

Bilance prvních deseti let 21. století v oblasti vědeckých informací

Jaroslav Šilhánek

Vysoká škola chemicko-technologická v Praze

silhanek@vscht.cz

INFORUM 2010: 16. konference o profesionálních informačních zdrojích

Praha, 25. - 27. 5. 2010

Abstrakt.

Bylo-li možné posledních deset let 20. století označit jako nástup moderních informačních technologií, pak během prvních deseti let 21. století se tyto technologie staly dominantním nástrojem práce s informací vůbec. Oblast vědeckých informací, která představuje i významný segment informačního průmyslu, je pak vhodná pro pokus o celkovou bilanci tohoto období. Současně jde rovněž o bilanci desetileté systematické státní podpory přístupu k vědeckým informacím. Kromě více či méně subjektivních hodnocení lze popsat vývoj i kvantitativně na základě statistik, a to především pro nejdůležitější oblast vědeckých informací, vědecká periodika. Statistické ukazatele jsou velmi impresivní a svědčí o prakticky stoprocentním přechodu do digitálního světa včetně prokazatelně výrazně vyššího využívání vědeckých informací, znamenající podstatně zvýšenou efektivitu celého procesu sdílení a šíření výsledků vědecké práce. Současně se ale jedná i o finančně náročnou aktivitu, a to jak pro finance veřejné, tak i privátní, a tvrdíme-li, že přechod na elektronické verze je v podstatě dokončen, nelze totéž říci o mechanismech financování na straně uživatelů a návratnosti investic na straně producentů. Současná etapa je zřetelně ve stadiu hledání „business“ modelu, který by odpovídal současnému digitálnímu světu vědeckých informací. Referát se v tomto ohledu pokusí o nástin tendencí a probíhajících aktivit i o prognózu vývoje v nejbližších letech.

Úvod.

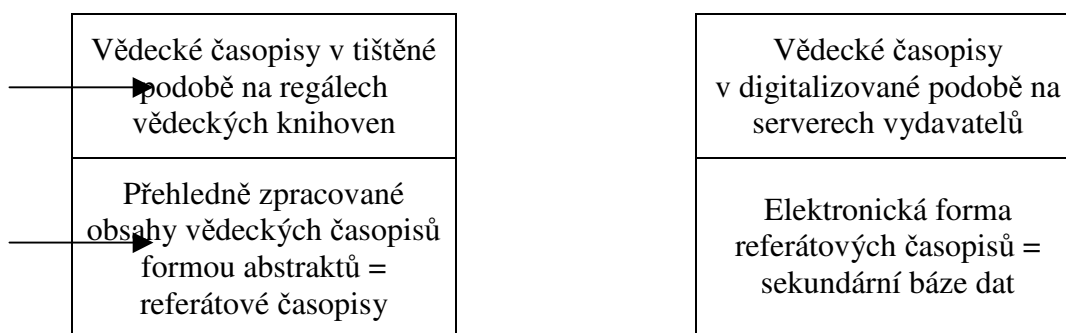
Rok 2010 je sice prvním „kulatějším“ datem v novém století, ale hlavní motivace pro pokus o sumarizaci stavu v oblasti vědeckých informací je především realita práce v této oblasti, která je dnes, *de facto* velmi přibližně po deseti letech, na úrovni, velmi výrazně odlišná od té, na jakou jsme byli zvyklí do konce minulého století. Pro pamětníky, kterých je snad ještě hodně, připomínáme, že posledních deset let minulého století se naprostá většina vědeckých pracovníků poprvé setkala s elektronickou podobou vědeckých textů a ke konci tohoto období dokonce už řada z nich posílala své práce redakcím časopisů v této podobě a hlavně se poprvé setkala s existencí počítačových sítí a pro mnohé se šokující možností získat požadovaný článek touto cestou prakticky okamžitě a odkudkoliv. Řada vědeckých oborů se také začala seznamovat s možnostmi, které digitální podoby vědeckých informací nabízí pro jejich obor, především ve většině přírodních věd. Stačí uvést ukládání a manipulaci s chemickými vzorci v elektronické podobě, digitalizovaná spektra, geografické informace nebo celé nové vědecké obory jako je dnes bioinformatika.

Přibližně lze konstatovat, že těchto posledních deset let 20. století bylo přípravným nebo snad testovacím obdobím, nejenom pro elektronické vědecké informace jako takové, ale i pro realizaci jejich vzájemných vazeb bezprostředně a logicky využívající skutečnost, že tyto

informace mohou být právě vybudovanou strukturou počítačových sítí navzájem mimořádně snadno identifikovatelné, naležitelné a jejich propojení realizovatelné řádově v sekundových časových dimenzích. Současně probíhal rychlý vývoj aplikací, jejich určitá standardizace nebo alespoň snaha pro maximální omezení problémů na straně uživatelů vyplývající z různých platforem, formátů i dalších aspektů souvisejících s digitalizovanou verzí. I když určitě nelze tvrdit, že do počátku 21. století bylo dosaženo ideálního stavu, přece jenom bylo možné zahájit práci s elektronickými vědeckými informacemi na široké platformě s jednoznačným cílem realizovat přechod od tištěných k elektronickým formám. A právě prvních deset let tohoto století je příležitostí podívat se jak daleko tento proces dospěl. Budiž ale připomenut jeden, poměrně málo respektovaný aspekt. Věda a logicky vědecké informace jsou těsně svázány s příslušným vědním oborem a jejich vnitřní struktura je na daném oboru silně závislá. Následující úvahy a bilance vycházejí z orientace autora a jsou proto zaměřeny převážně na přírodovědné, nebo spíše ještě úžeji, chemické disciplíny a uváděné závěry je nutné zohlednit prismatem daného oboru.

Vědecké časopisy

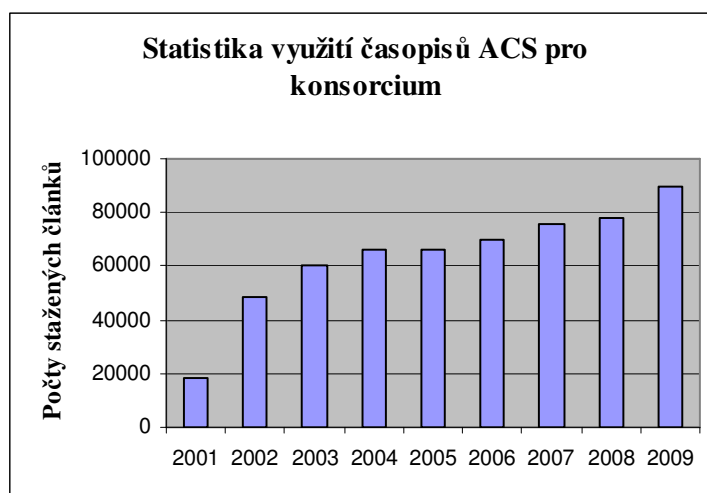
Vědecké časopisy představují nepochybně nejdůležitější složku vědeckých informací, jednoduše proto, že vědecký článek je rozumným kompromisem mezi snahou sdělit světu své nové poznatky co nejdříve a současně v ucelenější a přiměřeně rozsáhlé podobě. Je tomu tak již několik století a v žádném případě se nezdá, že by role časopisů jako souborů vědeckých článků směřovala ke svému konci bez ohledu na změnu technologie. Co se ale naprosto zásadně změnilo, je fyzická podoba vědeckého časopisu a tedy i vědeckého článku. Možnost pracovat v elektronické podobě je v této oblasti tak výhodná a efektivní, že přechod od tištěné podoby vědeckých časopisů na jejich elektronické verze je nepochybně nejdůležitější a nejvýraznější změna v každodenní vědecké praxi. Následující schéma se pokouší stručně sumarizovat tuto změnu, přičemž je nutné velmi zdůraznit, že tato změna proběhla téměř v úplnosti právě během prvního desetiletí 21. století:



Korektněji řečeno, elektronické verze referátových časopisů byly digitalizovány a zpřístupněny v řadě případů dávno předtím, ale v pravém slova smyslu bezproblémový síťový přístup k nim, který teprve umožnil v plné míře využít možnosti za dlouhou dobu nashromážděného materiálu, byl realizován až koncem 20. a hlavně v průběhu prvních 10 let 21. století.

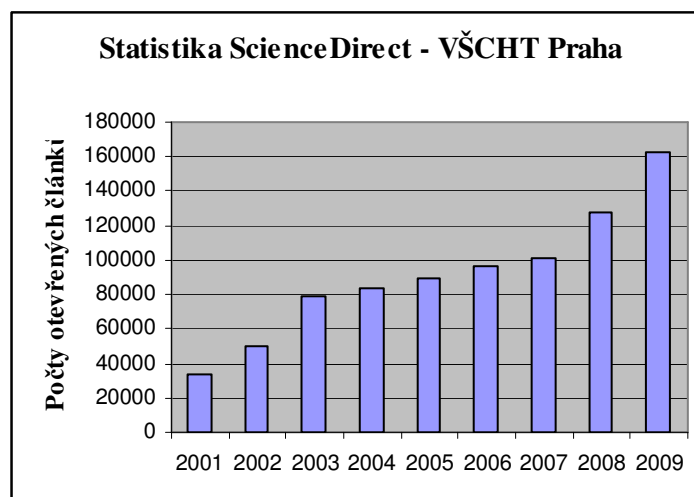
Nejnázornější ilustrace této zásadní změny je velmi markantní úbytek vědců ve vědeckých a univerzitních knihovnách, především z oblasti přírodních věd. Že tento úbytek návštěvníků neznamená úbytek využívání vědeckých informací je možné velmi snadno dokumentovat statistikami využívání elektronických verzí vědeckých časopisů. Tyto údaje mohou být dnes

získávány velmi jednoduše a současně přesně, což v případě tištěných svazků na regálech knihoven bylo možné jen nepřímo a v každém případě pouze přibližně. Vzhledem k určitým omezením, které vydavatelé uplatňují na zveřejňování statistiky, jsou k dispozici jen výjimečně statistiky pro oborová konsorcia nebo pro vlastní instituci. Následující tabulky tak demonstrují nárůst využívání elektronických verzí časopisů Americké chemické společnosti, celkem více než 35 titulů, konsorciem českých chemických fakult a vysokoškolských pracovišť (Obr. 1).



Obr. 1

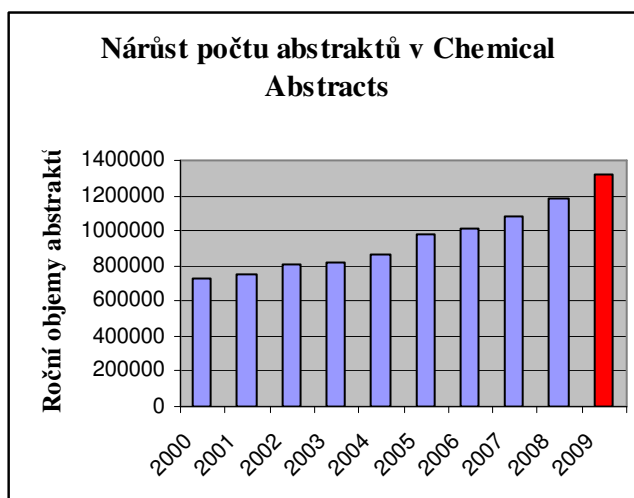
Nárůst za uplynulých 10 let 21. století je velmi markantní, i když odhlédneme od r. 2001 jako „testovacího“, ukazuje statistika prakticky nárůst 100%. Ještě dramatičtější je nárůst počtu otevřených vědeckých článků pro VŠCHT Praha a soubor ScienceDirect (Obr. 2):



Obr. 2

Tyto údaje je nutné porovnat s celkovým nárůstem objemu vědeckých článků za toto období. V zájmu oborové srovnatelnosti uveďme data pro nejdůležitější chemickou bázi dat, Chemical Abstracts. I zde je prokazatelný velký nárůst počtu zpracovaných vědeckých

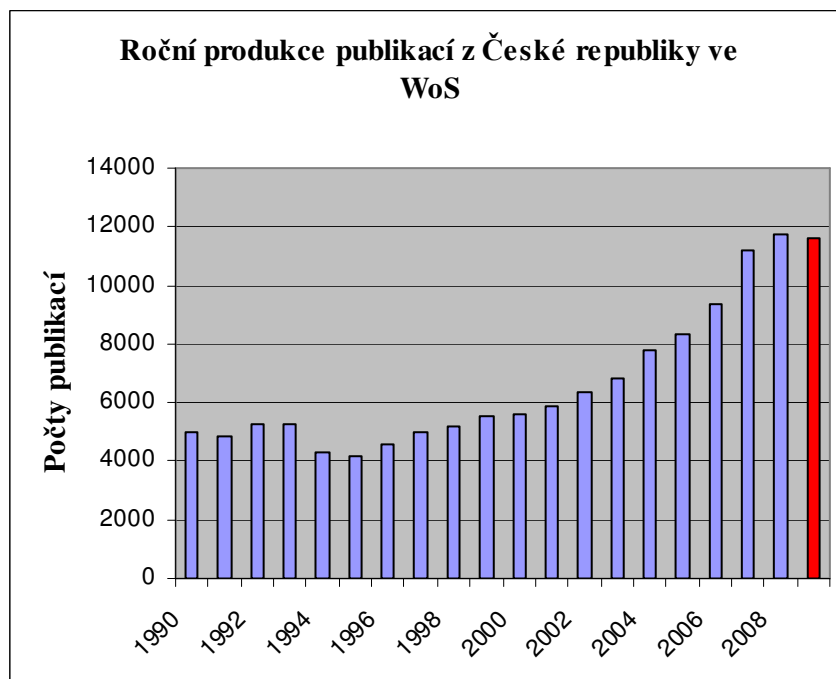
článků, přičemž je otázkou, zda využívání klasickou formou by mohlo tento nárůst absorbovat (Obr. 3):



Pro rok 2009 nejsou zatím úplná data.

Obr. 3

Celkem pochopitelně je v tomto případě nárůst relativně pravidelný, protože vzhledem k velkému počtu zpracovávaných periodik, jsou tak kompenzovány změny u vydavatelů. Můžeme doplnit tyto statistiky i z druhé strany, tj. jako nárůst publikační aktivity, v našem případě omezený na publikační aktivity vědců s adresou v České republice získaný z báze Web of Science (Obr. 4). V tomto případě je graf rozšířen i na posledních 10 let 20. století, kdy je názorně vidět určitá stagnace či snad váhání v devadesátých letech následovaná výrazným nárůstem opět během prvních 10 let 21. století.

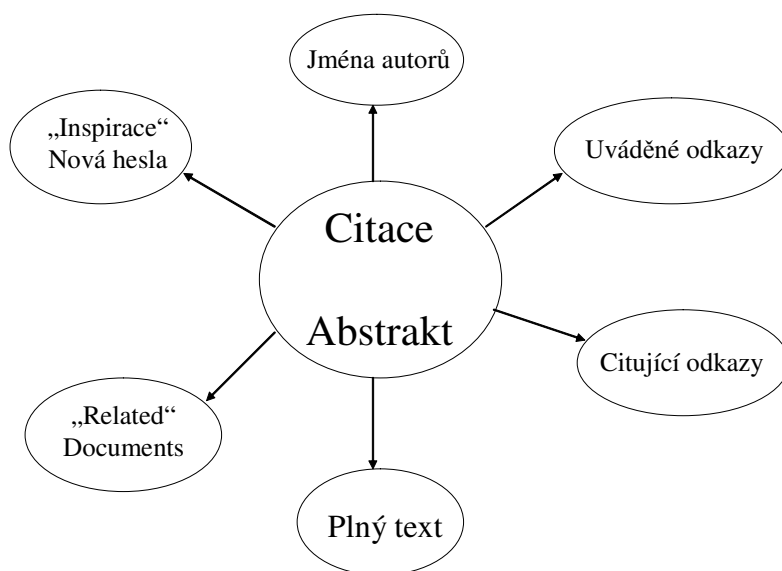


Údaj za rok 2009 není konečný.

Obr. 4

Zatímco nárůst publikační aktivity je srozumitelný a má určitě řadu důvodů, prokazatelný nárůst využívání vědeckých informací ze strany vědců je překvapující a otevírá otázku, jaký je důvod. Samozřejmě tato otázka nemůže být zodpovězena jen na základě úvah nebo i konkrétních zkušeností, ale v každém případě je možné ukázat, že se s velkou pravděpodobností nejedná jen o nárůst počtu „přečtených“ (nebo prostudovaných) vědeckých článků, ale že tento určitě potěšitelný ukazatel může mít řadu velmi důležitých aspektů.

Jedním z nich je nepochybně skutečnost, že elektronické verze velmi usnadnily vzájemné propojení informací, které ve vědeckých dokumentech vždy existovalo formou vzájemné citační praxe. Elektronická forma vědeckého článku umožňuje především velmi efektivní propojení na odkazované práce zpracované ve stejném sekundárním zdroji, tedy něco, co v době jen tištěných článků bylo možné až po získání plného textu a tam uvedených odkazů. Jména autorů sice byla přímo „viditelná“ už v sekundárním zdroji, ale znamenala poznamenat si je a přejít do jiného, autorského rejstříku. Z výsledku hledání v sekundárním zdroji bylo možné dedukovat proč byl ten či onen odkaz nalezen, ale nebylo bezprostředně vidět jakými dalšími hesly/termíny, je daný odkaz opatřen. Tato okamžitě viditelná informace je bezprostředně použitelná pro upřesnění daného souboru odkazů nebo jako „inspirace“ pro jiné zadání či pro reformulaci už použitého. A konečně pouze v elektronických verzích se mohly uplatnit nástroje jako vyhledání „related documents“ nebo okamžité nalezení dokumentů, které daný článek citují a představují tedy s velkou pravděpodobností příbuzné dokumenty. Tyto možnosti jsou schematicky znázorněny na Obr. 5 a s velkou pravděpodobností přispívají k vysokým číslu statistik na straně jedné a v každém případě ilustrují efektivitu práce s vědeckými informacemi, která byla za posledních 10 dosažena.



Obr. 5

Tyto možnosti vyplývají přímo ze záznamů uvedených v sekundárních zdrojích, tedy na úrovni abstraktu a jsou pochopitelně závislé na daném konkrétním zdroji, ale v dnešní době prakticky většina vědeckýchází dat využívá možností současných informačních technologií a uvedené cesty jsou nabízeny téměř ve všech bázích. Principiálně nemusí jít jen o vědecké zdroje v užším slova smyslu, ale o univerzální nástroje dnešních bází dat.

Daleko větší možností se nabízejí, jakmile otevřeme plný text, na který sekundární zdroj odkazuje. Zde se pochopitelně bude jednat v první řadě o vědecký text z daného vědního oboru a ve druhé řadě o to, z jakého období je předmětný vědecký článek. Starší, jen ve formátu pdf dostupné texty, mají pochopitelně menší možností než dnešní text ve formátu HTML nebo ve formátech odvozených. Na následujícím Obr. 6 jsou jen náznakově uvedeny některé typické příklady z chemických disciplín, jako jsou bezprostřední informace o experimentálních postupech, které nás mimořádně zajímají, tabelárně zpracované přehledy připravených sloučenin podle daného schématu včetně kvantitativních detailů, reakční schémata, která jsou velmi ilustrativní a vypovídají okamžitě o nejdůležitějších detailech, které bychom získali teprve po obstarání kopie článku v tištěné podobě nebo např. infračervená spektra, která jsou přímo použitelná pro bezprostřední srovnání s těmi, které jsme získali sami pro naše sloučeniny.

Následující Obr. 7 pak doplňuje možnosti ilustrací faktografických údajů pro fyzikálně-chemickou veličinu, spektrum nukleární magnetické resonance, ale i sekvenci fragmentu mRNA získanou propojením bibliografické báze na plný text a na genovou databanku GenBank. Podobných příkladů by bylo možné uvést daleko více, samozřejmě z dalších přírodovědných, ale i ostatních vědeckých disciplín, ale snad tento stručný ilustrativní výčet dostatečně dokumentuje rozsah změn v práci s vědeckými informacemi, který byl realizován na uplynulých 10 let.

Experimental Section

General Procedure for the Asymmetric Michael Reaction

The α,α -disubstituted aldehyde (**4a** or **4b**) (0.80 mmol) was added to a mixture of catalyst **3** (13.6 mg, 0.04 mmol), benzoic acid (5.9 mg, 0.04 mmol) and corresponding nitroalkene (0.20 mmol). The reaction mixture was stirred at 0°C for the requisite times as indicated in Table 1, Table 2 and Table 3. After the nitroalkene had been consumed as shown by TLC analysis, the reaction mixture was subject to flash column chromatography on silica gel (eluent: acetone/hexanes: 1/10) to afford the pure Michael product.

2,2-Dimethyl-4-oxo-2-pentenal (**4a**): ¹H NMR (400 MHz, CDCl₃): δ = 9.49 (s, 1H), 7.34–7.29 (m, 1H), 7.28–7.22 (m, 2H), 7.22–7.15 (m, 2H), 4.85 (dd, *J* = 13.1 and 11.3 Hz, 1H), 4.69 (dd, *J* = 13.1 and 4.2 Hz, 1H), 3.78 (dd, *J* = 11.3 and 4.2 Hz, 1H), 1.12 (s, 3H), 0.96 (s, 3H); ¹³C NMR (100 MHz, CDCl₃): δ = 264.1, 135.1, 126.9, 128.5, 127.9, 76.1, 48.1, 48.0, 21.4, 18.6. The enantiomeric excess was determined by chiral HPLC analysis using a Chiralcel OD-H column (i-PrOH/hexanes: 2/98), flow rate: 0.5 mL min⁻¹, λ = 254 nm; t_R = 13.6 min (major), 18.5 min (minor).

Table 2. Substrate studies of the Michael addition of **4a to various β -nitroalkenes (**5a-m**) under solvent-free conditions at 0°C.^[a]**

Entry	S	R	Product 6	Time (days)	Yield [%] ^[b]	ee [%] ^[b]
1	5a	C ₆ H ₅	6a	1.0	88	85
2	5b	4-Me-C ₆ H ₄	6b	0.5	89	84
3	5c	2-MeO-C ₆ H ₄	6c	3.0	92	71
4	5d	3-MeO-C ₆ H ₄	6d	0.5	88	85
5	5e	4-MeO-C ₆ H ₄	6e	0.5	96	83
6	5f	2-Cl-C ₆ H ₄	6f	3.0	66	52
7	5g	3-Cl-C ₆ H ₄	6g	0.5	88	85
8	5h	2-Br-C ₆ H ₄	6h	1.0	88	60
9	5i	3-Br-C ₆ H ₄	6i	0.5	89	85
10	5j	4-Br-C ₆ H ₄	6j	0.5	98	84
11	5k	3-Cl-C ₆ H ₄	6k	0.5	92	85
12	5l	4-Cl-C ₆ H ₄	6l	0.5	88	84
13 ^[a]	5m	PhC(CH ₃) ₂	6m	3.0	56	90

Tabulka připravených sloučenin

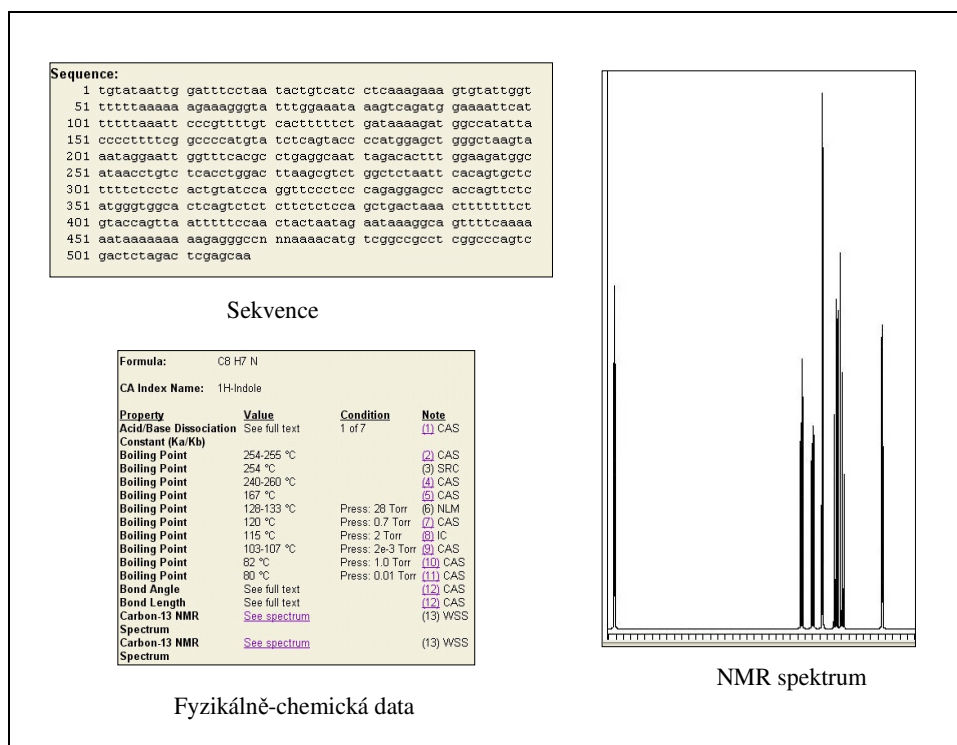
Experimentální postup

NOTES: Michael addition, stereoselective.
Reactants: 2, **Catalysts:** 1, **solvents:** 1,
Steps: 1, **Stages:** 1

Reakční schéma

IČ spektrum

Obr. 6



Obr. 7

Referátové a další zdroje vědeckých informací.

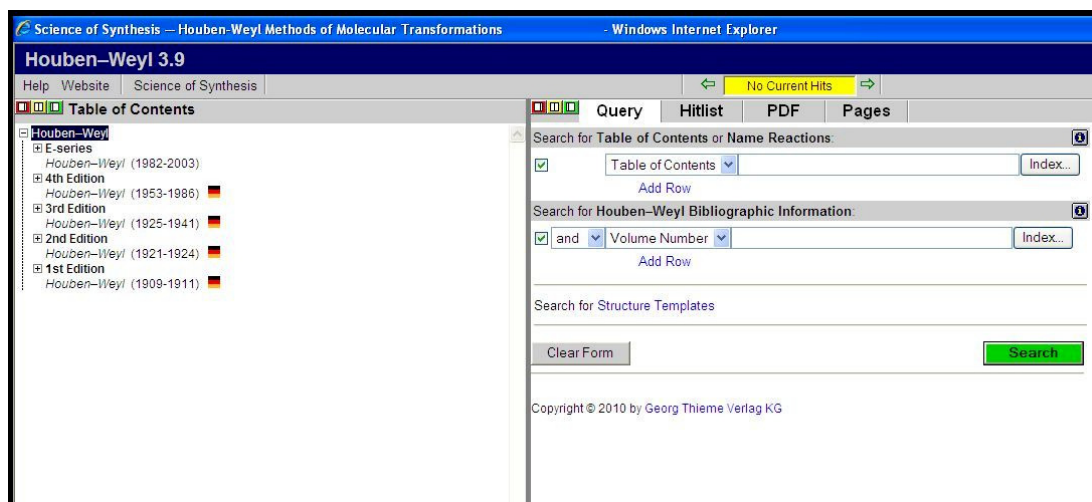
Otevřeně řečeno, v případech dalších vědeckých informačních zdrojů bychom většinou opakovali to, co bylo řečeno v případě vědeckých časopisů. Referátové zdroje typu Web of Science, Scopus, Inspec, Biological Abstracts, Chemical Abstracts, Food Science and Technology a celá řada dalších, byla přístupná v elektronických verzích prostřednictvím sítě už v průběhu 90. let minulého století. Je to logické, protože je jednalo o převážně bibliografické báze dat, kde elektronická verze je typickým využitím informačních technologií vyvíjených od samého počátku nástupu počítačové éry. V průběhu prvních deseti let tohoto století docházelo pak už k jisté nadstavbové etapě, tj. vzájemné propojování referátových, resp. sekundárních zdrojů s primárními zdroji, tedy plnými texty vědeckých časopisů a dalšími zdroji, jak bylo podrobněji dokumentováno v předchozí kapitole. Obdobně bychom mohli dokumentovat nárůst využívání prostřednictvím statistik, které by byly stejně impresivní, jako již uvedené. Skutečnost, že pro tyto vědecké zdroje už jiný než elektronický přístup prakticky neexistuje, velmi názorně dokumentuje příklad Chemical Abstracts, vůbec největší vědecké báze dat, kde v prosinci 2009 vyšel poslední tištěný sešit po 103 letech nepřetržité existence. Nadále je tento zdroj dostupný jen elektronicky, buď prostřednictvím sítě nebo na médiu CD-ROM. Ale i tato forma spěje ke svému konci.

Je ale nutné upozornit na zásadní rozdíl mezi elektronickými verzemi časopisů a referátových zdrojů, a to jejich financování, resp. úhradu licenčních poplatků za přístup a využívání. Zatímco v případě vědeckých časopisů je možná buď forma předplatného na stálý přístup v rozsahu titulů podle zájmu a finančních možností dané instituce nebo požadavek na zaslání

individuálních článků s úhradou jen za daný dokument, přístup do sekundárních zdrojů nelze principiálně dělit např. jen na části, které konkrétní obor zpracovávají. Vždy je nutné hradit formou předplatného buď celý nebo žádný přístup. Jediným vstřícným krokem ze strany producentů mohou být licenční poplatky přizpůsobené velikosti dané instituce, tedy např. univerzity. Pokud producent neumožňuje přístup ve veřejné knihovně, pak je dostupnost těchto zdrojů značně omezena.

Malou, ale důležitou kapitolou jsou nejrůznější encyklopedie, příručky, tabulky, soubory přehledných referátů apod. Přesto, že i tyto zdroje byly během uplynulých deseti let ve velké části převedeny do elektronické formy, praktické využívání si stále vyžaduje i existenci tištěných verzí. Určitým kompromisem, vycházejícím ale i z možností převodu do digitálních verzí skenováním, jsou pak elektronické verze těchto vědeckých informačních zdrojů, které obsahují jak elektronickou verzi v plném slova smyslu jako bázi dat a současně i naskenovaný obsah buď stejného zdroje nebo častěji starších svazků téhož zdroje, které nejsou převedeny do databázového modelu a mohou být využívány v pdf formátu. Na Obr. 8 je ilustrována taková situace na rozsáhlé encyklopedii známé v chemii pro názvem Houben-Weyl a pokračující jako encyklopedie Science of Synthesis, kde vidíme na levé straně nabídku všech předchozích tištěných vydání v pdf formátu a na pravé straně moderní databázový nástroj.

Otázka elektronických knih je v současné době velmi aktuální a v každém případě je důležitým výsledkem rozvoje přístupu k vědeckým informačním zdrojům během deseti let 21. století a evidentně vzhledem ke zkušenostem se elektronickými verzemi periodik se velmi rychle rozvinul, daleko rychleji než elektronické verze časopisů. Např. Springer uvádí už více než 30 tisíc E-books s ročním nárůstem 5%, Wiley uvádí zatím 8.500 titulů, ale knihovní katalog Max Planck Society uvádí už více než 40 tisíc e-books, tedy více než seznam titulů periodik. Zdá se toto tempo bude pokračovat digitalizací starších titulů, ale i do budoucna lze předpokládat paralelní existenci e-book a klasická tištěná verze, nakupovaná především z hlediska způsobu využívání, na který jsme jednoduše zvyklí a který je v řadě případů stále praktický.



Obr. 8

Shrnutí a představy o dalším vývoji.

Není pochyby o tom, že je možné snášet další a další důkazy jak dramatickým způsobem zasáhl vývoj informačních technologií metodiku, styl a hlavně efektivitu práce s vědeckými informacemi s pochopitelnými dopady na vlastní vědeckou práci. Bilance uplynulých prvních deseti let 21. století tak může být velice příznivá a záleží jen na literární licenci toho, kdo o ní bude referovat, jak vzletnými slovy to vyjádří. To, co je skutečně důležité, je skutečnost, že toto období jednoznačně prokázalo plnou funkčnost využívání elektronických forem a jejich síťového sdílení a z fáze testování a více či méně opatrného zkoušení, jsme se dostali do stadia rutinního provozu. Co tedy dál ?

Je samozřejmé, že vědecká komunita bere za samozřejmé, že dosažený stav bude dále pokračovat a bude dále rozvíjen v zásadě ve směrech, které byly v předchozích kapitolách naznačeny. Zbývá jediná otázka, a to je otázka financování celého systému, o čemž panuje jen obecné povědomí, že je to velmi drahé a že vydavatelé neúměrně vydělávají. Podívejme se na tuto klíčovou otázku alespoň trochu detailněji.

Zpracovávání, šíření a i do značné míry uchovávání vědeckých informací, je předmětem činnosti, která je buď ryze komerční, tedy zisková nebo nezisková, ale i zde musí docházet minimálně k pokrytí nákladů na celý proces. Ať tak či onak, jedná se v podstatě o průmyslovou činnost, která je dnes chápána jako Scientific Information Industry s tzv. „annual revenue“ 12 miliard USD. Tento průmysl vychází se „surovin“, což jsou rukopisy publikací, které zpracuje, doplní přidanou hodnotu a prodává jako svůj produkt. A v tomto okamžiku nastává dnes složitá situace, protože klasický model je založen na návratnosti vložených prostředků formou *ex post*, tedy předplatným, které ale vychází z dnes již téměř neexistující situace, kdy produkt, typicky vědecký časopis, stál na regálech knihoven, kam každý zájemce musel dojít. Jak to měl daleko, pak ve skutečnosti určovalo, kolik knihoven vědeckých institucí předplatí kolik časopisů a tedy jak velký bude příjem vydavatelů. Maximálně snadná dostupnost elektronických verzí tento koncept zpochybnila a to, čeho jsme byli svědky v uplynulých deseti letech, bylo hlavně hledání jiného obchodního modelu, který by zajistil více méně trvalý zdroj příjmů, aby celý systém mohl dále fungovat. Ale to už je téma tak široké, že není možné jej v této souvislosti detailněji diskutovat. Nicméně je a bude to hlavní téma následujících let 21. století.

Použitá literatura:

1. Ware M., Mabe M.: The STM Report, An overview of scientific and scholarly journals publishing, International Association of Scientific, Technical and Medical Publishers, Oxford, UK 2009.
2. Schonfeld R.C., Housewright R.: Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies, Ithaca S+R Organization 2010.
3. Hey T., Tansley S., Tolke K (Eds.): The Fourth Paradigm, Microsoft Research, Redmond, Washington, 2009.