

# Dilemata vědeckých bází dat

Jaroslav Šilhánek

Vysoká škola chemicko-technologická v Praze

# Co to je vědecká báze dat ?

*Obecně zdroj vědeckých informací, v užším slova smyslu referátový nebo bibliografický soubor odkazů na primární vědecké prameny, případně doplněný faktografickými daty.*

## V tištěné formě:

referátové časopisy, bibliografické soupisy, kompendia, „Handbuchy“ apod.

## V elektronické podobě:

báze dat – Science Citation Index (Web of Science)  
Chemical Abstracts, Compendex, Medline,  
Excerpta Medica, Inspec, Biosis, GenBank, FSTA,  
Metadex a mnoho dalších ....

# Dilema první

*Do jaké hloubky nebo z  
jakých hledisek budou  
primární zdroje zpracovávány*

# Možnosti:

- Bibliografická citace včetně názvu
- Abstrakt (anotace)
- Klíčová slova
- Předmětová hesla
- Fráze vystihující nejdůležitější výsledky dané práce
- Faktografické údaje
- Grafy, tabulky, schémata, vzorce apod.
- Procesy, reakce

# Důsledky:

*jednou nastavená koncepce hloubky zpracovávání se jen velmi obtížně mění a má proto zásadní význam.*

## Příklad:

*v r. 1907 se rozhodl producent referátového časopisu Chemical Abstracts, že bude z každého primárního zdroje excerpovat všechny studované chemické sloučeniny.*

**Důsledek:** *Po 100 letech je proto dnes k dispozici soupis všech známých chemických sloučenin*

*Dnes to ale znamená enormní náklady*

# Dilema druhé:

*Kdo a jak bude zpracovávání  
primárních zdrojů provádět ?*

## Dvě zásadní alternativy:

- *Více či méně automatizované vyhledávání a shromažďování materiálu s minimálním intelektuálním zásahem člověka nebo vůbec bez něho*
- *Více či méně podrobné prostudování daného primárního zdroje, tedy jako intelektuální činnost člověka*



## Alternativa první:

- Typický příklad webových technologií
- Google Scholar
- Scirus
- apod.

*Plně opodstatněný přístup, ale představující naprosto samostatný a oddělený svět potenciálně vědeckých informací.*

## Alternativa druhá:

- *Stále standardní způsob vytváření vědeckých bází dat*
- *Zásadní otázkou je zodpovězení prvního dilematu*
- *Organizace přípravy dat*
- *Kvalifikace zpracovatelů dat*
- *Interní nebo externí pracovníci*
- *Stále stoupající nároky na kvalifikaci*
- *Stále stoupající náklady na tvorbu vědeckých bází dat*

# Dilema třetí:

*Rozsah zpracovávaných  
vědeckých disciplín a typů  
dokumentů*

## Možnosti:

- *Báze dat pokrývající celou oblast vědy – prakticky jen Web of Science*
- *Báze zpracovávající jeden nebo několik blízkých vědeckých disciplín – Chemical Abstracts, Medline, Metadex apod.*
- *Zahrnutí nebo opomíjení patentových dokumentů*
- *Velmi úzce zaměřené vědecké báze dat – typicky báze dat faktografických údajů*

# Dilema na straně uživatele:

- *Volit spíše široce orientovanou vědeckou bázi dat – pravděpodobnost nalezení interdisciplinárních informací*
- *Volit úzce zaměřenou nebo specializovanou bázi dat nebo více takových bází dat – pravděpodobnost nalezení relevantních informací odpovídajících očekávání*

Zdá se, že postupně začíná převažovat orientace na jen několik úžeji zaměřených bází dat, se kterými se uživatel postupně podrobně seznámí a naučí se je efektivně využívat.

## Dilema čtvrté:

*Jak bude báze dat  
zpřístupňována – jakým  
způsobem bude k ní uživatel  
přístupovat*

## Lokální nebo vzdálený přístup:

- *Báze dat je instalována na lokálním počítači, resp. na serveru v lokální síti*
- *Přístup je realizován na vzdálený server prostřednictvím sítě*



Současný vývoj jednoznačně preferuje vzdálený přístup oproti instalaci na lokálním serveru

*Tento trend spolu s inherentními vlastnostmi elektronických bází dat vede k situaci, kdy po ukončení přístupu je veškerý materiál ztracen.*

# Realizace finanční úhrady:

- *Úhrada skutečných přímých nákladů – doba připojení, počet zadaných hesel, počet „hitů“, počet zobrazených odkazů aj. – zcela univerzální, typická forma pro komerční a průmyslové instituce*
- *Paušální platba formou licenčních poplatků nebo předplatného na roční bázi – prakticky jediná možnost pro akademické instituce*
- *Kombinace obou možností*

# Dilema paušální platby:

- *Producent musí stanovit určitý minimální roční licenční poplatek, ať podle velikosti instituce nebo počtu současně pracujících uživatelů*
- *Principiálně je možné porovnat paušální úhradu s přímou platbou – v praxi hraje roli řada dalších faktorů*
- *Pro menší univerzity nebo i komerční organizace může být i minimální poplatek nepřijatelný*

# Důsledek:

*Vědecký pracovník, který nepracuje na instituci předplácející licenční přístup k té či oné vědecké bázi dat, nemá k tomuto zdroji prakticky žádný přístup. Ale i pracovníci ve výdělečné sféře, kteří nepracují v několika málo největších světových firmách udržujících silná výzkumná oddělení, nemají rovněž možnost celou řadu současných vědeckýchází dat využívat*

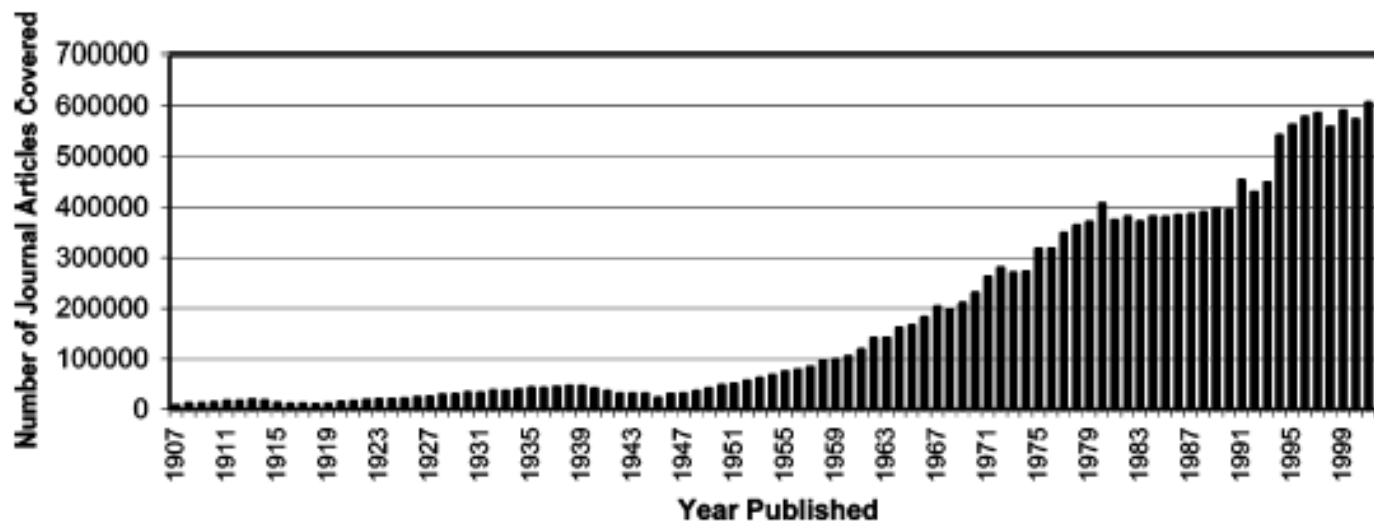
*Tištěná verze Chemical Abstracts  
(odhad 2006):*

Týdenní sešity	52 svazků
Roční rejstříky	68 svazků
Celkem	120 svazků

# Dilema páté:

*Jak řešit stávající a stále rychleji narůstající celkový objem vědeckých informací?*

# Nárůst počtu článků abstrahovaných v CA



Acc.Chem.Res. March 2004, pp. 141-148.

# Stav počtu záznamů o struktuře chemických sloučenin ke dni 19.5.2006

## The Latest CAS Registry Number® and Substance Count

<b>Date</b>	<b>Fri May 19 12:03:10 EDT 2006</b>
<b>Count</b>	<b>28,101,157</b> organic and inorganic substances
	<b>57,500,369</b> sequences
<b>CAS RN</b>	<b>884905-05-9</b> is the most recent CAS Registry Number

85,6 miliónů

Stav v dubnu 2002

37 miliónů

Nárůst 48,6 miliónů



Neustálý rozvoj vědeckých disciplín a stoupající finanční podpora vědy a výzkum vede ke ...

*... stále náročnějším požadavkům na zpracovávání výsledků vědecké činnosti do vědeckých bází dat a tím ke stále stoupajícím požadavkům na kvalifikaci pracovníků, kteří se touto činností zabývají a logicky také ke stoupajícím nákladům na tuto činnost.*

Objem dat, náklady na vytváření  
a cena za zpřístupňování budou  
proto nadále stoupat !!!

*Do jisté míry mohou tento vývoj  
regulovat grantové agentury,  
budou-li klást větší důraz na kvalitu  
vědeckých publikací oproti jejich  
kvantitě.*

# Mají vědecké báze dat budoucnost?

- *V současné době jsou nezastupitelným nástrojem vědecké práce.*
- *Představují souhrn vědeckých poznatků za velmi dlouhá časová období, a to dnes i v digitální podobě*
- *Představují velice efektivní propojení na digitální primární zdroje, což je konečný cíl vyhledávání vědeckých informací,*
- *Proto lze na položenou otázku odpovědět s největší pravděpodobností ano.*

# Realita práce s vědeckými bázemi dat:

