium Industry Abstracts** – zástupce bibliografické databáze určené pro výzkumníky, studenty a odborníky na relativně úzkou oblast zájmu (vše o hliníku, zpracování rudy, použití, vlastnosti, testy, metalurgie atd.), obsahuje ovšem i obchodní informace, které mohou být k užítku i managementu podniku.
- **Copper literature (COPPERLIT)** – podobná databáze zaměřená na zpracování mědi a jejích slitin, neuvádí ale obchodní informace, proto bude patrně využívána na chemii.
- **DECHEMA Thermophysical Property database (DETERM)** – faktografická databáze termofyzikálních dat. Jedna z databází využitelných odborníky v provozech chemických podniků.

### Databáze lékařských, farmaceutických a farmakologických informací:

- **Derwent Drug File (DRUGU)** – zejména pro farmakovigilanční pracovníky, farmaceuty a lékaře.
- **Medline** – rozšířený zdroj lékařských, farmaceutických, farmakologických a toxikologických informací, hojně využíván pro potřeby farmakovigilance.
- **Excerpta Medica (EMBASE)** – podrobná indexace účinných látek poskytuje vhodný zdroj pro farmakovigilanční pracovníky a farmaceuty.

### Toxikologické databáze:

- **Aquatic Toxicity Information Retrieval (AQUIRE)** – faktografická databáze účinků látek na vodní organismy, zdroj využívaný často pro potřeby registrací látek, doplnění údajů do bezpečnostních listů.
- **Hazardous Substance Databank (HSDB)** – informace o velké množině různých vlastností a aspektů, které se dané látky týkají, je vhodným zdrojem pro studenty i veřejnost (její verze je zpřístupněná i na internetu).
- **Registry of Toxic Effects of Chemical Substances (RTECS)** – databáze testů na škodlivé vlastnosti látek (toxicita, karcinogenita, mutagenita atd.), využívá se např. pro doplnění bezpečnostních listů, registrace atd. (její verze je zpřístupněná i na internetu).

### Patentové databáze:

- **International Patent Documentation DataBase (INPADOC)** - patentové databáze obecně mohou být důležitým zdrojem jednak pro odborníky z chemických provozů, jednak pro management (dlouhodobé plánování výroby, akviziční strategii atp.). Pro tuto druhou skupinu shrnuje Giragosian [Giragosian, 1978] sedm oblastí, kde se dají patentové informace využít. Jde o **vlastní nevyužité patenty** a rozhodnutí, co s nimi (např. prodat jiné firmě), dále posouzení, zda vlastní patentovanou technologii prodat v licenci nebo ji zavést do své výroby. Další oblastí je zmiňovaná **akviziční politika** (zda má kandidátská firma mnoho patentů před vypršením atd.). Velmi důležitým nástrojem marketingu je i analýza konkurence na základě patentů, kterou oblast zájmu považují za klíčovou, v kterých zemích patenty vlastní atp. Patentové informace slouží samozřejmě ke **sledování nových trendů** a potenciálních mezer na trhu. S tím souvisí i vývoj nových produktů, který může být korigován podle výsledků **patentové analýzy konkurence**. Autor dále zmiňuje i oblast zahraničního obchodu a nutnost ověřit patentovou situaci v daných zemích ještě před začátkem expanze.
- **Derwent World Patents Index** – pro záznamy z této databáze je možno využít k vizualizaci a dalšímu zpracování pomocí nástroje AnaVist (bude pojednáno dále).



### Databáze bezpečnostních informací a regulace:

- **Material Safety Data Sheets (MSDS)** – databáze bezpečnostních listů může poskytnout důležité informace např. pro výrobce bezpečnostních listů ve své firmě, ale není vhodná např. pro pracovníky zodpovědné za registraci, ani příliš pro vědecké účely, poněvadž neobsahuje odkazy na primární literaturu a není tak možno určit spolehlivost zdroje, případně ověřit správnost dat
- **Regulated Chemicals Listing (CHEMLIST)** – poskytuje informace o různých seznamech (nebezpečných, obchodovatelných...) látek, které jsou uznávány v jednotlivých zemích, zdroj vhodný pro tvůrce bezpečnostních listů, registrátory i management (podmínky prodeje určité látky se mohou v jednotlivých zemích dost podstatně lišit)

### Katalogy:

- **Chemical Catalogs Online (CHEMCATS)** – databáze informací o dodavatelských chemických látek je přínosná pro obchodní oddělení podniků, i pro zásobovatele výzkumných pracovišť
- **Merck** – katalog jednoho z nejznámějších dodavatelů chemikálií, obsahuje informace o vlastnostech látek i o některých legislativních aspektech

## 4 REŠERŠNÍ STRATEGIE V OBLASTI CHEMIE

Damon Ridley [Ridley, 2002] vidí zajímavou analogii v procesu vyhledávání informací a laboratorním experimentem:

EXPERIMENT	VYHLEDÁVÁNÍ INFORMACÍ
1. znalost tématu	1. znalost pokrytí a obsahu databází, alternativních zdrojů a možností vyhledávání
2. potřeba nových znalostí	2. potřeba informací na dané téma
3. návrh na výzkum či experiment	3. formulace výchozího dotazu
4. provedení experimentu	4. provedení vyhledávání
5. zaznamenání a interpretace výsledků	5. pozorné zkoumání výsledků
6. revize a vylepšení experimentu	6. revize dotazu na základě prozkoumaných výsledků
7. provedení revidovaného experimentu	7. objevení nových vyhledávacích možností
8. výsledný výstup experimentu	8. konečná sada odpovědí

V chemii, patrně více než v jakémkoliv jiném oboru, jsou na vysoké úrovni možnosti vyhledávání faktografických informací. Je to dáno stále narůstajícím počtem nově syntetizovaných látek a potřebou se v nich orientovat, proto již na konci 19. stol. vznikala rozsáhlá díla, která se soustředila na výčet známých látek a jejich vlastností. Můžeme tedy říci, že mezi specifika vyhledávání chemických informací patří právě orientace na faktografické informace, na něž se zaměříme i v dalších částech textu.



U vyhledávání faktografických informací můžeme narazit na dvě varianty:

1. **hledání vlastností určité látky** – to je jednodušší případ, kdy je (kromě výběru vhodného zdroje) jedinou starostí správně látku v databázi identifikovat, protože všechny dostupné informace o vyhledané látce se obvykle lehce lokalizují pomocí odkazů vedoucích přímo z hlavního záznamu vybrané látky.
2. **hledání látek, které vykazují určité hodnoty jednotlivých veličin** – tj. například dotaz typu: vyhledejte všechny kapaliny s bodem varu mezi 100 °C a 150 °C. Obvykle se využívají kombinace jednotlivých polí podle nabídky konkrétní databáze. Je třeba vědět, co přesně můžeme od toho kterého selekčního pole očekávat a v jaké formě se požadovaná data zadávají. Zajímavostí je, že v zadáních komerčních rešerší se tento typ dotazu prakticky nevyskytuje,\* půjde o dotazy směřované spíše na databáze v akademickém prostředí.

Další odlišnost plyne ze samé podstaty chemie a to je **potřeba vyhledávat látky a sloučeniny** (chem. látka je obecné označení, dále je dělíme na směsi, které lze jen obtížně jednoznačně vyjádřit např. vzorcem a tzv. chemická individua, čili sloučeniny nebo prvky), které tvoří naprostou většinu objektů zpracovávaných do databází a jsou jednoznačně určeny pouze svým chemickým názvem, identifikátorem nebo strukturním vzorcem.

S chemickými látkami souvisí další specifikum a to je možnost **vyhledávání chemických reakcí**. Ukládání chemických reakcí v databázích má kratší historii než ukládání samotných struktur. Je to dáno větší náročností na následné vyhledání, protože k hledání např. textových informací stačí identifikovat určitý řetězec nebo podřetězec. Při vyhledávání sloučenin podle strukturního vzorce je postup podobný, ale k efektivnímu vyhledání podobných reakcí nestačí jen srovnat struktury reaktantů a produktů. Chemici za podobné reakce považují i ty, kterých se účastní často velmi rozdílné substráty [Behnke, 1990].

Pokud jde o **bibliografické rešerše**, dá se říci, že většina témat těchto rešerší je spojena s konkrétní látkou nebo skupinou látek, je tedy opět nutné nějakým způsobem látku identifikovat. Některé databáze (Chemical Abstracts, Embase atd.) nabízí poměrně podrobnou indexaci, kde jsou k abstraktům přiřazeny názvy látek (popř. registrační čísla) spolu s **rolemi**, v jakých v textu vystupují. Pokud je třeba vyhledat práce o vedlejších účincích nitroglycerinu, články o syntéze téže látky nás patrně vůbec nebudou zajímat. V tomto případě ale stačí zadat roli (např. adverse effects) a dokumenty o syntéze a jiných aspektech by se ve výsledku neměly vůbec objevit.

#### 4.1 OBECNÉ VYHLEDÁVÁNÍ NA ZÁKLADĚ IDENTIFIKACE SLOUČENINY

Jakákoliv identifikace sloučeniny nebo látky je obvykle **základní selekční možností** v přístupu k chemickým databázím a to jak zejména faktografických, tak často i bibliografických (někdy vzájemně propojených odkazy, např. CA PLUS a CAS Registry).

V bázích **faktografického charakteru** je samozřejmě možností víc, vedle chemického názvu a registračního čísla CAS, ještě např. pole pro sumární vzorec, možnost grafického zadání strukturního vzorce atd.

**Bibliografické báze** (a zejména ty, jejichž předmětem zájmu není pouze chemie) se obvykle omezují jen na základní selekční pole týkající se chemických dokumentů. Např. v databázích Chemical Safety NewsBase a Analytical Abstracts lze vyhledávat sloučeniny podle registračního čísla CAS a podle názvu. Víceoborová databáze Inspec má své zvláštní selekční pole Chemická indexace, kde je možné specifikovat celou sloučeninu, část molekuly, složku slitiny atd.

Následující diagram (0) ukazuje návrh pro obecný postup vyhledávání pomocí identifikace látky. Vlastní postup samozřejmě závisí na mnoha faktorech. V souhrnu jsou to zejména tyto:

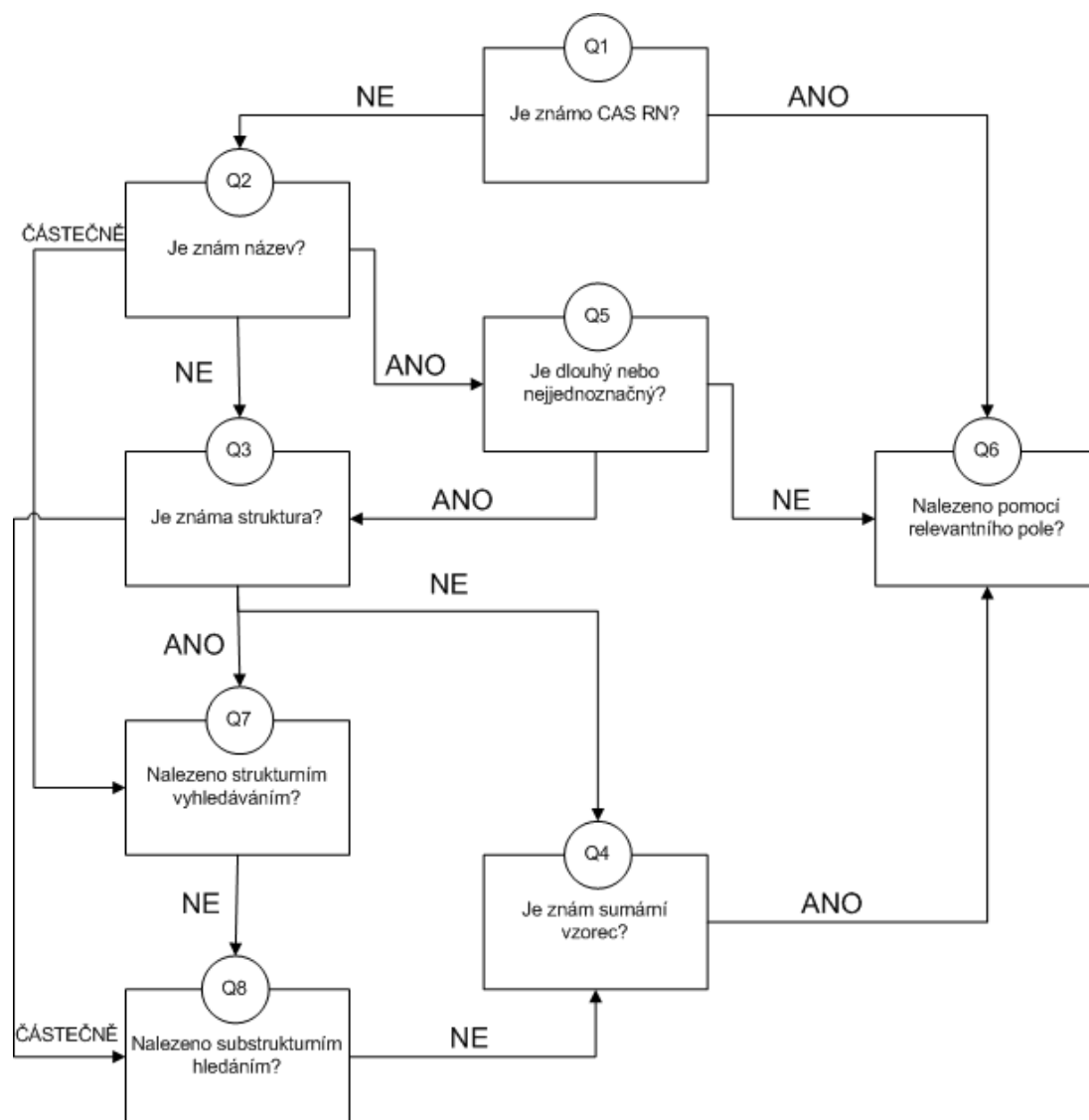
1. **Dostupnost selekčních polí** – tento diagram počítá s víceméně ideální situací, kdy je dostupné kromě klasických polí i strukturní vyhledávání.
2. **Cenová politika databáze** – pokud půjde o hledání v komerčních databázích, uživatel se velmi pravděpodobně pokusí získat registrační číslo CAS v některém zdarma dostupném zdroji, tak aby další možnosti vyhledávání

---

\* BOČEK, Bohumil : červenec 2008 (Praha, osobní rozhovor).

vůbec využít nemusel a ušetřil tak. Při určitém typu dotazu, číslo získat jednoduše nelze, ale i v tomto případě se bude uživatel zřejmě snažit obejít např. strukturální hledání kombinací jiných způsobů hledání.

3. **Znalosti uživatele** - a to jak znalosti oboru, tak znalosti praktické. Pokud si uživatel není jistý správným názvoslovím nebo pokud neví, jak vypadá struktura, kterou chce graficky znázornit, nebude moci využít všechny kroky v diagramu a vyhledávání mu bude trvat déle nebo bude neúspěšné. Znalosti praktické se vztahují k systému, který pro vyhledávání využíváme. Pokud se uživatel neorientuje ve strukturálním editoru a nakreslení složitější molekuly mu zabere hodně času, zřejmě zvolí raději vyhledávání podle názvu nebo sumární vzorec. Zkušenější uživatel naopak od vágního názvu raději přistoupí rovnou k zadání strukturálního dotazu.
4. **Povaha sloučeniny nebo látky** – často se stane, že látku nelze vyjádřit rozumným strukturálním nebo i sumárním vzorcem a uživatel je pak odkázán pouze na vyhledávání podle chemického názvu.

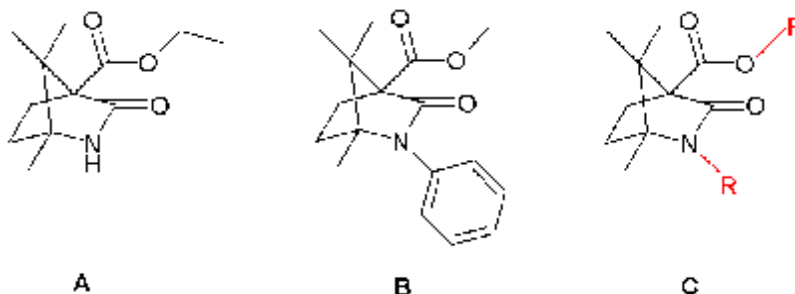


Obr. č. 2: Návrh obecného postupu pro vyhledávání na základě identifikace látky

#### 4.1.1 Strukturální a substrukturální vyhledávání

Pro tento typ vyhledávání jsou uživatelům k dispozici obvykle přímo v databázovém rozhraní tzv. **strukturální editory**. V nich se pomocí jednoduché palety nástrojů nakreslí požadovaná struktura a vloží se do dotazu, obvykle v kombinaci s dalšími hledanými údaji, zadanými tentokrát v textové podobě. Pokud uživatel hledá přesnou strukturu, tak jak ji zakreslil, jedná se o **strukturální hledání**, pokud hledá jiné molekuly, které jsou v nějakém vztahu se zakreslenou strukturou, jde o **substrukturální vyhledávání**.

Příklad z praxe uvádí Judith Currano [Currano, 1999]. Vědecký tým chce syntetizovat látku A, která ovšem ještě nikde v literatuře nebyla popsána. Vědí, že už kdosi připravil podobnou látku B a informace o ní by jim byly k užitku, dají tedy vyhledat strukturu C, aby našli nejen látku B, ale i ostatní jí podobné. Využijí substrukturní hledání a k tomu určené nástroje, které jsou analogií funkce rozšíření (tzv. wildcards) v dotazovacím jazyku pro textové vyhledávání a specifikují místa, kde může být navázán jakýkoliv substituent (pomocí písmene R s vazbou na dusík a kyslík).



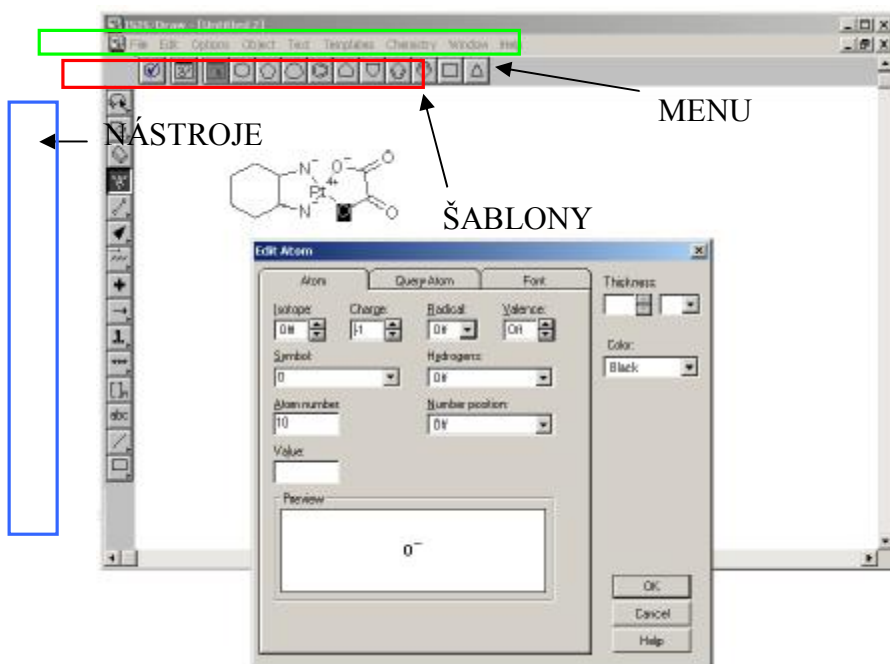
Obr. č. 3: Vzorce požadovaných struktur (převzato z: [Currano, 1999])

#### 4.1.1.1 Strukturní editory

Strukturní editor je program, který umožňuje **grafickou formou vyjádřit strukturu** sloučeniny. Pokud je zároveň spojen s databází, je možné jej využít jako vstupní „pole“ pro zadávání dotazů. A to samozřejmě dotazů na **strukturu samostatných sloučenin**, ale i **reakcí** (po použití příslušných spojovacích znamének). Jinak se editory hojně používají pro nejrůznější **publikační činnost**. Mezi nejrozšířenější patří např. *ISIS Draw 2.5* (MDL Information System) [3].

Tyto programy mohou fungovat samostatně nebo být součástí nějakého balíku služeb – např. databází nebo modelovacích programů. Některé obsahují další možnosti:

- převod dvojrozměrných obrázků do trojrozměrné podoby
- generování chemických názvů podle pravidel IUPAC z nakreslených struktur a naopak
- kontrola správnosti nakreslené struktury (např. správný počet vazeb vycházejících z atomu atd.)



Obr. č. 4: Ukázka editace léčivé látky oxaliplatinu ve strukturním editoru ISIS Draw 2.5 Standalone

## 5 PŘÍSTUP K CHEMICKÝM INFORMACÍM

### 5.1.1 Zdarma dostupné databáze

V drtivé většině jde o databáze přístupné přes **webové rozhraní**. Motivace vzniku jsou různé. Často jde o snahu informovat veřejnost o vlastnostech a účincích nebezpečných látek na člověka a životní prostředí. Neplacené a zároveň velmi užitečné databáze vznikají ruku v ruce s **chemickou legislativou**, na univerzitách potom vznikají např. **databáze pro podporu výuky** dostupné i široké veřejnosti.

Pokud jde o obsah, také se velmi různí, někde dostaneme jen stručnou informaci o **identifikaci látky** (název, vzorce, identifikační čísla, molární hmotnost), někde i podrobnější zprávu o **vlastnostech**. Typické pro tento druh databází je provázanost odkazy na další zdroje. Stejně jako u ostatních internetových zdrojů je potřeba nejprve posoudit, zda je věrohodnost zdroje dostatečná pro naplnění uživatelské informační potřeby.

Uvedme zde jednoduché rozdělení podle skupin producentů s několika málo příklady:

- státní orgány a instituce
  - US. National Library of Medicine – databáze PubChem
  - U.S. Environmental Protection Agency - databáze toxikologických informací ECETOX
  - Ministerstvo životního prostředí ČR – Integrovaný registr znečišťování
  - Evropská chemická kancelář – Evropský informační systém chemických látek (ESIS)
- univerzity
  - Oxfordská univerzita - databáze bezpečnostních listů
  - Vysoká škola chemicko-technologická v Praze – databáze termodynamických vlastností CDATA
- prodejci chemikálií
  - Sigma-Aldrich, Fisher Scientific aj. – databáze bezpečnostních listů

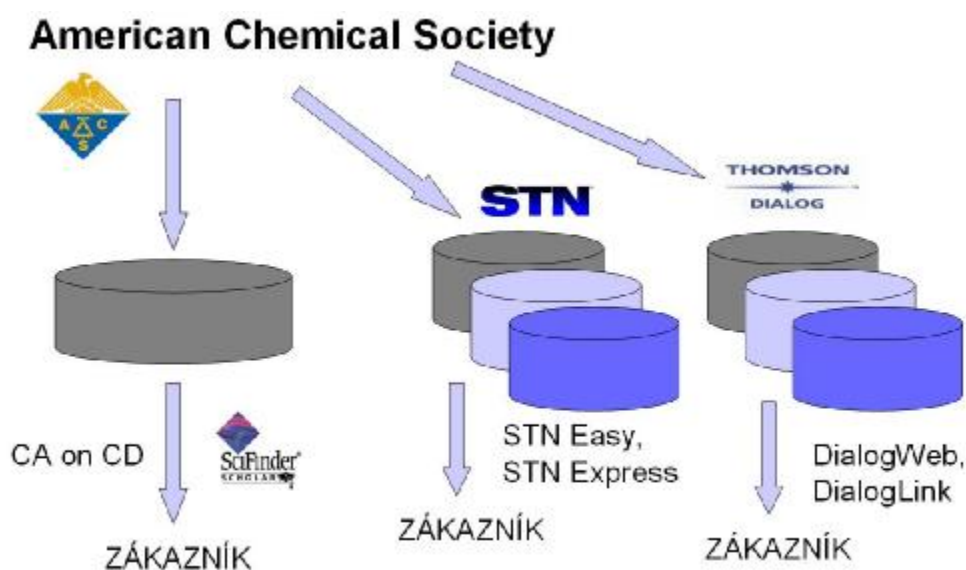
- jednotlivci
  - Mark Winter – vlastnosti prvků periodické tabulky (WebElements)
  - RNDR. Pavel Piskač, Vilém Čermák - Ekotoxikologická databáze

### 5.1.2 Placené databáze

Z hlediska zpřístupnění můžeme tyto databáze rozdělit na:

- samostatně fungující databáze – jde většinou o jednu samostatnou bázi dat zaměřenou na určitý druh informací, která má své vlastní vyhledávací rozhraní a licenční politiku, např.:
  - Royal Society of Chemistry – bibliografická databáze Analytical Abstracts
  - Medistyl, s r.o. – databáze nebezpečných látek Medis-Alarm
- báze dat přístupné v rámci databázového centra – databázi nepřístupňuje výlučně producent, ale jiný subjekt, který disponuje širokou nabídkou databází, má vlastní vyhledávací rozhraní (obvykle několik typů), jež funguje nad všemi zpřístupňovanými bázemi a také si sám řídí cenovou nabídku.
  - STN International
  - Thomson - Dialog

U některých databází fungují tyto dva přístupy zároveň. Důvodem je obvykle snaha vyjít vstříc všem skupinám zákazníků a služby jim takzvaně ušít na míru. Na následujícím obrázku je vidět, jaké jsou základní možnosti přístupu k databázi Chemical Abstracts Americké chemické společnosti.



**Obr. č. 5:** Nabídka přístupů k databázi Chemical Abstracts (od producenta a prostřednictvím databázového centra)

### 5.1.3 Vyhledávací prostředí

Ať už jde o databáze placené nebo neplacené, existuje široké spektrum rozhraní, prostřednictvím kterých lze vyhledávat. Je to velmi důležitá součást celého systému a ohlédneme-li se do minulosti, byl to zejména rozvoj **grafického prostředí** (GUI), který zpřístupnil vyhledávání v databázích široké veřejnosti. Přínos byl a je nesporný, zejména pro neprofesionálního rešeršera.

Podle způsobu přístupu k databázím můžeme prostředí dělit následovně:

- **tenký klient** - není potřeba instalovat žádný speciální program, tento způsob je typický pro zdarma dostupné databáze na internetu
  - **aplikace v rámci webového prohlížeče** – např. PubChem [1], NIST Chemistry Webbook [2]
- **tlustý klient** – vyžaduje instalaci zvláštního softwaru, obvyklé pro placené databáze
  - **komunikační software** – např. STN Express, který umožňuje nejen zadávat dotazy, ale svoji strategii si nejdříve připravit, to znamená např. strukturu hledané sloučeniny nakreslit v offline režimu a následně vložit do dotazu, čímž se výrazně šetří finanční prostředky; poté lze výsledky dále zpracovat, analyzovat atd. (viz. dále kapitolu 6.2.1)
  - **klientský software s lokálním přístupem** – např. Ultranet pro jednoduchou práci s virtuálními CD ROM, Crossfire Commander pro databáze Beilstein a Gmelin

Podle způsobů zadávání dotazů lze databázová rozhraní rozdělit na:

- systémy využívající **dotazovací jazyky** – typické pro databázová centra, je nutné je dokonale zvládnout, uživatel není nijak veden, celou rešerši řídí sám zadávanými příkazy typu *s (oxaliplatin# (w) cancer) and 2005/py and english/la* (najdi požadované termíny kdekoli v záznamu, přičemž primární dokument má být vydán v roce 2005 v anglickém jazyce); *d all 2-3* (zobraz celý obsah vyhledaného záznamu č. 2 a 3)
  - systémy využívající **vyhledávací pole** (tzv. „menu driven“), někdy sdružené do formulářů, pokud záznamy v databázi obsahují rozvinutější strukturu selekčních údajů. Užívané zejména v databázích instalovaných lokálně a databázích veřejně přístupných. Jsou uživatelsky přívětivější, předpokládají nižší míru znalostí o struktuře záznamů, použitých heslech věcného popisu atd. Typické je pro ně to, že mnoho kroků ve vyhledávacím procesu se provede automaticky bez zásahu uživatele.

Směr, kterým se moderní systémy vydávají, je vyhledávání ve více bázích najednou (v prostředí služeb databázových center pro profesionály je to již dlouho známý „multifile searching“). Je aplikován např. v klientu **Crossfire Commander** (Elsevier MDL) pro společně prohledávání databází Beilstein a Gmelin (pokud je hledání omezeno na společná selekční pole) nebo v programu **SciFinder Scholar** (Americká chemická společnost), kde je možno si vybrat z celkem 6 bází dat (CAPLUS, CAS Registry, CASREACT, CHEMCATS, CHEMLIST a Medline).

## 5.2 VLIV TYPU PŘÍSTUPU NA REŠERŠNÍ STRATEGIE

Jak je zřejmé z výše uvedených kapitol, způsobů jak přistupovat k chemickým informacím je celá řada, navzájem se navíc kombinují a bylo by velmi zjednodušující předpokládat, že nijak neovlivní používané rešeršní strategie.

Pokud budeme chtít porovnávat rozdíly v rešeršních strategiích v současnosti, nejzřetelněji je uvidíme na systémech placených systémem pay-as-you-go (pojmenujme je pro jednoduchost „**komerční**“) ve srovnání s databázemi hrazenými formou předplatného či dostupnými zdarma (nazvěme je „**akademické**“).

Při vyhledávání v komerčních bázích dat se rešeršér řídí jednoduchým pravidlem získat co nejvíce relevantních informací za co nejméně peněz. Aby toho dosáhl, musí brát při sestavování své strategie v potaz cenovou politiku stanovenou pro jednotlivé databáze a využít ji ve svůj prospěch. Někdy postupuje i složitěji, než by postupoval v databázích akademických, právě proto, aby minimalizoval náklady. Oproti tomu při vyhledávání v akademických databázích tento problém odpadá a má to i psychologický efekt, kdy se uživatel nemusí bát, že utratí zbytečně moc peněz. Rešeršní strategii si také nemusí tak pečlivě připravovat dopředu a během vyhledávání může své zadání mnohokrát upravovat

Další rozdíly v rešeršních strategiích netkví ani tak ve způsobu placení jako **v typu vyhledávacího prostředí**, které je s databází asociováno. Jak už bylo zmíněno, s komerčními databázemi je spojováno (ne výlučně) prostředí využívající **dotazovací jazyk**, naopak akademické databáze využívají hojně prostředí s vyhledávání pomocí polí a formulářů („**menu-driven**“). V těchto systémech lze pracovat často jen s omezeným množstvím selekčních polí a obvykle je k dispozici pouze základní hledání pomocí operátorů, někdy i třídění a řazení záznamů podle zvolených kritérií, čímž nabídka oproti komerčním databázím končí. Komerční databáze používající dotazovací jazyk jdou v tomto dále a menší uživatelskou přívětivost vyvažují širší nabídkou selekčních polí a větší variabilitou nástrojů.

Určitý dopad na rešeršní strategii má i možnost prohledávat více databází najednou („**multifile searching**“), pokud využíváme služeb databázového centra. Je třeba se ale nejdřív ujistit, zda všechny vybrané databáze podporují selekční pole, které chce rešeršér využít. Poté je možno provést účinnou deduplikaci záznamů, což je v akademických databázích problematické.

## 6 VYHLEDÁVÁNÍ INFORMACÍ V DATABÁZOVÝCH CENTRECH

### 6.1 ASPEKTY REŠERŠNÍCH STRATEGIÍ \*

Jak již bylo naznačeno, v komerčních databázích placených systémem pay-as-you-go se rešeršní strategie musí přizpůsobit jejich **cenové politice**. Uvedme zde nejdůležitější aspekty, ke kterým musí uživatel přihlížet.

#### Počet vyhledávaných termínů

Pokud jej rešeršér odhadne dopředu správně, může využít správný typ účtování vybrané databáze, a tak ušetřit. Např. v STN International lze vstoupit do databáze Chemical Abstracts ve dvou variantách. Jednou variantou je databáze „ZCA“, kde se platí za vyhledávaný termín 2,08 € a nic za čas v databázi strávený, druhou variantou je báze „HCA“, kde je to naopak, to znamená, že se vyplatí v ní vyhledávat, pokud má rešeršér před sebou komplikovaný dotaz se spoustou klíčových slov, které jsou potom zadarmo (ovšem čas je účtován částkou 207 € za hodinu<sup>†</sup>).

#### Cena selekčního pole

Je dobré zjistit, zda nejsou některá pole v rámci databáze zvýhodněna (např. vyhledání registračního čísla v Registry je zdarma) a ty potom s výhodou využít. Na druhou stranu pokud jsou některá pole neúměrně drahá, je na schopnostech rešeršéra, zda vymyslí jak toto pole obejít a požadovanou informaci přitom získat.

#### Typ dotazu

Tento aspekt trochu souvisí s cenou selekčního pole a jedná se zejména o strukturní vyhledávání, které je v komerčních databázích poměrně drahé (70-104 € za dotaz v reakčních databázích STN International). Celkem úspěšně jej však lze v mnoha případech nahradit vyhledáváním v polích typu „CNS“ (Chemical Name Segment) a „CRN“ (Component Registry Number), kterými lze identifikovat část struktury hledané látky. Účinné je např. spojení se sumárním vzorcem (viz. rešerše č. 3 v Ukázkách vyhledávání).

#### Cena příkazu

Často se liší cenou i jednotlivé příkazy, které se při vyhledávání použijí. Obecně platí, že čím více informací pomocí něj získáme nebo čím je časově náročnější, tím je dražší. Např. pomocí příkazu EXPAND můžeme zjistit, kolikrát se dané předmětové heslo v databázi vyskytuje, ale žádné záznamy není možné jen na základě „expand“ zobrazit, nelze použít „wildcards“, dotaz se nedá nijak zkombinovat, proto je příkaz zdarma, na rozdíl od samotného SEARCH, případně nadstavbových příkazů pro deduplikaci, třídění atp. Opět se musí rešeršér rozhodnout, jaký postup mu v dané situaci přinese nejvyšší užitek při minimálních nákladech.

#### Typ zobrazovaného záznamu

Je nutné se zorientovat v různých formách výstupů, které vybraná databáze poskytuje – mnohé nabízí jeden formát zdarma (obvykle dostane uživatel jen základní informaci – např. název dokumentu a přiřazenou indexaci). Užitečný, a ne drahý, je formát, který ukáže pole, ve kterém se nachází hledaný termín. Ve faktografických databázích lze často zobrazit pouze vlastnost, nebo skupinu vlastností, které uživatele zajímají a samozřejmě nejdražší je výpis celého záznamu. Obvyklým postupem je nevypisovat hned po získání požadované množiny výsledků celý záznam, ale pouze některý z levnějších formátů. Z nich si poté vybrat ty opravdu pertinentní, které se nechají zobrazit.

---

\* Tyto aspekty jsou odvozeny od cenové politiky a vyhledávacích možností databázového centra STN International

† Ceny jsou platné pro rok 2008 (viz. [FIZ, 2008a])





## 6.2 SLUŽBY POSKYTOVANÉ DATABÁZOVÝMI CENTRY NA PŘÍKLADU STN INTERNATIONAL

Toto téma je poměrně široké, omezím se proto jen na současnost a základní služby, které poskytuje databázové centrum STN International jako patrně nejdůležitější databázové centrum pro technické vědy, medicínu a chemii.

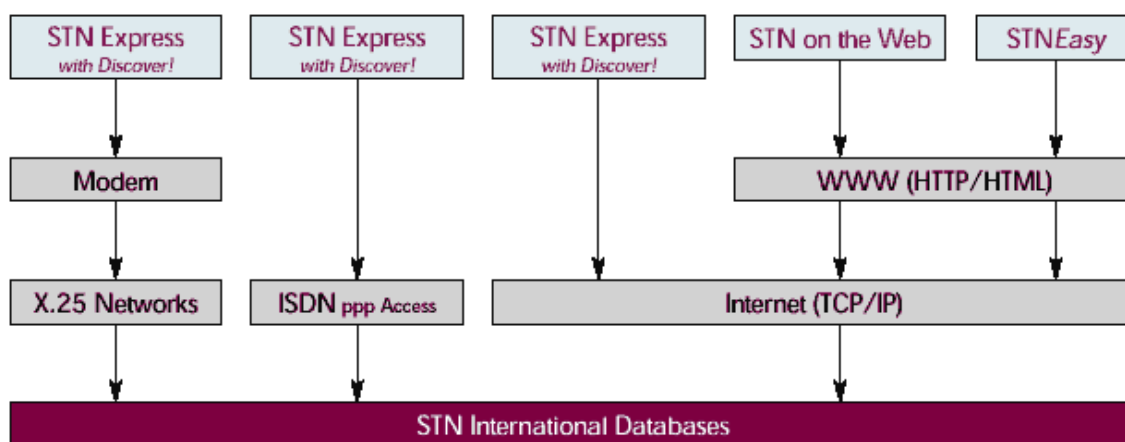
Aby centrum vyšlo vstříc různým skupinám uživatelů, nabízí několik typů přístupů ke svým databázím. Jde především o přístup a vyhledávání prostřednictvím komunikačního software **STN Express** (patří mezi tzv. tlusté klienty, je potřeba zvlášť nainstalovat a nastavit připojení k databázovému centru). Vzhledem k tomu, že tento software využívá pro vedení rešerší dotazovací jazyk, je vhodný spíše pro profesionály. Pro zájemce o přímý přístup přes webové rozhraní je nabízen **STN on the Web**, který se však ve svých funkcích příliš neliší od vyhledávání pomocí STN Express (následné úpravy výstupu lze dělat v STN Expressu). Pro širší odbornou veřejnost je určen přístup **STN Easy**, který znalost dotazovacího jazyka od uživatele nevyžaduje a jeho vyhledávací prostředí je podobné těm, které využívají báze akademické, jak jsme je nazvali v předchozí kapitole. STN nabízí i variantu STN Easy pro intranet, která umožňuje jeho administrátorovi nastavit systém tak, aby vyhovoval maximálně různým skupinám uživatelů v instituci, kde funguje. Odborník, který STN Easy spravuje tak může poskytnout koncových uživatelů lepší podporu při práci.

V následující tabulce je uvedeno srovnání vlastností a funkcí jednotlivých přístupů. Na dalším obrázku je vidět technické řešení připojení k databázovému centru STN International.

	 (STN Easy)	 (STN on the Web)	 (STN Express with Discover!)
Počet přístupných databází	>100	>200	>200
Vyhledávání chem. látek	pomocí polí Jméno a Vzorec	pomocí polí Jméno a Vzorec, (sub)strukturním hledáním	pomocí polí Jméno a Vzorec, (sub)strukturním hledáním
Vyhledávání biosekvencí	ne	ano	ano
Potřebný software	webový prohlížeč	webový prohlížeč	STN Express with Discover!
Selekční jazyk	„menu-driven“ rozhraní, 4 módy hledání (jednoduché, pokročilé, podle CAS čísel a patentové)	Selekční jazyk STN (Messenger)	Selekční jazyk STN (Messenger)
Operátory	booleovské	proximitní, numerické, booleovské	proximitní, numerické, booleovské
Možnost SDI	ne (pouze v řešení pro intranet)	ano	ano
Vyhledávání ve více bázích najednou („multifile searching“)	ano	ano	ano
Řazení dle relevance	ano	ano	ano
Odstranění duplicit	ne (ale lze je	ano	ano

	identifikovat)		
Možnost zužování množiny výsledků	ano	ano	ano
Hypertextové odkazy	ano (např. hledání citací, citované reference, látky indexované v dokumentu)	ano (např. odkazy na registrační čísla, hledání citací, URL, pomocné zprávy, odkazy mezi záznamy z Derwent)	ano (např. odkazy na registrační čísla, čísla patentů, URL)
Formáty výstupu rešerše	html, pdf, rtf	html (zip, exe), rtf (zip, exe, rtf), pdf	trn, rtf
Dostupnost výstupu	4 dny	4 dny	dlouhodobě (uloženo lokálně na PC uživatele)
Možnost následné úpravy výstupu	ne	ano (výstup rtf pomocí STN Express)	ano (různé nástroje pro úpravu výstupů – Custom Report Tool atp.)
Ceny	jednotná částka za dotaz plus částka za zobrazení (závislé na databázi), oprava dotazu zdarma	standardní účtování (pay-as-you-go, tzn. částka za čas připojení, hledaný termín a formát zobrazení)	standardní účtování (pay-as-you-go, tzn. částka za čas připojení, hledaný termín a formát zobrazení)

Obr. č. 6: Srovnání funkcí jednotlivých přístupů k databázím STN International [STN, 2007]



Obr. č. 7: Různé typy připojení k databázím centra STN International [STN, 200?a]

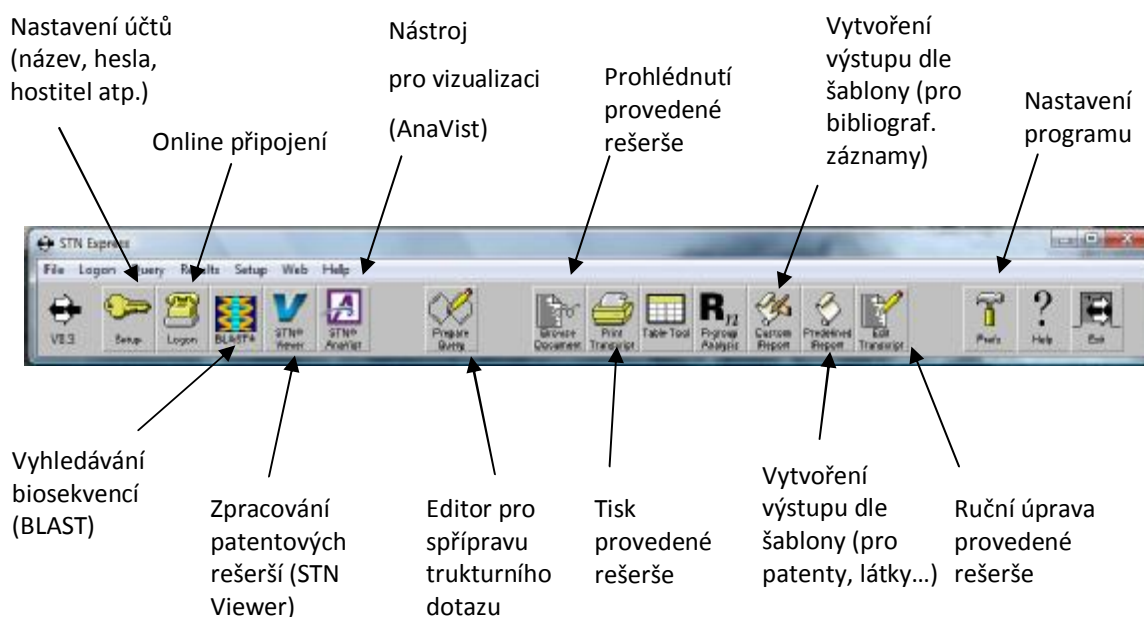
### 6.2.1 STN EXPRESS

Plným názvem STN Express with Discover!, současná verze 8.3 je základním softwarem pro přístup k databázím STN.

#### Základní funkce

- Přihlášení do on-line režimu
  - nabídne místo, kam uložit průběh rešerše (nativním formátem je textový soubor s koncovkou .trn)
  - možnost zvolit si z přednastavených účtů
  - automaticky předvyplní jméno a heslo.

- Vedení rešerše
  - ve formě příkazového řádku
  - podpora dotazovacího jazyka Messenger
  - možnost spuštění předem připraveného dotazu (jednak v textovém formátu s koncovkou .pss, jednak strukturního dotazu, který je možno si připravit v integrovaném editoru)
  - možnost využít tzv. „wizard“ interface, který napomůže při neznalosti dotazovacího jazyka
  - informace o průběžné ceně.
- Ukončení rešerše
  - možnost uložení na 120 minut, po které je možné se opětovně připojit a využít všechny již provedené dotazy (označené písmenem L a číslem).
- Následné zpracování
  - probíhá offline
  - export uložené rešerše např. do rtf (text a obrázky standardně uložené zvlášť se tak rychle spojí do jednoho dokumentu)
  - vytváření přehlednějších výstupů (podle šablon, kde si uživatel sám nadefinuje které pole z nalezeného záznamu má být zobrazeno, jakým fontem, která slova mají být zvýrazněna atp.)
  - takto zpracované dokumenty mají funkční odkazy na full-text a URL odkazy
  - možnost ukládání do formátů xls, html, ascii a rtf.

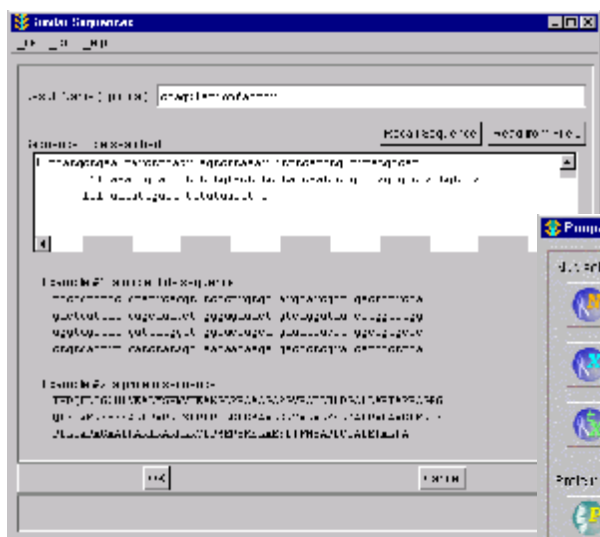


**Obr. č. 8:** Základní panel programu STN Express

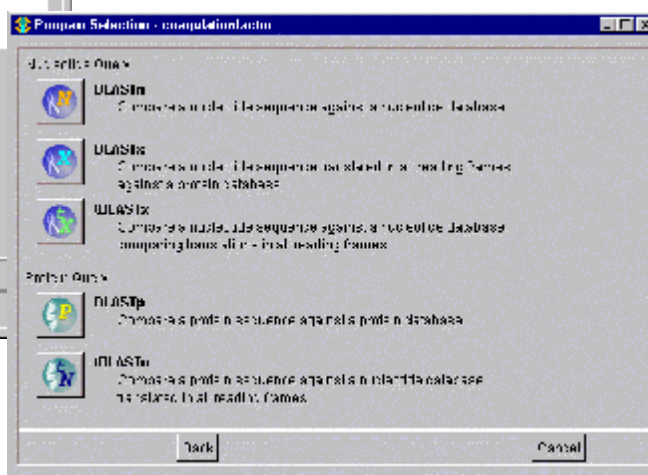
## 6.2.2 Další služby integrované do STN EXPRESS

### 6.2.2.1 CAS Registry BLAST (Basic Local Alignment Search Tool)

Služba pro vyhledávání nukleových kyselin a bílkovinných sekvencí v databázi Registry.



Obr. č. 9: Výběr následujícího zpracování. Převzato z: [STN, 200?c]



Obr. č. 10: Zadání sekvence (v příkladu 1 – pomocí zkratk vyjádřené složení nukleotidů v DNA, v příkladu 2 sekvence zkratk aminokyselin, tvořící bílkoviny) Převzato z: [STN, 200?c]

V následujících krocích je možno si vybrat, zda má být hledání provedeno v celé databázi, nebo jen některé její části a lze zadat, nakolik bude hledání citlivé (zda chceme více nebo méně výsledků). Součástí je i přehledná tabulka jednotlivých hledání a seřídění výsledků. Jednoduše lze vyhledat registrační čísla sekvencí nalezených BLASTem v databázi Chemical Abstracts.

### 6.2.2.2 STN Viewer

Nástroj pro zpracování patentových rešerší, které byly vedeny pomocí STN Express. Uspadňuje hodnocení **pertinence patentů** (pomocí různého značkování), umožňuje vyhledané záznamy třídit a filtrovat. Je možné si vytvářet pomocí tohoto nástroje vlastní projekty a sdílet je s kolegy. Dále umožňuje snadné **získání full-textů patentů**. V první fázi je provedena patentová rešerše např. v Chemici Abstracts nebo World Patent Index. Kliknutím na číslo množiny výsledku (L#) se spustí STN Viewer a k záznamům, kde jsou k dispozici, stáhne plné texty [CAS6].

Seznam patentů lze pak **seřadit podle různých kritérií** (datum publikování, patentová rodina, zaměření, hodnocení uživatelem atd. Každý jednotlivý patent je také možno odeslat do založeného projektu. V rámci projektu si uživatel může nastavit slova, která chce barevně vyznačit a vytvořit si jakousi **mapu**, kde na první pohled uvidí, jak často se která slova a v jakých patentech vyskytují. Ke každému záznamu lze přidávat i poznámky, což může usnadnit komunikaci s případným zákazníkem rešeršera, který mu takto výsledky komentuje. Pokud má zákazník, případně kolega atd. zřízení účet STN, je možné sdílet výsledky zpracované v projektu jednoduše zasláním URL, kde je projekt uložen.

Co se týče poplatků, v současné době se platí za detailní zobrazení záznamu, zobrazení informace o patentové rodině a za uložení projektu.



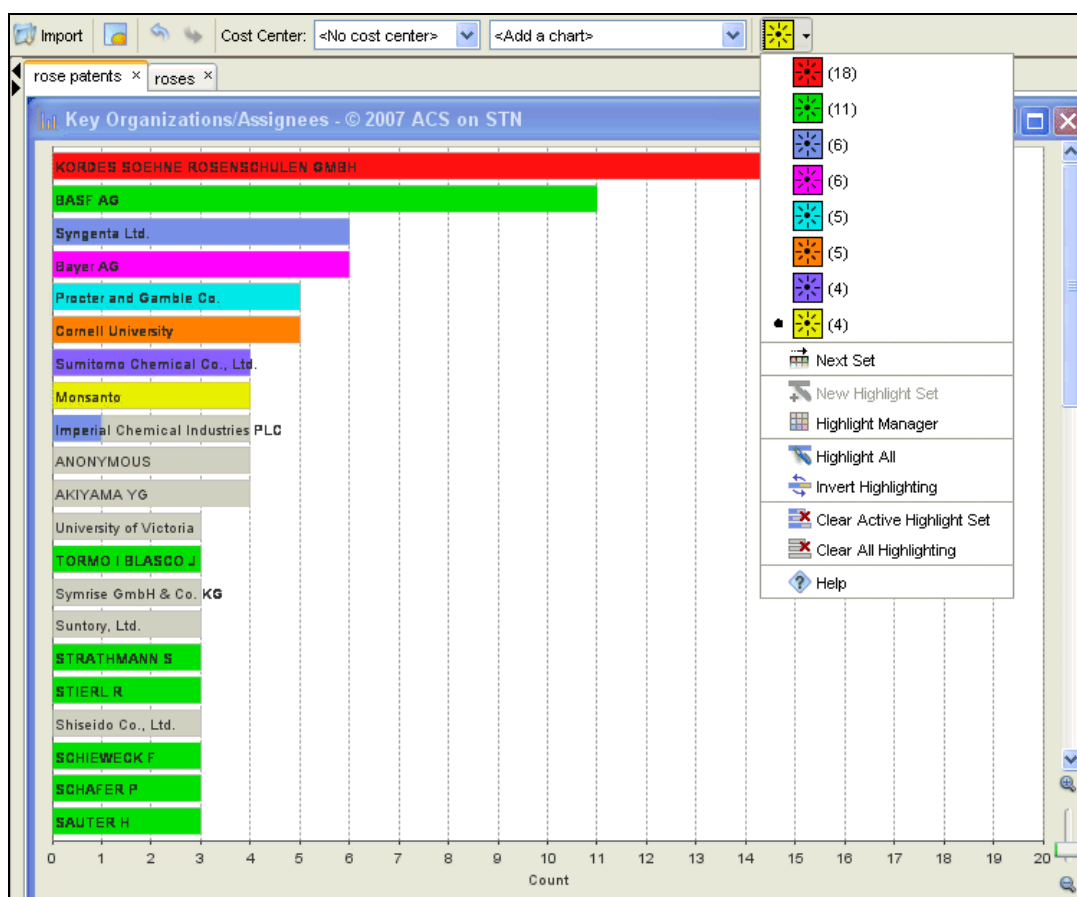
Obr. č. 11: Ukázka označování slov v STN Viewer [Převzato z: Rahman, 2007]

### 6.2.2.3 STN AnaVist

Nástroj pro analýzu a vizualizaci vědeckých a zejména **patentových informací** představený v roce 2005. Pomocí různých funkcí umožňuje odhalit trendy a vztahy spojené s uživatelským dotazem. Konkrétně se může stát důležitým nástrojem pro informační specialisty a jiné odborníky ve firmách, kteří se zabývají metodami **konkurenčního zpravodajství** (Competitive Intelligence), zjišťováním **patentové situace produktů**, vymýšlením nových aplikací již existujících technologií, strategickým plánováním atd. Software jako takový je zdarma pro vlastníky účtu u STN International [CAS7].

Postup práce je následující :

1. Je potřeba připojit se pomocí STN Express, provést hledání a požadovaný dotaz s výsledky naimportovat do STN AnaVist. Databáze, ze kterých mohou být výsledky importovány, zahrnují CAPLUS, PCTFULL, USPATFULL a DWPI (Derwent World Patents Index).
2. Pro vizualizaci je nutné vybrat pole, podle kterých budou dokumenty rozřazeny: titulky / abstrakt, hlavní nárok a kód mezinárodního patentového třídění. Podle nastavených kritérií se dokumenty určitým algoritmem setřídí a na mapě jsou pak příbuzné dokumenty umístěny blízko sebe. Příbuznost se hodnotí na základě konceptů odvozených z textu [CAS8].
3. Výsledkem je mapa zobrazující tzv. výzkumné fronty - zde např. dokumenty vyhledané na základě slova „calendula“ (rodové jméno rostliny měsíček)



Obr. č. 12: Ukázka z programu AnaVist – setřídění přihlašovatelů patentů. Převzato z: [CAS9]



Obr. č. 13: Mapa zobrazující výzkumné fronty („research landscape“). Převzato z: [ACS, 2008b]

Mapu je možné zobrazit i trojrozměrně, kde jsou vidět (jako vrcholy) i četnosti dokumentů v jednotlivých skupinách (klastrech).

Dále se objeví seznamy výzkumníků a organizací, kteří jsou klíčoví pro náš dotaz. Jde v podstatě o setřídění autora/organizací podle četností nalezených dokumentů v jednotlivých letech. Podobných tabulek je možno vytvořit si vícero, jednak podle selekčních polí (kódy patentů podle Derwentu, MPT, země vydávající patent) a jednak podle vlastního označení, který uživatel jednotlivým patentům přiřadí.

V tabulkách se dá lepe orientovat a sledovat různé vztahy pomocí nástroje, kterým se **barvou označí firma** v seznamu klíčových organizací a stejnou barvou zůstane zvýrazněna (resp. její zaměstnanci) i po přepnutí tabulky např. do seznamu vynálezů setříděných podle četnosti přihlášek v jednotlivých letech.

Jak bylo zmíněno, dokumenty se dají označit i vlastním heslem a na základě něho pak vytvářet filtrované seznamy apod.

Tyto mapy, seznamy a tabulky lze poté zpracovat pomocí předdefinovaných šablon do dvou typů zpráv – **přehledové a detailní** (liší počtem tabulek a informací o zpracovávaných dokumentech). Takto vytvořené projekty lze ukládat, tisknout i sdílet s kolegy (případně klienty), kteří mají vedený účet u STN a nainstalovaný AnaVist.

**Cena** se odvíjí **podle počtu výsledků**, které byly nalezeny v rámci dotazu importovaného do AnaVist. Např. pro množinu 2-1000 kusů jde o 195 € (bez daně) za vizualizaci, tj. za všechny výše popsané akce. Zvlášť se platí zobrazení záznamů patentů (liší se podle využití databáze – od 1,20 do 6,38 € za kus) [FIZ, 2008].

## 7 VYHLEDÁVÁNÍ CHEMICKÝCH INFORMACÍ NA VOLNÉM WEBU

Podobně jako je tomu u jiných vědních disciplín, mnoho více či méně hodnotných odborných informací je dostupných **zdarma prostřednictvím internetu**. Problémem je, že velká část těchto informací je součástí tzv. **hlubokého webu**, který je pro nespécializované vyhledávací roboty špatně nebo jen málo dostupný. Podívejme se tedy na **vyhledávací stroje** (search engines) a **katalogy** (directories), na jejich možnosti lokalizace chemických informací (zejména informací o vlastnostech látky) a na trendy, které v této oblasti panují.

Jedním z viditelných trendů je právě **snaha o vyšší využití informací z hlubokého webu**, o zpřístupnění obsahů těchto zdrojů v rámci jednoho vyhledávacího prostředí (projekty ChemFinder, ChemSpider, ChemID a další). Druhým a prolínajícím se směrem je **vize kompletní indexace sloučenin** nacházejících se na webu pomocí jejich počítačově čitelné reprezentace (kódy InChI a SMILES) a využití stávajících (nebo i vylepšených) funkcí vyhledávacích strojů.

Důkazem, že informace publikované na webu nabývají nejen na rozsahu, ale i na významu, je fakt, že **Chemical Abstract Service** v posledních letech indexuje molekuly, které se objevují nejen v odborných časopisech a seznamech typu EINECS (Evropský seznam obchodovatelných látek), ale i ve vybraných souborech dostupných na internetu.

### 7.1 VŠEOBECNÉ VYHLEDÁVACÍ STROJE

Altavista, Yahoo!, Excite, HotBot, Northern Lights a zejména progresivně se rozvíjející Google patří mezi zavedené vyhledávací stroje, které indexují webové stránky v řádech stamiliónů. Nicméně vyhledávání chemických informací se všemi jejich specifiky je poměrně náročný úkol.

Pokud bychom chtěli postupovat podle doporučeného postupu (viz. kapitola 4.1), problém nastane hned u zadání registračního čísla CAS. Podobný formát zápisu jako číslo CAS mohou mít např. výsledky sportovních utkání, soutěžní časy, kódy výrobků atp. Pokud se však přidá ještě slovo „CAS“ má to na relevanci výsledků velmi pozitivní vliv [Glander-Höbel, 2001]. Jak ale uvidíme dále, stejnou roli, jakou má CAS číslo pro specializované aplikace a databáze, by mohl mít pro web identifikátor InChI.

Pokud vyhledáváme podle chemického názvu, obvykle se zavádějící odkazy nevyskytnou, ale setkáme se s problémem stejným jako u méně dokonalých chemických databází a to, že se do vyhledávání nezahrou automaticky všechny tvary a synonyma názvů, které se mohou v dokumentech vyskytovat.

Hledání podle sumárního vzorce funguje (alespoň v Googlu), poměrně dobře, samozřejmě u jednodušších vzorců (např. CH<sub>2</sub>O) může dojít k záměně např. se zkratkou nesouvisející se sloučeninou, ale jinak je poměrně dobrá odezva od různých encyklopedií a na prvních místech se umísťují i stránky, kde byl strukturní vzorec vyhledán jako součást identifikátoru InChI. Strukturní hledání ve všeobecně zaměřených vyhledávacích možných není, jednak kvůli absenci strukturního editoru, ale zejména kvůli zatím poměrně malému zastoupení reprezentace struktury pomocí některého počítačově čitelného kódu. Struktury jsou na webu přítomny nejvíce v obrázkových formátech, které lze vyhledávat standardně zase jen podle názvu.

### 7.2 SPECIALIZOVANÉ VYHLEDÁVACÍ STROJE A KATALOGY

Tyto vyhledávače indexují ve srovnání se všeobecně zaměřenými stroji jen zlomek webových stránek, jejich výhoda ovšem tkví v tom, že je pečlivě vybráno, jaké zdroje budou uživatelům zprostředkovávat. Obecně platí, že čím více je látka známa, tím lepší informace o ní lze nalézt pomocí specializovaných vyhledávačů, pokud jde o málo známou látku, je lepší použít všeobecný vyhledávací stroj [Glander-Höbel, 2001]. V chemicky zaměřených vyhledávacích je obvykle možno



použít i strukturní hledání. Specializované předmětové katalogy fungují na stejném principu jako ty všeobecné, liší se od nich tím, že zpracovávají vybrané stránky s chemickou tematikou.

- **ChemFinder** je zajímavým a poměrně dlouho (od roku 1995) trvajícím projektem. Tento vyhledávač informací o látkách vyvinula firma CambridgeSoft a je zvláštní tím, že si tvoří vlastní databázi látek na základě informací získaných vyhledávacím strojem z kolekce veřejně dostupných webových zdrojů. Proces indexace je v ChemFinderu doplněn o standardizaci chemického názvu., ukládá si i základní fyzikální vlastnosti jako je třeba bod tání a varu [Brecher, 1998]. Tato podoba ChemFinderu je ovšem v současné době vytlačována v rámci CambridgeSoft službou ChemBioFinder, která výrazně omezila množství informací dostupných zdarma a její odkazy vedou v podstatě pouze do placených databází. Původní ChemFinder (s copyrightem datovaným do roku 2004) je ovšem stále dostupný [5].
- **ChemIDplus** [6] – podobný projekt na pomezí databáze a vyhledávače. Je součástí služby Toxnet Národní lékařské knihovny USA a poskytuje jednak základní informace o látce a jednak přímé odkazy na záznam látky v jiných zdrojích. Těchto zdrojů ChemIDplus zahrnuje na 70 a na rozdíl od ChemFinderu jsou mezi nimi i databáze bibliografické. Jde většinou o produkty amerických státních institucí.
- **ChemSpider** [Chyba! Nenalezen zdroj odkazů.] - poměrně čerstvý příklad snahy o vytvoření jednotného rejstříku struktur všech molekul dostupných na internetu (představený v březnu 2007). Autoři ho popisují jako chemický search engine, který byl vytvořen za účelem indexace a shromažďování chemických struktur a jejich vlastností do jediného repozitáře, ve kterém lze vyhledávat [ChemZoo, 2007]. Každá látka má svůj záznam se základními vlastnostmi, odkud se odkazuje na další zdroje, ve kterých lze nalézt další informace. Těchto zdrojů zpracovává několik desítek a jsou mezi nimi i placené produkty, které pro další postup vyžadují login. Zároveň poskytuje i textové hledání ve více než 50 000 odborných článcích. Podobnou zdarma dostupnou službou je **PubChem** [1], o němž je blíže pojednáno v Ukázkách vyhledávání.

Jako příklady typičtějších (nicméně ne tak známých) vyhledávačů a katalogů jmenujme alespoň:

- **LabCrawler** [8] – vyhledávací stroj pro laboratorní informace
- **Chemie.de** [9] – obecně pro chemii, možnost zvolit si druh dokumentů, ve kterých vyhledávat (články, encyklopedie, novinky, katalogy atp.), dále nabízí seřazený katalog
- **ChemIndustry** [10] – vyhledávací stroj, součástí je i katalog zaměřený spíše na obchodní informace

### 7.3 SÉMANTICKÝ WEB A ROLE KÓDŮ INCHI, SMILES

Chemický sémantický web je vize, ve které jsou všechny chemické informace okamžitě přístupné a zpracovatelné pomocí nástrojů založených na W3C protokolech\* [Coles, 2005].

V kapitole 2.2.1.2 byla řeč o identifikátorech automaticky generovaných počítačem a zahrnujících v sobě informaci o struktuře, příp. dalších strukturálních vlastnostech. Je to právě jeden z nich – InChI, který by měl do budoucna pomoci vizi chemického sémantického webu naplňovat. Směřuje totiž k větší viditelnosti volně dostupných databází roboty typu Google. IUPAC vidí jeho využití v komunikaci mezi databázemi, spojování sbírek dat, které byly vytvářeny pomocí rozdílných systémů, udržování laboratorních inventářů chemikálií atd. [Rovner, 2005].

Autoři Coles a kolektiv podotýkají, že kdyby se všechny molekuly, které se na webu vyskytnou, označily odpovídajícím způsobem (konkrétně InChI), mohl by se internet stát jednou velkou znalostní bází s vynaložením mnohem menšího podílu lidské práce, než jaká je potřeba na druhotné sdružování informací, tak jak funguje dnes. To znamená vynechání fáze, kdy jsou informace abstrahovány z publikovaných zdrojů a soustředěny do centralizované databáze [Coles, 2005]. Jistě bychom našli mnoho výhod, které právě tyto „centralizované“ databáze mají, nicméně myšlenka spolehlivého hledání látek na volném webu, je lákavá.

---

\* Protokoly vyvíjené v rámci otevřené spolupráci, kterou koordinuje World Wide Web Consortium (W3C) za účelem vývoje a podpory standardizace služeb rodiny WWW. Jde např. o protokoly HTML, XML, RDF atd.

Výhodou tohoto řešení je to, že v podstatě neklade žádné nároky na vyhledávací stroje, využívá pouze jejich běžné funkce. Překážkou je ovšem maximální délka slova, kterou mají roboti nastavenou a jež kódy InChI mohou přesahovat. Pokud ale provozovatelé vyhledávacích strojů vyjdou InChI identifikátorům vstříc a zároveň i nakladatelé dovolí vyhledávacím strojům indexovat své dokumenty s přiřazenými kódy, bude přesnosti i úplnosti hledání stoprocentní.

## SEZNAM POUŽITÉ LITERATURY

Záznamy odkazující obecně pouze na webové prezentace jsou odděleny a číslovány.

- [ACS] *Teaching Chemical Information: tips and techniques* [online]. American Chemical Society, 1998 [cit. 23-07-2006]. Dostupný na WWW: <http://library.stanford.edu/depts/swain/hosted/cinf/workshop98aug/stratac1.html>
- [ACS, 2008b] *AnaVist Help : Research Landscape: What is It?* [online]. American Chemical Society, c2008 , 06/30/2008 [cit. 2008-10-02]. Dostupný na WWW: <http://www.cas.org/stnavist/help/v2.0/index.htm>
- [Behnke, 1990] BEHNKE, C; BARGON, J. Computer- Assisted Topological Analysis and Completion of Chemical Reactions. *Journal of Chemical Information and Computer Science*. 1990, vol. 30, no. 3, s. 228-237.
- [Brecher, 1998] BRECHER, Jonathan S. The ChemFinder WebServer : indexing chemical data on the internet. *Chimia*, 1998, vol. 52, no. 11, s. 658-663.
- [CAS5] *CAS Registry Number and Substance Counts* [online]. Chemical Abstract Service, [200?], 2006 [cit.2006-08-13]. Dostupný na WWW: <http://www.cas.org/cgi-bin/regreport.pl>
- [CAS6] *Introducing STN Viewer : a feature to enhance full-text patent evaluation and navigation* [online]. Chemical Abstracts Service, 2007, 12/12/2007 [cit. 2008-08-20]. Dostupný na WWW: [http://www.stn-international.de/archive/presentations/online\\_information07/STNViewer.pdf](http://www.stn-international.de/archive/presentations/online_information07/STNViewer.pdf)
- [CAS7] *STN AnaVist, Version 2.0 : Quick Start Guide* [online]. Chemical Abstracts Service, [200?] , 07-Sep-07 [cit. 2008-09-05]. Dostupný na WWW: <http://www.stn-international.de/stninterfaces/stnavist/pdf/anavistquickstart2.pdf>
- [CAS8] *How the Research Landscape is Created* [online]. Chemical Abstracts Service, [200?], 06/30/2008 [cit. 2008-09-05]. Dostupný na WWW: <http://www.cas.org/stnavist/help/v2.0/visualizing/understand.htm>
- [CODEN] *IEEE Xplore. Glossary : CODEN* [online]. Institute of Electrical and Electronics Engineers, [2004], 2006 [cit.2005-04-12]. Dostupný na WWW: <http://ieeexplore.ieee.org/lpdocs/epic03/xplorehelp/CODEN.htm>
- [Coles, 2005] COLES, Simon J. et al. Enhancement of the chemical semantic web through the use of InChI identifiers. *Organic and Biomolecular Chemistry*, 2005, vol. 3, s. 1832-1834.
- [Currano, 1999] CURRANO, Judith. 'Making a Sound' in Chemical Information: The Importance of a Structure Editor in Information Retrieval. *Katharine Sharp Review*. 1999, no. 8, Dostupný na WWW: <http://alexia.lis.uiuc.edu/review.old/8/currano.html>
- [Donner, 1996] DONNER, Wolfgang T. Economic Aspects of Chemical Information. *Journal of Chemical Information and Computer Science*. 1996, vol. 36, s. 937-941.
- [FIZ, 2008a] *Price List : euro* [online]. Karlsruhe : FIZ Karlsruhe, January 1, 2008 [cit. 2008-08-25]. Dostupný na WWW: <http://www.stn-international.de/service/prices/EUROen.pdf>
- [FIZ, 2008b] *STN AnaVist : 2008 Price List* [online]. FIZ Karlsruhe, c2008, 05/15/2008 [cit. 2008-10-10]. Dostupný na WWW: [http://www.stn-international.de/service/prices/anavist\\_prices.html](http://www.stn-international.de/service/prices/anavist_prices.html)
- [Fugmann, 1985] FUGMANN, Robert. Peculiarities of Chemical Information from a Theoretical Viewpoint. *Journal of Chemical Information and Computer Science*. 1985, vol. 25, s. 174-180.

- [Giragosian, 1978] GIRAGOSIAN, Newman H. Patent Information Needs from a Marketing Manager's Point of View. *Journal of Chemical Information and Computational Science*. 1978, vol. 18, no. 3, s. 121-122.
- [Glander-Höbel, 2001] GLANDER-HÖBEL, Cornelia. Searching for hazardous substances on the Internet. *Online Information Review*. 2001, vol. 25, no. 4, s. 257-260.
- [ChemZoo, 2007] *FAQs for ChemSpider* [online]. ChemZoo, [2007] [cit. 2008-09-10]. Dostupný na WWW: <http://www.chemspider.com/FAQ.aspx>
- [IUPAC] *The IUPAC International Chemical Identifier (InChI™)* [online]. International Union of Pure and Applied Chemistry, c2005, 2006 [cit. 09-07-2006]. Dostupný na WWW: <http://www.iupac.org/inchi/>
- [JCHIM] *Journal of Chemical Information and Modeling*. American Chemical Society. 2005-, vol. 45, no. 1-1x za 14 dní. Dostupný na WWW: <http://pubs.acs.org/jcics>. ISSN 1549-9596.
- [JMB] *Instructions to Authors* [online]. Elsevier, 200?, 2006 [cit. 09-07-2006]. Dostupný na WWW: [http://www.elsevier.com/framework\\_products/promis\\_misc/130606yjmibi.pdf](http://www.elsevier.com/framework_products/promis_misc/130606yjmibi.pdf)
- [Rahman, 2007] RAHMAN, Basim. *STN Viewer : more efficiency in patent information management* [elektronicky]. [Karlsruhe] : STN International, 2007.
- [Rovner, 2005] ROVNER, Sophie L. Chemical 'Naming' Method Unveiled. *Chemical & Engineering News* [online]. 2005, vol. 83, no. 34 [cit. 2008-09-10], s. 39-40. Dostupný na WWW: <http://pubs.acs.org/email/cen/html082205061024.html>. ISSN 0009-2347.
- [Ridley, 2002] RIDLEY, Damon. *Information Retrieval : SciFinder and SciFinder Scholar*. Chichester : Wiley, c2002. ISBN 0-470-84350-0.
- [STN, 200?a] *All Roads Lead to STN International* [online]. FIZ Karlsruhe, [200?] [cit. 2008-08-10]. Dostupný na WWW: [http://www.stn-international.de/guided\\_tour/roads.html](http://www.stn-international.de/guided_tour/roads.html)
- [STN, 200?b] *STN Express with Discover!* [online]. FIZ Karlsruhe, [200?] , 06/09/2008 [cit. 2008-09-07]. Dostupný na WWW: [http://www.stn-international.de/stninterfaces/stnexpress/stn\\_exp.html](http://www.stn-international.de/stninterfaces/stnexpress/stn_exp.html)
- [STN, 200?c] *Conducting a CAS Registry BLAST® search* [online]. STN International, [200?] [cit. 2008-08-05]. Dostupný na WWW: <http://www.stn-international.de/stninterfaces/stnexpress/expressblast.html>
- [STN, 2007] *STN Interface Comparison Sheet* [online]. FIZ Karlsruhe, c2008 , 09/26/2007 [cit. 2008-08-10]. Dostupný na WWW: <http://www.stn-international.de/stninterfaces/ifcc.html>
- [STN, 2008] *STN News 2/2008 : Finding information about substances that do not have associated references* [online]. STN International, 2008 [cit. 2008-10-10]. Dostupný na WWW: [http://www.stn-international.de/archive/stnews/recent\\_articles/STN%20Ask%20Reggie.pdf](http://www.stn-international.de/archive/stnews/recent_articles/STN%20Ask%20Reggie.pdf)
- [Šilhánek, 2002] ŠILHÁNEK, Jan. *Chemická informatika*. Praha : Vydavatelství VŠCHT, 2002. ISBN 80-7080-465-3.
- [UC] *Unofficial InChI FAQ* [online]. University of Cambridge. Murray-Rust Research Group, c2004 [cit. 2006-01-14]. Dostupný na WWW: <http://wwwmm.ch.cam.ac.uk/inchifaq/>
- [UTA, 200?] *Beilstein Crossfire* [online]. University of Texas at Austin, [200?], September 16, 2008 [cit. 2008-08-30]. Dostupný na WWW: <http://www.lib.utexas.edu/chem/xfire.html>
- [Vlasák, 1999] VLASÁK, Rudolf. *Světový informační průmysl*. Praha : Karolinum, 1999. ISBN 80-7184-840-9.
- [Warr, 1999] Balancing the need of recruiters and aims of the educators. In *Book of Abstracts*, 218<sup>th</sup> ACS National Meeting, 1999, New Orleans, LA, USA. Washington, DC: Wendy Warr & Associates, [1999].
- [Willet, 2005] WILLET, Peter. Searching Techniques for Databases of Two- and Three-Dimensional Chemical Structures. *Journal of Medicinal Chemistry*. 2005, vol. 48, no. 13, s. 4183-4199.
- [Williams, 2008] WILLIAMS, Anthony. ChemSpider and Its Expanding Web : Building a Structure-Centric Community for Chemists. *Chemistry International* [online]. 2008, vol. 30, no. 1 [cit. 2008-09-10]. Dostupný na WWW: [http://www.iupac.org/publications/ci/2008/3001/ic\\_chemspider.html](http://www.iupac.org/publications/ci/2008/3001/ic_chemspider.html)

### **Webové prezentace:**

1. PubChem [online]. c2006 [cit.2006-08-10]. Dostupný na WWW <http://pubchem.ncbi.nlm.nih.gov/>.
2. *NIST Chemistry Webbook* [online]. c1991 [cit.2006-10-30]. Dostupný na WWW <http://webbook.nist.gov/chemistry/>.
3. *MDL Information Systems* [online]. c2006 [cit.2006-08-10]. Dostupný na WWW <http://www.mdl.com>.
4. *Protein Data Bank* [online]. [200?] [cit.2006-08-10]. Dostupný na WWW: <http://www.pdb.org>.
5. ChemFinder [online]. Cambridge Soft, c2004 [cit.2008-09-10]. Dostupný na WWW: <http://chemfinder.cambridgesoft.com/reference/chemfinder.asp>.
6. ChemIDplus [online]. National Library of Medicine, USA, c2004 [cit.2008-09-5]. Dostupný na WWW: <http://chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp>
7. ChemSpider [online]. ChemZoo, [2007] [cit.2008-09-10]. Dostupný na WWW: <http://www.chemspider.com/>
8. Lab Crawler[online]. Beyer Internet-Beratung, 2006 [cit.2008-09-12]. Dostupný na WWW: <http://www.labcrawler.com/>
9. Chemistry Search engine [online]. Chemie.DE Information Service, 1997 [cit.2008-09-12]. Dostupný na WWW: <http://www.chemie.de/search/?language=e>
10. ChemIndustry.com [online]. ChemIndustry, 1999 [cit.2008-09-13]. Dostupný na WWW: <http://www.chemindustry.com>