



SciFinder a Reaxys – porovnání dvou nejdůležitějších chemických bází dat

Jaroslav Šilhánek

Současná reálná situace

- Z celooborového hlediska jsou nepochybně nejdůležitějším zdrojem chemických informací báze dat Chemical Abstracts – dnes nejefektivněji dostupné prostřednictvím programu SciFinder.
- Z hlediska klasické chemie organické a anorganické jsou nezastupitelné báze dat Beilstein a Gmelin – dnes dostupné jako klíčová součást databázového systému Reaxys.

V čem se liší?

- Tištěné verze

v tištěné podobě diskuse o rozdílech prakticky nepřicházela v úvahu, charakter obou zdrojů a práce s nimi byly evidentní.

- Elektronické verze

elektronické verze rozdíly stírají a umožňují doplňovat informace, které v tištěných verzích nepřicházely v úvahu.

SciFinder®

Explore References Explore Substances Explore Reactions

Welcome Jaroslav Silhanek | Sign Out

Explore References ?

Research Topic Research Topic ? Search

Examples:
The effect of antibiotic residues on dairy products
Photocyanation of aromatic compounds

- Research Topic
- Author Name
- Company Name
- Document Identifier
- Journal
- Patent
- Tags

SciFinder®

Explore ▾ Saved Searches ▾ SciPlanner

REFERENCES

- Research Topic
- Author Name
- Company Name
- Document Identifier
- Journal
- Patent
- Tags

REFERENCES: RESEARCH TOPIC

Examples:
The effect of antibiotic residues on dairy products
Photocyanation of aromatic compounds

Search

Advanced Search

Substances

- Chemical Structure
- Markush
- Molecular Formula
- Property
- Substance Identifier

Reactions

- Reaction Structure

Publication Years Examples

Document Types


REAXYS®

Query Results Synthesis Plans History Report My Alerts

Standard **Advanced**

Search

Search in Reactions Substances Literature

selected query editor:
 MarvinSketch by ChemAxon

Search as / by

- Product
- Starting material
- Reagent / Catalyst
- Any role
- As Drawn
- Substructure
 - on heteroatoms
 - on all atoms
- Similarity

STRUCTURE EDITOR

Create Structure Template from Name

Výchozí rozdíly obou zdrojů

Chemical Abstracts

- Chemical Abstracts je od svého založení v r. 1907 typickým bibliografickým zdrojem se snahou zachytit co nejširší, v ideálním případě kompletní produkci chemických a příbuzných časopiseckých a patentových zdrojů. Od samého počátku ale registruje všechny chemické sloučeniny, které se v původní literatuře vyskytují.

Beilstein, Gmelin

- Díla Beilstein a Gmelin byla koncipována jako pokud možno úplný soupis známých organických a anorganických sloučenin od počátku vědecké chemie a průběžně doplňovaný vždy k určitému datu. Ve svých tištěných podobách nikdy nebyly chápány jako průběžně vydávané referátové zdroje.

Tyto principiální výchozí koncepce se promítají i do elektronických verzí jako chemických bází dat, ale jsou postupně modifikovány jak díky technologickým možnostem, tak konkurenčním tlakem, čímž dochází k duplicitě informací.

Vyvstává proto otázka, do jaké míry jsou oba zdroje duplicitní a zda je účelné využívat a hlavně investovat do obou hlavních chemických bází dat.

Oblasti porovnávání:

- Práce s textovými informacemi, bibliografických odkazů, využívání předmětových hesel, klíčových slov apod. - „*Explore References*“, „*Literature*“.
- Vyhledávání chemických sloučenin na základě grafické reprezentace, ale i nomenklatury – „*Explore Substances*“, „*Substances*“.
- Vyhledávání přeměn chemických sloučenin – reakcí = „*Reactions*“, „*Explore reactions*“

Dotaz: *Stereoselective hydrogenation with ruthenium catalysts*

nebo: *stereoselective AND hydrogenation AND ruthenium AND catalysts**

- CA = SciFinder
- nejvyšší relevance = 29 odkazů
- průměrná relevance = 1009 odkazů
- Beilstein, Gmelin = Reaxys
- pouze jedna úroveň relevance = 34 odkazů.

Dotaz: *Analysis of phenols in waste water*

nebo: *analysis AND phenol* AND waste AND water**

CA = SciFinder:
střední relevance = 1512 odkazů

Reaxys
jedna relevance = 934 odkazů

Pozor:

Reaxys je propojen s řadou dalších chemických i nechemických zdrojů, které produkuje nakladatelství Elsevier, které ale většinou nebyly a i v současnosti nejsou zpracovávány do bází Beilstein nebo Gmelin. Takže ve výsledcích dotazu zadaného prostřednictvím předmětových hesel je velká většina odkazů na dokumenty, ze kterých nejsou do báze převzaty žádné sloučeniny ani reakce.

Typickým příkladem je báze dat Scopus, dále např. Ullmann, HoubenWeyl a řada monografií a knižních řad typu „Advances in ...“, Progress of ...“ apod.

Proč jsou tak velké rozdíly?

Záznam z díla Beilstein:

4-Brom-1-acetoxy-benzol, Essigsäure-[4-brom-phenylester], [4-Brom-phenyl]-acetat
 $C_8H_7BrO_2$, Formel XVIII (R = CO-CH₃) (H 200). B. Aus 4-Brom-phenol und Acetylchlorid (KLARMANN, Mitarb., *Am. Soc.* **55** [1933] 4657, 4661). Durch Erwärmen von Essigsäure-phenylester mit *N*-Brom-succinimid in Tetrachlormethan (BÜY-HOL, *A.* **556** [1944] 1, 7). — K_p : 100° (KL., Mitarb.). — Beim Erhitzen mit AlCl₃ auf 150–160° entsteht 5-Brom-2-hydroxy-acetophenon (KL., Mitarb.).

Do elektronické báze dat Beilstein (Reaxys) nemohly přejít žádné jiné textové informace, než jaké jsou uvedeny v záznamech u individuálních sloučenin.

Pro novější záznamy je k dispozici abstrakt, který je naprosto stejný v obou bázích.

A synthesis of (S)-N-acetylcolchicinol I (R = H), a key intermediate in the synthesis of the pharmaceutically useful phosphate prodrug ZD 6126 I (R = PO₃H₂), via catalytic asym. hydrogenation of enamide II was developed. After screening a range of metal and ligand combinations it was found that (S,S)-iPr-FerroTANE Ru(methallyl)₂ and [(S,S)-tBuFerroTANE Rh(COD)]BF₄ gave both high enantioselectivity (>90% ee) and high catalyst utility (molar S/C = 1000).

Záznam v bázích CA (SciFinder)

Kromě abstraktu obsahuje obsáhlý popis předmětovými hesly:

Concepts	
Alkaloids	
asym. synthesis of N-acetylcolchinol, a key intermediate for ZD 6126, via rhodium or ruthenium catalyzed stereoselective enamide hydrogenation	
Reactant; Synthetic preparation; Preparation; Reactant or reagent	
Stereoselective synthesis	
of N-acetylcolchinol, a key intermediate for ZD 6126, via rhodium or ruthenium catalyzed stereoselective enamide hydrogenation	
Hydrogenation	Hydrogenation catalysts
stereoselective; asym. synthesis of N-acetylcolchinol, a key intermediate for ZD 6126, via rhodium or ruthenium catalyzed stereoselective enamide hydrogenation	

894762-65-3 Q
asym. synthesis of N-acetylcolchinol, a key intermediate for ZD 6126, via rhodium or ruthenium catalyzed stereoselective enamide hydrogenation
Reactant; Reactant or reagent

Věcný (heslový) popis obsahu dokumentu

CA - SciFinder

- Každý záznam je vytvářen na základě vytištěné kopie článku nebo patentu zkušeným odborníkem se širokým spektrem znalostí a jak předmětová hesla, tak i jednotlivá klíčová slova a upřesnění role chemických sloučenin jsou tak maximálně autentické.

Reaxys (Beilstein, Gmelin)

- Obdobným způsobem jsou vytvářeny záznamy pro báze Beilstein a Gmelin jen pro vybraný okruh časopisů (cca 400) a patentů (ale jen EP, WO a US od r. 1976 a třídy C07, A01N a C09B). Výsledky dotazů zahrnují i odkazy z dalšíchází dat, které produkuje nakladatelství Elsevier, např. Scopus. Ty ale nejsou „ručně“ indexovány odborníky.

Závěry testování předmětového vyhledávání informací

- Program SciFinder pracuje s materiálem, který byl a stále je vytvářen jako bibliografická báze dat a záznamy jsou vytvářeny „ručně“ specialisty. Proto jsou výsledky dotazů formulovaných pomocí hesel výrazně relevantnější.
- Systém Reaxys vychází z děl charakteru soupisu chemických sloučenin, a i když jsou tyto báze rozšiřovány a propojovány s dalšími bibliografickými bázemi dat zatím nemohou konkurovat v případě vyhledávání předmětovým popisem problému ke srovnatelným výsledkům jako SciFinder.

Substances (Compounds) Search

- Obě báze umožňují prakticky identické nástroje, sumární vzorce, názvy(-sloví), grafické zadání.
- Obě báze pracují se stejnou koncepcí, z publikace jsou excerpovány všechny studované (použité) sloučeniny.
- CA vždy zpracovávaly co nejširší okruh periodik a patentů, Beilstein a Gmelin jen výběrově.
- Toto může být jediným důvodem pro rozdílné výsledky.

Substances

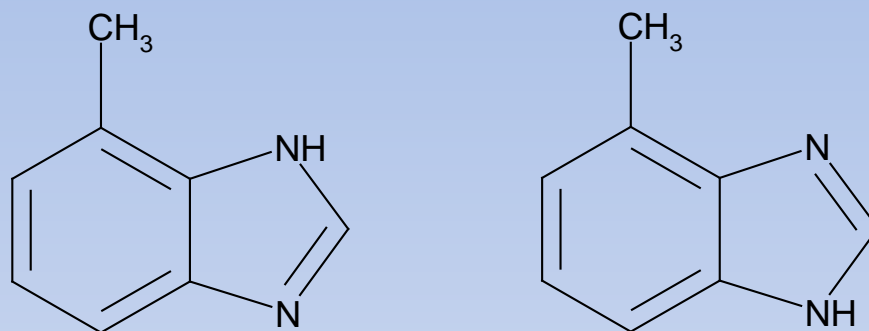
SciFinder

- 1H-benzimidazol = 9311 sloučenin, s vyloučením solí = 30
- hexakyanidoželezitan draselný jako jednoduchá sůl = 34 sloučenin
- Mirtazepine, uzavřená struktura = 16 sloučenin
- aminoferrocene = 4 sloučeniny.
- $\text{Ag}(\text{NH}_3) = 603$

Reaxys

- 1H-benzimidazol, s vyloučením solí = 18
- hexakyanidoželezitan draselný jako jednoduchá sůl = 17 sloučenin
- Mirtazepine, uzavřená struktura = 12 sloučenin
- aminoferrocene = 1 sloučenina
- $\text{Ag}(\text{NH}_3) = 97$

Tautomerism:



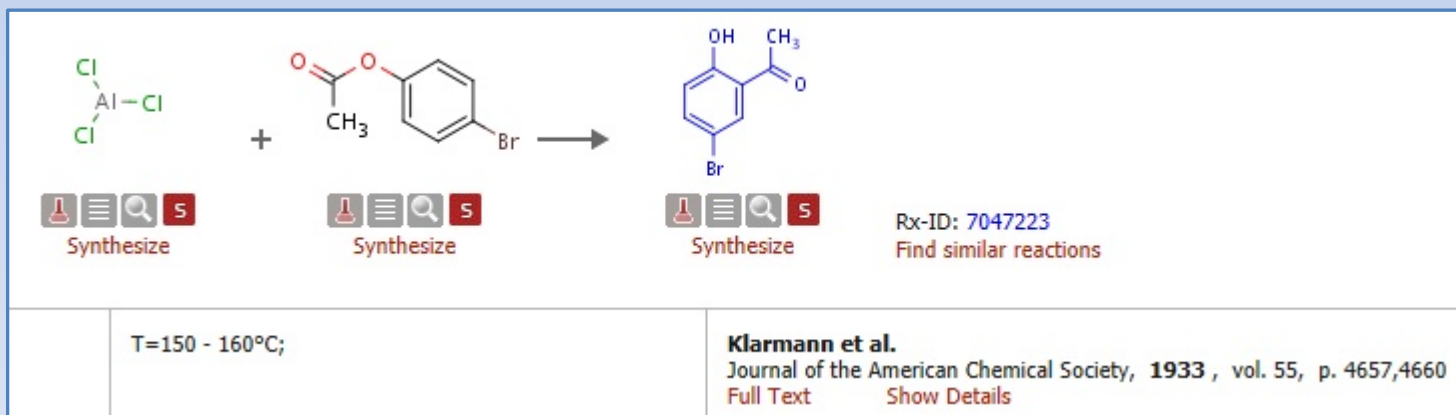
SciFinder = pouze jeden tautomer

Reaxys = oba tautomery s rozdílnými vlastnostmi

Reakce (*Explore Reaction, Reaction*)

Beilstein i Gmelin = pro každou sloučeninu jsou zaznamenány reakce její přípravy i přeměny formou věty !!!

4-Brom-1-acetoxy-benzol, Essigsäure-[4-brom-phenylester], [4-Brom-phenyl]-acetat
 $C_8H_7BrO_2$, Formel XVIII (R = CO-CH₃) (H 200). *B.* Aus 4-Brom-phenol und Acetylchlorid (KLARMANN, Mitarb., *Am. Soc.* **55** [1933] 4657, 4661). Durch Erwärmen von Essigsäurephenylester mit *N*-Brom-succinimid in Tetrachlormethan (BUU-HOI, *A.* **556** [1944] 1, 7). — Kp_2 : 100° (KL., Mitarb.). — Beim Erhitzen mit $AlCl_3$ auf 150–160° entsteht 5-Brom-2-hydroxyacetophenon (KL., Mitarb.).



Důležitý rozdíl:

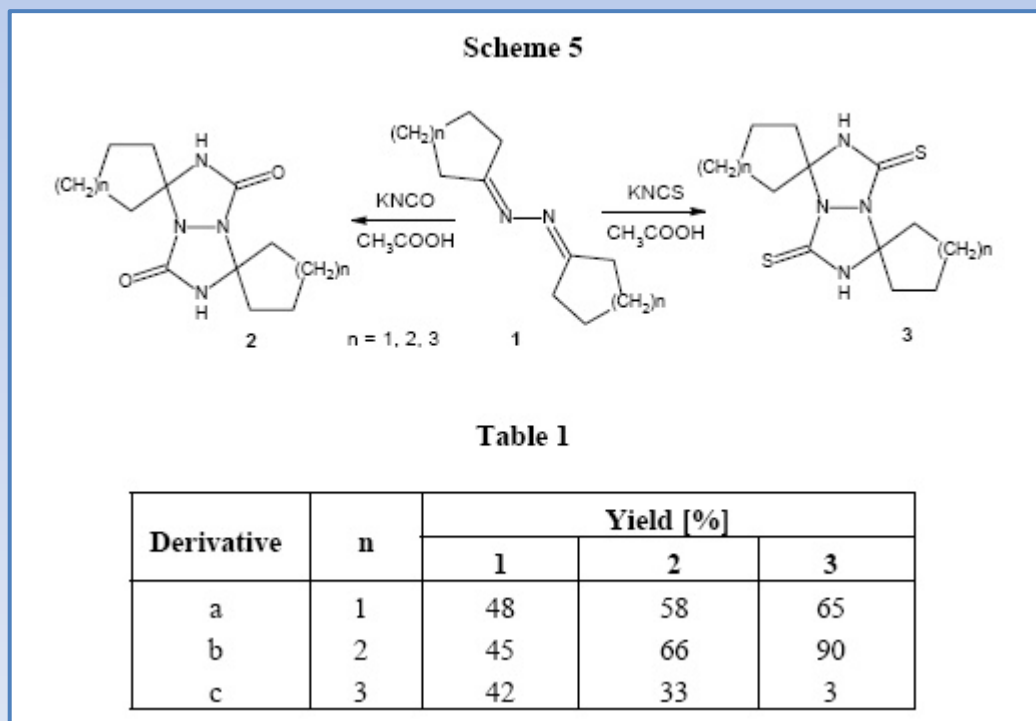
- Systém Reaxys (báze dat Beilstein a Gmelin) pokračuje ve zpracovávání se stejnou koncepcí a pokrývá tak období od 1771 do současnosti.
- SciFinder zahrnuje báze CASREACT, jejíž budování bylo zahájeno koncem osmdesátých let a zatím je jen omezeně doplňována zpětně.
- CASREACT je omezena jen na některé sekce organické a farmaceutické chemie, přírodní látky apod.
- Reaxys zahrnuje organickou i anorganickou chemii, kterou CASREACT nemá.

Ilustrace problematiky vícestupňových reakcí

Publikace:

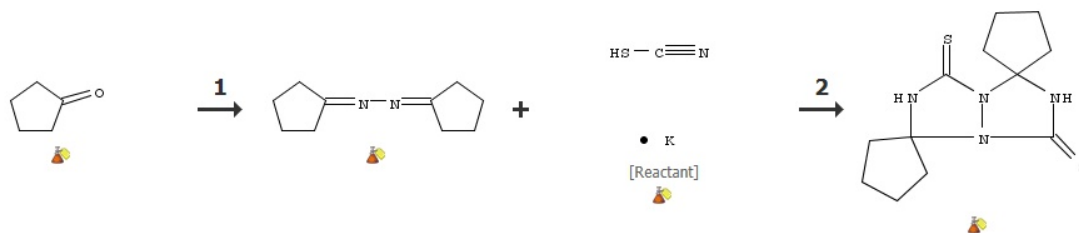
Jiří Verner, Milan Potáček

Criss-cross Cycloadditions on Ketazines Derived from Alicyclic Ketones.



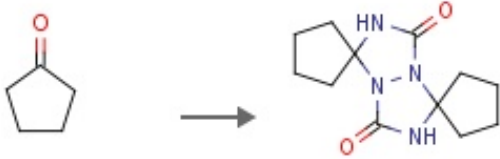


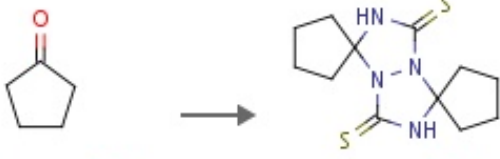


Zpracování - SciFinder

3. 2 Steps *Hover over any structure for more options.*



Steps	Stages	Notes	Yield
1	1.1 R:N ₂ H ₄ -H ₂ O, S: Benzene, 5 h, reflux	Dean-Stark apparatus used, Reactants: 1, Reagents: 1, Solvents: 1, Steps: 1, Stages: 1	48%
2	2.1 S: AcOH, 1 h, rt	Reactants: 2, Solvents: 1, Steps: 1, Stages: 1	65%

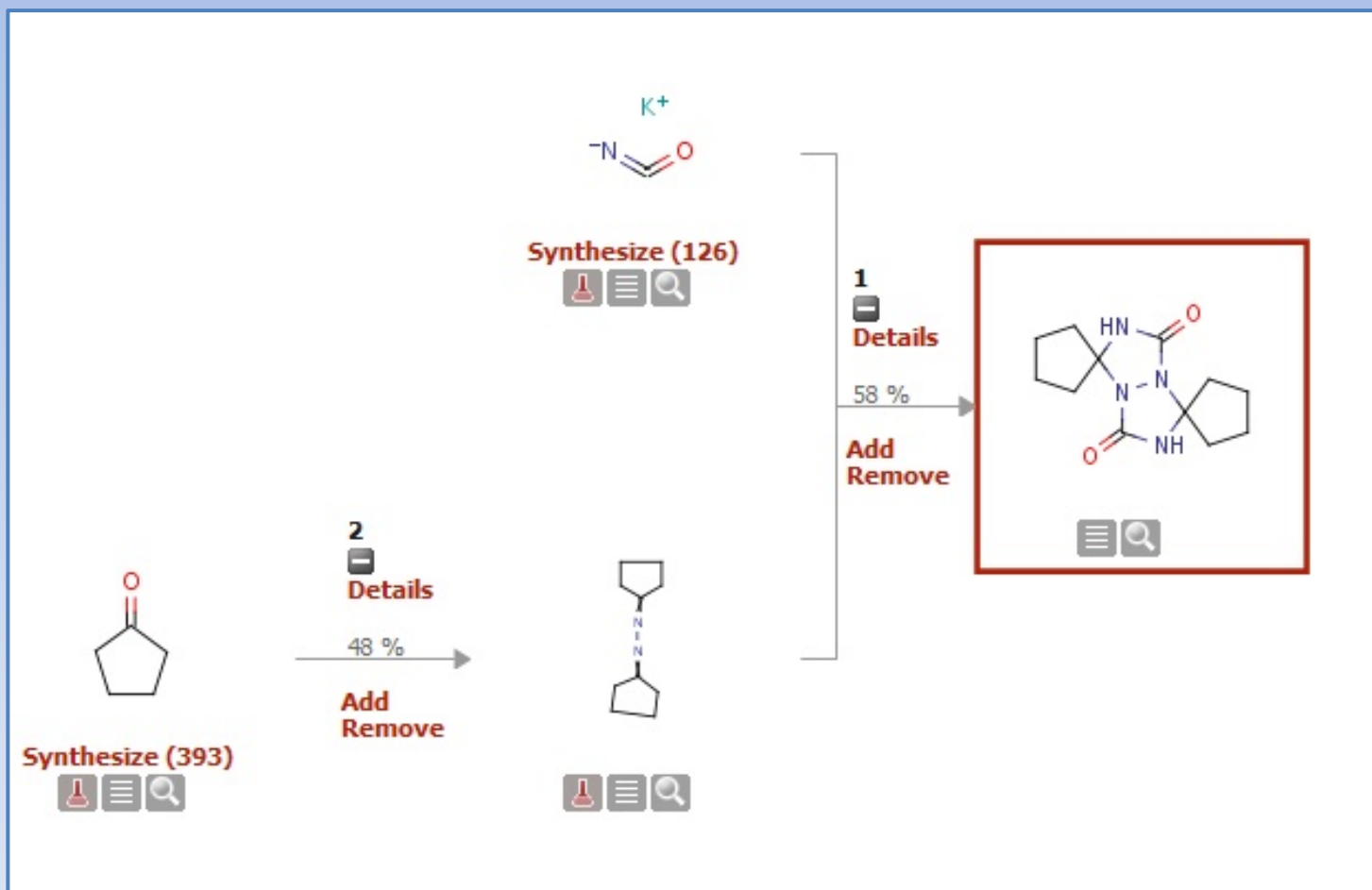
Zpracování - Reaxys

	 Synthesize	 Synthesize	Rx-ID: 12657850 Find similar reactions
Yield	Conditions		
	Multi-step reaction with 2 steps 1: 48 percent / hydrazine hydrate / benzene / 5 h / Heating 2: 58 percent / AcOH / H ₂ O / 0.25 h / 0 °C View Scheme		
	 Synthesize	 Synthesize	Rx-ID: 12657854 Find similar reactions
Yield	Conditions		
	Multi-step reaction with 2 steps 1: 48 percent / hydrazine hydrate / benzene / 5 h / Heating 2: 65 percent / AcOH / 1 h View Scheme		

Projekt ChemEIZ, OP VaVpI

CZ.1.05/3.2.00/12.0231

Zpracování - Reaxys



Principy záznamů vícestupňových reakcí

Popis experimentů v původní publikaci:

Sloučenina 1 na sloučeninu 2, tedy $1 \text{ ----> } 2$ a dále $2 \text{ ----> } 3$, nebo $2 \text{ ----> } 4$ a za alternativních podmínek $4 \text{ ----> } 5$, celkem 4 reakce.

SciFinder/CASREACT k těmto popsaným reakcím doplní dvě 2-stupňové reakce:

$1 \text{ ----> } 2 \text{ ----> } 3$, $1 \text{ ----> } 2 \text{ ----> } 4$ a $2 \text{ ----> } 4 \text{ ----> } 5$ a konečně jednu

třístupňovou reakci:

$1 \text{ ----> } 2 \text{ ----> } 4 \text{ ----> } 5$

celkem je tedy pro danou publikaci zaznamenáno v bázi SciFinder/CASREACT 8 reakcí !!!

Báze Reaxys zaznamenává reakce podobným způsobem, zatím se zdá, že méně systematicky

Výše uvedená práce Verner,Potáček: Molecules

Počet reakcí popsanych v experimentální publikace	12
Počet reakcí zaznamenaných v SciFinder/CASREACT	21
Počet reakcí zaznamenaných v bázi Reaxys	16

Anorganické reakce:

Výsledky hledání způsobů přípravy kyseliny tetrathionové v bázích SciFinder a Reaxys

	CAS RN	SciFinder	Reaxys
Volná kyselina	13760-29-7	Žádná reakce	64 prep ze 71 reakcí
Její dianion	15536-54-6	1 prep z 8 reakcí	224 prep z 334 reakcí
Draselná sůl	13932-13-3	Žádná prep z 20 reakcí	21 prep ze 75 reakcí

Závěry:

- Přes formální podobnost se jak při práci, tak i ve výsledcích projevují zásadní rozdíly výchozích tištěných zdrojů, na nichž jsou obě báze budovány.
- SciFinder profituje z relativně bohatého věcného popisu dokumentů, který je navíc vytvářen z originálních článků.
- Do jisté míry se projevují i důsledky rozdílné šíře záběru zpracovávaných dokumentů, v Reaxysu převažuje klasická organická a anorganická chemie.
- Bezkonkurenční předností Reaxysu je bohatství faktografických údajů pro všechny sloučeniny v bázi.
- Záznamy reakcí jsou srovnatelné vyjma neexistence anorganických reakcí v SciFinderu.
- Chemické reakce jsou sice v bázi SciFinder zpracovány za kratší období, ale zatím evidentně systematičtěji.
- Naproti tomu anorganické reakce zde chybějí, zatímco v Reaxysu jsou zaznamenány ve značném rozsahu.

Celkový závěr:

Je evidentní, že obě chemické báze dat se sice do určité míry překrývají, ale v daleko významnější míře se navzájem doplňují. Ještě podstatnější je skutečnost, že principiálně rozdílné koncepce vedou k rozdílné strategii řešeršních studií, což je opět mimořádně důležitý aspekt pro získání skutečně vyčerpávajících výsledků řešerší. Orientace na obě tyto nejdůležitější chemické báze dat je tak plně odůvodněná.