

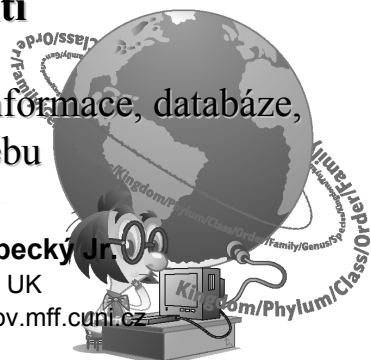
Věda v síti

aneb
vědecké informace, databáze,
etc., na webu

Vladimír Kopecký Jr.

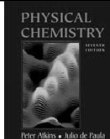
Fyzikální ústav UK

kopeccky@karlov.mff.cuni.cz



Knihy na webu

- **Učebnice vytvářené pro web**
 - Většinou jako projekty krachují nebo jde spíše o poznámky...
 - The Biophysical Society – <http://www.biophysics.org/>
 - The Sherill Group – <http://vergil.chemistry.gatech.edu/notes/>
- **Historické archivy**
 - Oregon State University Special Collections – <http://osulibrary.oregonstate.edu/specialcollections/>
- **Doplňky tištěných knih**
 - Řada nakladatelství zveřejňuje doplňkové informace k jednotlivým knihám, Java aplety, animace, testy, errata etc.
 - P. Atkins, J. de Paula: Physical Chemistry – <http://www.whfreeman.com/pchem7/>
- **Elektronické verze tištěných knih**
 - W. H. Press et al.: Numerical Recipes – <http://www.nr.com>



Bookshelf

- **Kolekce** (nejenom) **biomedicínských knih, které je možno prohledávat přes klíčová slova**
- **Ne všechny jsou kompletně volně přístupné, nejde o archiv voně přístupných knih!**
- **Nemůžete si stáhnout celou knihu v elektronické podobě, ale pouze kýženou stránku či kapitolu!**
- **Webová adresa – <http://www.ncbi.nlm.nih.gov/> pak klikněte na odkaz Search: Books**

 **Bookshelf**



National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

Vědecké časopisy on-line

- **Většina vědeckých časopisů má svůj elektronický archiv**
- Přístup bývá **většinou placen** (nutno mít příslušná hesla nebo dobrou IP adresu)
- Odběr elektronické verze časopisu je často podmíněn odebíráním tištěné verze, nebo je **cena elektronické verze přinejmenším srovnatelná s verzí tištěnou**
- Elektronická verze obsahuje **doplňková data k článkům**
- Většina časopisů má některá čísla volně přístupná jako **ukázková čísla**
- Existují časopisy vydávané společnostmi či nevýdělečnými organizacemi, které mají s jistým **časovým zpožděním přístupný kompletní archiv**
- Do budoucna dojde možná k velkým změnám...

FEBS
Letters

Jde to i jinak – videožurnály

- **Journal of Visualized Experiments (JoVE)**
- **Založen v r. 2006 jako webový recenzovaný časopis** (indexován v PubMed, ale zatím bez impakt faktoru)
- Měsíčně publikuje 15 videí laboratorních protokolů z oblasti biochemie, mikrobiologie, biologie, medicíny a i z běžné laboratorní praxe
- **Přímé sledování laboratorních postupů usnadňuje experimentální praxi!**
- **Webová adresa –<http://www.jove.com>**

jove

Věda „na cestách“ – podcast

- **Audio/video soubory vydávané seriálově a dostupné prostřednictvím webu**
- Složenina z iPod a broadcast (případně Personal On Demand broadCAST)
- Řada vědeckých časopisů, především významných – publikujících články z více oborů, vydává vlastní podcasty
- **Vědecké podcasty obsahují souhrny nejdůležitějších článků, rozhovory s autory a komentáře vědecké veřejnosti**



Science @ Direct

- Nejrozsáhlejší **archiv elektronických článků a knih** z přírodních věd, technologie a medicíny
- Články již od roku 1823 ve formátu PDF i s doplňujícími daty
- Zahrnuje **>2500 časopisů, 9,5 milionů článků** (+0,5 milionu ročně) a **10 000 knih**
- Přístup do jednotlivých časopisů je **placený institucionálně** (přístupnost je dána IP adresou)
- Nastavení je možno **personalizovat**
- Databázi lze užít i ke **sledování citací** publikovaného článku
- Články a reference na ně jsou k dispozici již ve verzích „uncorrected proof“
- **Webová adresa – <http://www.sciencedirect.com>**



SCIENCE @ DIRECT

IoP Publishing

- **Profesní sdružení** ca. 37 000 fyziků z celého světa
- Zahrnuje **IoP nakladatelství** vydávající 40 z nejprestižnějších fyzikálních časopisů
- Placený přístup (přes IP adresu)
- **IoP Select – volná databáze vybraných článků**, s týdenní aktualizací
- Řada článků je volně dostupná do 7 dnů od publikování a některé i déle
- **Webová adresa – <http://www.iop.org>, <http://www.iop.org/Select/>**

Institute of **Physics** IoP select ::::

Ingenta

- **Ingenta Institute** založen v roce 1998 jako **instituce pro sdílení vědeckých informací**
- Zahrnuje **29 000 publikací** od **260 akademických a profesionálních nakladatelství, 6000 časopisů, 23 milionů článků**
- Přístupná v ca. 14000 institucích pro ca. 25 milionů uživatelů
- Články jsou k dispozici v PDF i s doplňujícími daty
- Přístup do jednotlivých časopisů je **placený institucionálně** (přístupnost je dána IP adresou)
- **Webová adresa – <http://www.ingenta.com>**
- Nastavení je možno **personalizovat**
- Rozsáhlé možnosti vyhledávání článků



arXiv.org

- Plně automatický elektronický archiv preprintů
- Internetový server arXiv.org založen Dr. Paulem Ginspargem (Cornell University) v roce 1991
- Původně zahrnoval pouze teoretickou fyziku, dnes pokrývá: **fyziku, matematiku, informatiku, lingvistiku, neurovědy a příbuzné obory**
- **Preprinty nejsou recenzovány**, existuje pouze sbor dohlážitelů nad arXiv.org
- V mnoha oblastech je e-print užíván jako forma publikace
- **Webová adresa – <http://arxiv.org>**



arXiv.org e-Print archive



PubMed Central



- **Volně přístupný archiv biomedicínských a přírodovědeckých časopisů**
- Projekt navržen v roce 1999 Dr. Haroldem Varmusem (bývalý ředitel NIH)
- **Spuštěno v únoru 2000**, za podpory National Institutes of Health a National Academy of Sciences
- Uveřejňuje články v průměru s **půlročním zpožděním** (min. neprodleně, max. 1 rok)
- **Velká negativní odezva nakladatelství**
- **Webová adresa – <http://www.pubmedcentral.gov>**



PubMed Central
An archive of life sciences journals

BioMed Central

- **Nakladatelství sdružující volně přístupné oponované časopisy v biomedicínských vědách**
- Založeno v roce 1999
- Nakladatelství vydává více jak 199 časopisů
- Podporuje „open access“ publikování
- Své „know-how“ dává volně k dispozici
- Snaží se vyvinout nové standardy hodnocení úspěšnosti vědecké publikace a nahradit tak impact factor, sic!
- **Webová adresa – <http://www.biomedcentral.com>**



BioMed Central
The Open Access Publisher

Svobodné publikování v síti?

- Dochází k **masivní podpoře svobodného přístupu k vědeckým informacím**, především ze strany bohatých zemí
- **Co je Open Acces Publication?**
 - Autoři a vlastníci © garantují úplný a neodvolatelný volný přístup, užití a šíření své práce
 - Komplettní verze práce musí být umístěna ve veřejném (institucionálně podporovaném) archivu
- **Webová adresa –**
<http://www.biomedcentral.com/openaccess>

Open Access now

Campaigning for freedom of research information



PLoS – Public Library of Science

- **Nevýdělečné sdružení vědců zaměřené na volné zpřístupnění vědecké a medicínské literatury**
- Založeno v říjnu 2000, zakládací dopis podepsalo 34000 vědců ze 180 zemí
- Od roku 2002 je projekt **open-access publication** financován z nadačních prostředků
- V roce 2003 založen časopis **PLoS Biology** a v roce 2004 **PLoS Medicine** (vychází elektronicky i tištěné)
- **Veškeré publikační náklady nese autor (zatím ca. \$1500 za článek)!**
- **Webová adresa –** <http://www.plos.org>



Co je DOI?

10.1006/jmbi.1995.0238
prefix suffix

- **DOI = Digital Object Identifier**, tj. systém unikátní identifikace dokumentů v elektronickém prostředí
- Interpretace DOI není závislá na velkých a malých písmenech
- **Prefix** – část přidělovaná agenturou CrossRef (placená služba)
 - Prefix začíná „10.“ k odlišení DOI od ostatních implementací
 - Následuje min. 4místný číselný identifikátor (obvykle přiřazen nakladateli, není však identifikací nakladatele!)
- **Suffix** – určován nakladatelem
 - Libovolná, pokud možno logicky konzistentní kombinace znaků, u knih často ISBN, u časopisů název, rok, ročník, stránky atp.
- **Volání DOI přes webové rozhraní –** <http://dx.doi.org/DOI>
- **Webová adresa –** <http://www.crossref.org>



Look for this.
It means CrossRef
technology is at work.



PubMed

- Databáze bibliografických informací
- Zahrnuje 19 milionů citací na biomedicínské a přírodovědecké práce od roku 1948
- Vytvořen v roce 1988 National Centre for Biotechnology Information při National Library of Medicine
- Součást vyhledávacího systému Entrez
- Citace zahrnují odkazy na příslušné články a další informační zdroje (3D struktury, sekvence atp.)
- Webová adresa – <http://www.pubmed.org>

PubMed



ProQuest 5000

- Webová kolekce databází článků
 - Zahrnuje několik tisíc současných i historických periodik (časopisy, noviny)
 - Obsahuje databáze z oblasti ekonomie, techniky, medicíny a farmacie
 - Historický archiv zahrnuje 5,5 miliardy digitalizovaných stránek
- Přístup do databází je **placený institucionálně** (přístupnost je dána IP adresou)
- Informace jsou denně aktualizovány
- Většina článků (především novinových) je k dispozici v plné verzi
- Webová adresa – <http://www.proquest.co.uk>

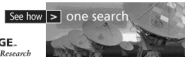
ProQuest ProQuest
COMPANY

ISI Web of Knowledge

- Webové rozhraní spojující všechny databázové a softwarové produkty firmy Thomson Reuters – <http://www.thomsonreuters.com/>
- Zahrnuje
 - Web of Science
 - ISI Proceedings
 - Derwent Innovations Index
 - Current Contents Connect
 - Journal of Citation Reports
- Integrovan je i management software – EndNote, ProCite...
- Partnery jsou např. Inspec, Biosys Previews...
- Webová adresa – <http://www.isiknowledge.com/>



ISI Web of
KNOWLEDGE
Transforming Research



Web of Science

- **Webový vyhledávač zahrnující přes 10 000 časopisů od roku 1900 a 110 000 konferenčních sborníků od roku 1990**
- **Databáze zahrnuje**
 - Science Citation Index Expanded
 - Social Sciences Citation Index
 - Art & Humanities Citation Index
 - Index Chemicus
 - Current Chemical Reactions
- **Databáze je kompletní až od roku ca. 1945**
- Ročně přibude přes 22 milionů referencí
- Umožňuje vyhledávání pomocí logických operátorů
- **Webová adresa – <http://www.isiknowledge.com/>**



ISI Web of KnowledgeSM

Take the next step

Current Contents

- **Webový vyhledávač zahrnující přes 8000 časopisů a 2000 knih**
- Vyhledávat je možné i přes „schválené“ **webové stránky** (5000 webů, 443 000 dokumentů)
- **Zahrnuje příspěvky od roku 1998**
- V každém ze 7 oborů přibude ca. 100 000 článků ročně
- Využívá se především pro rešerše
- Umožňuje vyhledávání pomocí logických operátorů
- **Webová adresa – <http://www.isiknowledge.com/>**



ISI CURRENT
CONTENTS
connect.



Powered by
ISI Web of KnowledgeSM

ISI Journal of Citation Reports

- **1963** – Institute for Scientific Information začíná na komerční bázi vydávat **Science Citation Index® (SCI®)**, ve kterém jsou zavedeny všechny základní indexy pro „měření vědy“
- **2009** – současná webová verze SCI pokrývá více než 10 000 časopisů z více než 150 vědeckých disciplín
- Webová databáze je dostupná od roku 1997
- **Databáze zahrnuje**
 - Science Edition
 - Social Sciences Edition
- **Webová adresa – <http://www.isiknowledge.com/>**



ISI JOURNAL
OF CITATION
REPORTS



Powered by ISI Web of KnowledgeSM

ISI Highly Cited

- **Volně dostupná** webová databáze **nejcitovanějších vědců** v letech 1981–1999, etc. (průběžně vždy za posledních 20 let)
- **Zahrnuje 21 vědeckých oborů** a v nich vždy **250 nejcitovanějších vědců** (ca. 0,5 % všech vědců)
- **Citovanost se měří jakou prostý součet všech citací autora za dané období**
- **Webová adresa** – <http://hcr3.isiknowledge.com/>



ISI Highly Cited.com.
Powered by ISI Web of Knowledge



Sharing Research Excellence

Scopus

- **Největší databáze abstraktů a referencí**
- **Zahrnuje 245 milionů citací, 250 milionů www stránek, 28 milionů abstraktů, 13 milionů patentů**
- Přes 18 tisíc žurnálů 5 tisíc nakladatelů (včetně 1200 Open Access, 600 sérií konferenčních sborníků a 350 knižních sérií, 3,6 milionu konferenční abstrakt)
- Vytvořen v roce 2005 nakladatelstvím Elsevier
- Přístup do databáze je **placený institucionálně** (přístupnost je dána IP adresou)
- **Abstrakty od roku 1996, reference od roku 1823**
- **60 % zdrojů mimo USA, 100% pokrytí PubMedu**
- **Webová adresa** – <http://www.scopus.com>

SCOPUS INFO

Entrez

- **Založeno v r. 1988 jako National Centre for Biotechnology Information**
- **Integruje veškeré databáze provozované NCBI a další veřejně dostupné databáze**
- **Integrované vyhledávání v databázích**
 - Strukturních
 - Sekvenčních
 - Genomových
 - Taxonomických
 - Bibliografických



Entrez, The Life Sciences Search Engine

PubChem

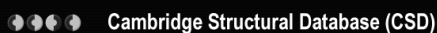
- Voně dostupná databáze malých malých molekul, založená NIH v roce 2004
- Databáze obsahuje informace o řádově milionech malých molekul
 - Názvosloví, vlastnosti, synonyma, základní komentáře, relevantní odkazy v databázi PubMed, integrace s dalšími volně dostupnými databázemi (e.g. NIAID)
- Umožňuje vyhledávat pomocí
 - Jména, 3D struktury, fragmentu, identifikátoru
 - Dokonalá podpora vyhledávání podobných struktur
- Webová adresa – <http://pubchem.ncbi.nlm.nih.gov>



PubChem

Cambridge Structural Database

- CCDC založena v roce 1965 jako nevýdělečná organizace na University of Cambridge
- Obsahuje přibližně 300 000 krystalových struktur především malých molekul (ročně přibývá ca. 20 000 nových struktur)
- Umožňuje vyhledávání pomocí ID, chemických názvů, časopisů, autorů, ale i nakreslených strukturních fragmentů
- Není volně přístupná přes web – přístup přes placené národní licence
- Webová adresa – <http://www.ccdc.cam.ac.uk>



Brookhaven Protein Databank

- Založena v roce 1971 (původně 7 struktur)
- Od r. 1998 pod patronací Research Collaboratory for Structural Bioinformatics (RCSB)
- Obsahuje ca. 60 000 struktur proteinů
 - 82 % struktur určeno pomocí X-ray krystalografie
 - 16 % určeno NMR spektroskopií
 - 2 % teoretické modely (již nelze deponovat)
- Pod PDB ID (identifikační kód) jsou uložena strukturní a sekvenční data
- Struktury molekul uloženy v PDB formátu
- Webová adresa – <http://www.pdb.org>



Nucleic Acid Database

- Založena v roce 1992 (odštěpením od PDB) za podpory National Science Foundation
- Obsahuje přes 4300 struktur nukleových kyselin (DNA, RNA, smyčky atp., s ligandy či proteiny)
- Možnost rozsáhlých způsobů vyhledávání (NDB ID, struktura, metodika, autor etc.)
- Zajímavostí je hudební atlas sekvencí
- NDB spravuje formát mmCIF (macrommolecular Crystallographic Informational File)
- Webová adresa – <http://ndbserver.rutgers.edu>



Genomové databáze

- **ExPASy** – Expert Protein Analysis Systém
 - Server provozován Swiss Institute for Bioinformatics
 - Databáze určená k analýze proteinových sekvencí, 3D struktur, 2D gelů etc.
 - Webová adresa – <http://www.expasy.org>
- **TIGR** – Institute for Genomic Research
 - Databáze soukromé nevýdělečné organizace
 - Obsahují řadu genomů různých organismů
 - Webová adresa – <http://www.tigr.org>



Databáze spekter

- Databáze spekter obsahující UV-Vis, FTIR, NIR, Raman, MS, NMR a flurescenční spektra různých molekul
- Obsahuje řádově tisíce spekter
- Založena v roce 1996 firmou ThermoGalactic
- Vyhledávání pomocí jména, ID, sumárního vzorce, molekulové váhy a porovnáním spektra
- Databáze je nyní placená systémem „pay per view“, tj. placení za vyhledání 1 záznamu
- Webová adresa – <http://www.ftirsearch.com>



▶ Ne všemu na webu se dá věřit!



MUSÍME SE SMTÍT S TĚM, ŽE NAŠÍM JEDINÝM BOHATSTVÍM
JSOU INFORMACE.

▶ Doporučená literatura

- Z. Šesták: Jak psát a přednášet o vědě. Academia, Praha 2000.
