

KNIHOVNY VE SVĚTLĚ 21. STOLETÍ: arXiv – INTERNETOVÁ DATABÁZE FYZIKÁLNÍCH ČLÁNKŮ A PREPRINTŮ

Gabriela Opletalová, Přírodovědecká fakulta, MU v Brně

1 Úvod

Elektronické informační zdroje nabývají v České republice stále větší důležitosti a v celosvětovém měřítku jsou již dlouho samozřejmostí. Internet se tak stává přirozenou a prvořadou informační bránou v širokém měřítku umožňující komplexní přístup k těmto informačním pramenům. O přínosu internetu v ekonomické, podnikatelské a všeobecné sféře v dnešní době již není možno pochybovat. Nicméně internet se díky přístupu k elektronickým informačním zdrojům stal nepostradatelným nástrojem i pro vědecké obory přírodních věd a pro vědce pracující v těchto disciplínách. To samozřejmě v první řadě znamená nutnost elektronizace hlavního zprostředkovatele těchto informačních zdrojů, kterými jsou v akademické oblasti univerzitní knihovny a dále pak on-line databáze s nejrůznějším zaměřením. Právě tato skutečnost se stává silnou motivací pro rozvoj univerzitních knihoven v České republice a dává i podnět k této přehledové práci.

Existuje několik nejvýznamnějších a nejpřednějších vědeckých časopisů zabývajících se matematickou fyzikou. Uvedme časopisy ze série *Physical Review* Americké fyzikální společnosti, časopisy Institutu fyzikálních publikací ve Velké Británii, nebo některé časopisy z nakladatelství Kluwer. Až na ojedinělé výjimky, naprostá většina českých vysokoškolských a univerzitních knihoven si tato periodika jednoduše nemůže dovolit odebírat. Nicméně především publikace v těchto časopisech určují trend vývoje soudobých vědních disciplín. Představme si knihovníka, jehož úkolem je tyto články na požádání získat. Zdá se, že je nucen pokusit se tyto publikace obstarat ze zdrojů některé z oněch několika českých knihoven, odebírajících daná periodika, anebo i z knihoven zahraničních. Tento proces získávání může trvat řekněme několik dnů nebo i týdnů, v závislosti od řady faktorů.

Ukazuje se však, že v oblasti matematické fyziky existuje jedinečný, zcela volný a kdykoli dostupný archiv nejnovějších článků, počínaje rokem 1991. Lze dokonce říci, že prakticky každá publikace za posledních deset let v těchto a příbuzných disciplínách, je obsažena v tomto ojedinělém elektronickém informačním zdroji - internetové databázi arXiv. Ačkoli je arXiv hojně využíván vědeckými a výzkumnými pracovníky univerzitních kateder a ústavů teoretické fyziky, kteří do něj i aktivně přispívají,

v širší knihovnické obci zřejmě ještě není příliš znám. Pro tyto a další důvody si databáze arXiv nepochybně zaslouží pozornost a následující strany jsou pokusem o takovéto pojednání.

Tento příspěvek je organizován následovně. V kapitole 2 se seznámíme s primárními charakteristikami arXivu a v kapitole 3 budou uvedeny základní vlastnosti a možnosti arXivu – prohlížení dokumentů a jejich zaslání na arXiv. V kapitole 4 se seznámíme s dalšími užitečnými možnostmi arXivu a kapitola 5 stručně zmiňuje návštěvnost arXivu a ukazuje grafy a statistiku přístupů. Kapitola 6 pak uvedená fakta shrnuje a podává ucelený obraz pilotní myšlenky arXivu, taktéž se zamýšlí nad jeho vývojem a perspektivami.

2 O arXivu obecně

arXiv je velmi dobře propracovaný plně automatizovaný, databázový systém článků z matematických a fyzikálních oborů. Umožňuje efektivně vyhledávat články z jednotlivých disciplín těchto oborů na základě požadovaných kritérií. Jedná se o otevřenou databázi, tzn. že do ní může v zásadě přispívat každý (tedy alespoň každý, kdo ovládá v jisté míře psaní souboru v TeXu). Jeho internetové adresa je <http://xxx.lanl.gov/>. Databáze je udržována skupinou vedenou při Cornellově univerzitě v USA. Veškerý její chod je plně automatizován. Databáze obsahuje články z následujících kategorií:

- Fyzika
- Matematika
- Nelineární vědy
- Počítačové vědy

Obsah arXivu je podřízen akademickým standardům Cornellovy univerzity. Je majetkem Cornellovy univerzity, což je soukromá nezisková vzdělávací instituce, která jej také vyvíjí a finančně podporuje. arXiv je navíc částečně financován Národní vědeckou nadací (National Science Foundation) ve spolupráci s lokální podporou Fyzikálního ústavu při Augsburské univerzitě. Protože s přibývajícím objemem růstu internetových odkazů roste též i počet autorů, kteří své články do této databáze zasílají, stává se arXiv ne-li hlavním, tedy alespoň rovnocenným partnerem klasických tištěných vědeckých časopisů a periodik.

arXiv byl zprovozněn v srpnu roku 1991. Jeho hlavním tvůrcem je P. Ginsparg. Původně byl zamýšlen jako experiment v oblasti částicové fyziky, který měl obejít nesnáze spojené s odbornými výzkumnými časopisy, především dlouhou dobou mezi zasláním článku a jeho publikováním v tištěné podobě. Radikální přijetí arXivu mezi členy vědecké komunity ovlivnilo jak zefektivnění přenosu informací pomocí internetu, tak i další technické a vědecké výhody, jak bude níže popsáno. Zprvu se jednalo o archiv článků z fyziky vysokých energií a fyziky elementárních částic,

nazvaný hep-th (high energy physics – theory). Postupně se přidávaly další fyzikální a matematické disciplíny spadající do jedné ze čtyř kategorií vypsanych výše.

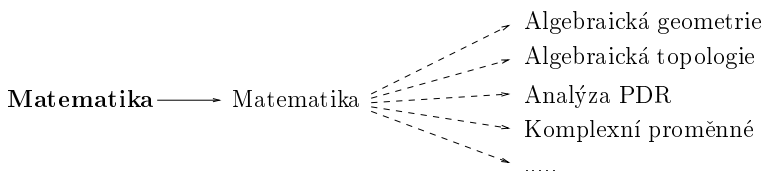
Redakční rada. Redakční radu arXivu tvoří dvě skupiny – pro fyzikální vědy a matematické vědy – a je sestavena ze známých vědců z těchto oblastí. Kromě toho má arXiv separátní redakční komisi.

3 arXiv – jeho základní vlastnosti

3.1 Struktura arXivu

V tomto odstavci uvedeme základní strukturu arXivu, tj. především to, jak je arXiv organizován po formální stránce. Jak již bylo řečeno, arXiv obsahuje články ze čtyř základních kategorií: fyziky, matematiky, nelineárních věd a počítačových věd. Každá z těchto kategorií obsahuje jednu nebo více sekcí (Fyzika jich obsahuje 12, ostatní po jedné). Dané sekce se pak mohou dělit dále na podsekce, tzv. subject-classes.

Příklad takovéto stromové struktury je na obr. 1.



Obr. 1: Příklad struktury členění článků z kategorie Matematika na sekci a její podsekce.

Stručně řečeno, arXiv jakožto elektronický informační zdroj poskytuje dvě hlavní služby. První z nich je prohlížení a stahování dokumentů, které jsou součástí této databáze. Druhou je možnost zaslání vlastního článku, respektive příspěvku do databáze. Dále pohovoříme podrobně o každé z těchto možností.

3.2 Zasilání článku

Možnost zaslání článků do databáze je otevřena každému a každý den po 24 hodin. Nicméně se tento příspěvek neobjeví na internetu okamžitě, ale až v 1 hodinu ráno středoevropského času v pracovní dny. Průměrně se denně objeví více než 100 nových příspěvků.

Necenzurované články. Články elektronicky zasílané na arXiv především nejsou cenzurované, není zde nikdo, kdo by daný příspěvek mohl zamítnout a nepřipustit k zobrazení na Internetu. To by se na první pohled mohlo zdát riskantní, neboť do arXivu pak může poslat každý co je mu

libo bez rizika ohrožení. Paradoxně však právě tato volnost dává databázi arXiv punc článků zpravidla vysoké kvality a nikdo ji nezneužívá.

V uplynulých letech byl postup většinou takový, že autoři posílali své články na arXiv teprve potom, až byly tyto práce přijaty do tisku. Dnes se již tento postup obrací: autor dopíše verzi rukopisu, kterou považuje za vhodnou k odeslání do zvoleného časopisu. Poté tuto verzi "zavěsí" na arXiv a čeká jeden až dva týdny na případnou odezvu jiných autorů publikujících popřípadě sledujících arXiv. A teprve poté, zaslá autor svůj článek oficiálně do onoho vybraného časopisu.

Mimochodem přítomnost článku na arXivu zjednodušuje i proces zaslání článku. Některé časopisy (např. Physical Review nebo Classical and Quantum Gravity) akceptují i alternativní formu zaslání článku, kdy autor uvědomí e-mailem redakci daného časopisu, že článek s daným číslem v arXivu nabízí k publikaci. Příslušný člen redakční rady pak tento článek stáhne z Internetu, a přitom má jistotu, že v něm nejsou technické chyby, které by TeXovský procesor nepřeložil.

Především je třeba vyjasnit v jakém formátu lze do arXivu posílat a v jakém formátu je stahovat (či prohlížet). Ve shodě s publikačními standardy naprosté většiny vědeckých periodik je formát článků přípustných pro uveřejnění na arXivu jednoznačně stanoven na TeXovské dokumenty. Jsou tedy přípustné články psané v LaTeXu 2.09, LaTeXu 2 ϵ , PlainTeXu atd. Alternativně je možné zasílat články přímo v PostScriptu či PDF a také jako soubory typu HTML. Nicméně protože většina příspěvků na arXivu je potažmo určena pro publikaci v tištěných recenzovaných časopisech, využívá mnoho autorů speciálních TeXovských stylů určených pro tato periodika. Autoři zasílají do databáze úplný zdrojový text včetně všech dalších doplňujících souborů, jsou-li nějaké (obrázky, vlastní makra, atd.). Poté TeXovský procesor na serveru arXivu dokument přeloží a automaticky vygeneruje soubory DVI, PostScript a PDF. Ty budou spolu se zdrojovým textem volně ke stažení. To je obrovská výhoda tohoto archívu, neboť většina článků zde uveřejněných bude nebo byla publikována v tištěné formě. Prakticky lze říci, že alespoň v ohledu na teoretickou fyziku, prakticky každý článek, uveřejněný v tištěné podobě, lze volně stáhnout z arXivu.

Vraťme se však k samotnému odesílání článků na arXiv. Rozhodne-li se někdo, že zašle svou práci na arXiv, navštíví adresu

<http://lanl.arXiv.org/help/submit>

kde najde základní pokyny pro odesílání článků. Nejprve je zapotřebí se zaregistrovat. K tomu je určen jednoduchý formulář se základními údaji, které je třeba vyplnit - jméno, e-mailová adresa, instituce a dále pak jedna ze sekcí, kam daný nový uživatel zamýšlí posílat své články. Uživatel si rovněž zvolí své heslo, které mu umožní přístup k jeho článkům. Jako potvrzení (a zároveň kontrola zadané e-mailové adresy) mu ihned přijde

automaticky generovaný e-mail obsahující ověřovací kód. Tento kód pak uživatel použije k navštívení www stránky s jeho vyplněnými údaji.

Dostupné jsou 3 základní možnosti odesílání:

- Přes web. Tento způsob je preferován a je i nejjednodušší. Proto se o něm zmíníme podrobněji. Pokud je autor zaregistrovaný, může přikročit přímo k odeslání článku na adrese <http://arXiv.org/uploads>. Zde je nejdříve vyzván, aby vyplnil své jméno a heslo, které použil při registraci. Poté je veden na stránku, kde:
 - A Musí vyplnit do kterého archívu je článek určen a eventuelně může změnit některé své registrační údaje (např. kategorii)
 - B Vyplní kolik souborů bude elektronicky zasílat. Např. obvykle daný článek obsahuje jeden zdrojový texovský soubor a pak může obsahovat několik obrázků, tedy souborů typu .eps apod.

Po vyplnění těchto údajů a jejich odeslání se objevuje formulář, kde je třeba vyplnit pole jako autor, název článku, abstract a některé základní údaje. Pomocí browseru pak vyhledá dané soubory a celý formulář odešle. Tvůrci arXivu mysleli v této fázi na to, že mnohdy na poslední chvíli udělá odesílající člověk chybu z nepozornosti. Např. gramatickou chybu v abstraktu apod. Proto se po stisknutí tlačítka odeslání článku objeví stránka arXivu ukazující, jak budou autorem vyplněné údaje zobrazeny na Internetu. Zejména lze získat PostScriptový soubor vygenerovaný z jeho TeXovského zdrojového souboru. Pokud je autor spokojen, potvrdí odeslání článku, čímž je vše dokončeno. Jako potvrzení přijetí článku pak obdrží autor automaticky generovaný e-mail, v němž jsou stanoveny podmínky publikace a dále e-mail obsahující ID a heslo k příslušnému článku ¹To je nezbytné tehdy, bude-li v budoucnu autor chtít změnit (tj. nahradit) svůj článek, provést modifikace, nebo jen přidat komentář.

- Pomocí FTP.
- Zasláním e-mailu, přičemž soubory potřebné k vygenerování článku v PostScriptu jsou přiloženy jako attachment.

Doplňování informací a nahrazování článků. Jak jsme si řekli, mnoho článků zasílají jejich autoři na arXiv ještě před tím, než tyto články oficiálně odešlou do tištěného recenzovaného periodika. To v praxi též znamená, že autoři mají možnost své články přepracovat, doplnit o nové výsledky, popřípadě dodatečně informace apod. Na arXivu je k tomuto účelu zřízen formulář, který umožňuje daný článek nahradit novější verzí. Nahrazování probíhá analogicky jako samotné odesílání článků. Přitom se stará verze článku smaže, nelze mít pod jedním číslem dvě různé verze, každá jiná verze by byla považována za samostatnou publikaci. Je-li článek

¹ K ID článků se ještě vrátíme v 3.3

úspěšně nahrazen, objeví se u jeho legendy na internetu kolonka, do níž autor může poznamenat krátký komentář o provedených změnách. Dále je též možné dodatečně přidat přesnou citaci článku v daném časopise (to v případě, kdy článek nebyl ještě publikován v době jeho zaslání na arXiv).

3.3 Prohlížení a stahování článků

Prohlížení článků dle jejich data zaslání. Hlavní stránka arXivu nabízí výběr ze všech čtyř hlavních kategorií a jejich sekcí. Specifikací jedné z těchto sekcí v dané kategorii jsme pak vedeni na stránku, kde lze zadat rok publikací, jež nás zajímají, a na další stránce taktéž měsíc vybraného roku, kdy byly příspěvky do arXivu zaslány. V této fázi se nám konečně zobrazí příspěvky v chronologickém pořadí tak, jak byly přijaty serverem. Přesněji, každý článek má své číslo dle sekce, ke které náleží, roku a měsíce, v němž byl zaslán, a dále dle jeho pořadí mezi danými příspěvky patřícího měsíce. Například může být článek referován jako gr-qc/0301023, ukazující, že článek náleží do sekce General Relativity and Quantum Cosmology (gr-qc), a byl zaslán v lednu roku 2003. Typická reference pak vypadá jako na obr. 2:

hep-th/0204212 [abs, ps, pdf, other]:

Title: Living with Ghosts

Authors: S.W.Hawking, Thomas Hertog

Comments: 13 pages; typo corrected, reference added\\

Journal-ref: Phys.Rev. D65 (2002) 103515

Obrázek 2: Příklad zobrazení reference k danému článku.

Prohlížení dle alternativních kritérií. Pokud někdo sleduje arXiv denně, nepotřebuje vyhledávat články dle jejich data zaslání, ale zajímají ho především nové příspěvky. K tomuto účelu zavedli autoři arXivu alternativní relativní členění článků na nové a nedávné. Nové jsou ty, které se objevily na internetu v aktuální den. Nedávné články jsou ty, jež byly zaslány nejdéle zhruba před týdnem. Výhoda u této alternativní formy je ta, že u každého nového článku v dané sekci je uveden i jeho abstrakt.

Nahrazené články a křížové reference. Prohlíží-li si někdo nové články v jisté sekci, zobrazí se navíc ještě dvě další skupiny článků. Jsou to články nahrazené předešlý den a dále křížové reference. Těmi míníme příspěvky z jiných sekcí, které mají nějaký vztah či vazbu k sekci právě prohledávané. Tedy každý pracovní den se nám současně zobrazí články nové (v dané sekci), nahrazené (v dané sekci) a křížové reference (příbuzné články z jiných sekcí). Toto štěpení je uživatelsky velmi nápomocné, navíc pak

odpadá nutnost prohledávat množství sekcí kvůli článkům z oblasti, kterou se sami aktivně zabýváme.

Stahování článků. Databáze arXivu obsahuje k červenci 2003, za 12 let své existence, přes 241 000 příspěvků. Řekli jsme si, že všechny dokumenty na arXivu jsou volně ke stažení. Zajímá-li nás daný článek a nalezneme na něj referenci, zobrazí se tato tak jak jsme uvedli na obr. 2. Nyní poklikáním na příslušný hypertextový odkaz si lze přečíst abstrakt nebo volně stáhnout (či prohlédnout) PostScriptovou či Pdf verzi článku² Poklikáním na odkaz „other“ se zobrazí další formáty dostupné ke stažení. Těmi je jednak DVI soubor a také samozřejmě zdrojový text se všemi obrázky atd.

Zůstaneme-li u příkladu z obr. 2, dodejme, že poklikáním na jednoho z autorů práce se zobrazí stránka obsahující všechny autorovy příspěvky v arXivu.

4 arXiv – výhody a další vlastnosti

Vyhledávání v arXivu. Jako každá moderní databáze publikací, arXiv disponuje pokročilým vyhledávacím systémem s řadou sofistikovaných funkcí. Zde se stručně zmíníme o základním vyhledávání. V principu lze v arXivu rozlišit dva základní typy vyhledávání. První je vyhledávání na základě specifikace popř. zúžení názvu článku, autora a samozřejmě zúžením na konkrétní disciplínu databáze. Na internetovou stránku formuláře jsme vedeni po stisknutí tlačítka „find“, které je umístěno na každé stránce. Jména autorů a názvy článků zadáváme jako obvykle pomocí regulárních výrazů a booleovských operátorů. Spíše než opakovat přehled používání booleovských operátorů, popř. regulárních výrazů ukažme na příkladě odlišnost oproti běžným vyhledávacím procesorům. Protože se jedná o databázi fyzikálních, matematických a počítačových publikací, museli autoři arXivu zajistit možnost vyhledávání některých matematických výrazů.

Druhým typem je vyhledávání na základě znalosti přesného čísla článku. Formulář k tomuto způsobu se objeví po stisknutí tlačítka „form interface“. Lze vyhledávat mezi novými články, články nahrazenými, taktéž je možno zobrazit například všechny autory publikující v daném měsíci v konkrétní sekci.

Citační statistika. Další velmi vítaný nástroj ocení všichni, kdo chtějí hlouběji proniknout do problematiky spadající pod některou ze sekcí na arXivu. Je nepopiratelné, že pokud chce kdokoli sledovat vývoj jisté disciplíny až k současnému stavu vědění, musí nezbytně projít a prostudovat množství alespoň těch nejpodstatnějších a nejnovějších článků, které se touto oblastí zabývají. Zde hrají životně důležitou úlohu citace uváděné

² Přesněji se stahuje komprimované soubory typu *.ps.gz nebo *.pdf.gz..

v každé práci, tj. odkazy na další práce z problematiky a jejich výsledky. A právě databázový systém arXivu nabízí efektivní a propracovaný systém zacházení s citacemi. Zejména jsou to dva typy plně automatizovaných služeb:

1) Zobrazení všech citovaných článků v dané práci. Navíc pokud jsou tyto články dokumentovány v některé z databází spojených hypertextově s arXivem, pak se zde tento hypertextový odkaz na tyto zdroje objeví. Těmito hypertextově propojenými databázemi jsou zejména stránky:

- Velké databáze fyzikálních a matematických článků – SPIRES HEP Literature Database. Zde je možno nalézt citace a reference k více než půl miliónu článků z oblasti částicové fyziky, teoretické fyziky, fyziky vysokých energií a příbuzných oblastí od roku 1974. Internetová adresa databáze SPIRES je

<http://www.slac.stanford.edu/spires/hep/>

Filozofie databáze SPIRES je sice poněkud odlišná od filozofie arXivu, nicméně tyto dvě databáze jsou těsně propojeny. Lze říci, že SPIRES obsahuje reference publikovaných článků a nabízí možnost jejich přímého stažení v případech, kdy jsou tyto publikace volně dostupné (což je téměř výlučně případ, kdy byl daný článek zaslán do arXivu.) Databáze SPIRES využívá indexování článků dvěma internetovými archívy: SLAC Library (Stanford Linear Accelerator Centre, <http://www.slac.stanford.edu/library/>) a německou databází DESY (Deutsches Elektronen-Synchrotron, viz <http://www-library.desy.de/homepage.html>).

- Americké fyzikální společnosti – APS (American Physical Society), která publikuje časopisy ze série Physical Review. Internetová adresa je <http://publish.aps.org/>.
- Institutu fyzikálních publikací - IOP (Institute of Physics Publishing), která publikuje několik desítek odborných fyzikálních časopisů. Internetová adresa je <http://www.iop.org/>.

2) K dispozici je dále citační služba, která umožňuje najít všechny články na arXivu, které náš daný článek citují. Za to jsou odpovědné mimo jiné křížové reference mezi jednotlivými sekcemi.

Tyto služby jsou pro aktivní autory velmi cenné, navíc silně povzbuzují internetovou komunikaci mezi členy vědecké komunity. Neocenitelná je právě možnost nalézt požadovaný článek, který je citován jistou prací, v několika okamžicích (za předpokladu, že je volně dostupný), nebo alespoň získat o něm základní informaci (např. abstrakt, který většina časopisů na svých internetových stránkách zpravidla uveřejňuje).

Dostupná makra, které lze stáhnout na arXivu. Jak jsme již uvedli, články do arXivu se zasílají v podobě TeXovských souborů a jsou auto-

matically přeloženy TeXovským procesorem, odtud je článek vygenerován v PostScriptu, popřípadě ve formátu PDF. Naprostá většina příspěvků je zamýšlena k publikaci v některém z vědeckých periodik, a proto zpravidla využívá speciálních TeXovských stylů těchto časopisů. Kdyby měl každý autor zasílat do arXivu kromě samotného zdrojového textu (a eventuelně obrázků) i tyto styly, byly by servery arXivu silně přetíženy. Proto jsou už dopředu implementovány všechny význačné a známé styly konkrétních vědeckých časopisů. Pouze pokud autor využívá vlastní makra, zašle je i s TeXovským zdrojovým textem.

Skutečnost, že arXiv disponuje naprostou většinou TeXovských stylů, je naopak velmi cenná i pro ty, kteří si daný článek chtějí stáhnout přímo jako textový soubor a vygenerovat PostScript či PDF samostatně. Všechna makra a styly jsou totiž na arXivu volně ke stažení, takže se nemůže stát, že by kdokoli nebyl schopen např. PDF vygenerovat. arXiv tak slouží zároveň jako archív mnoha zefektivnění, stačí si vybrat.

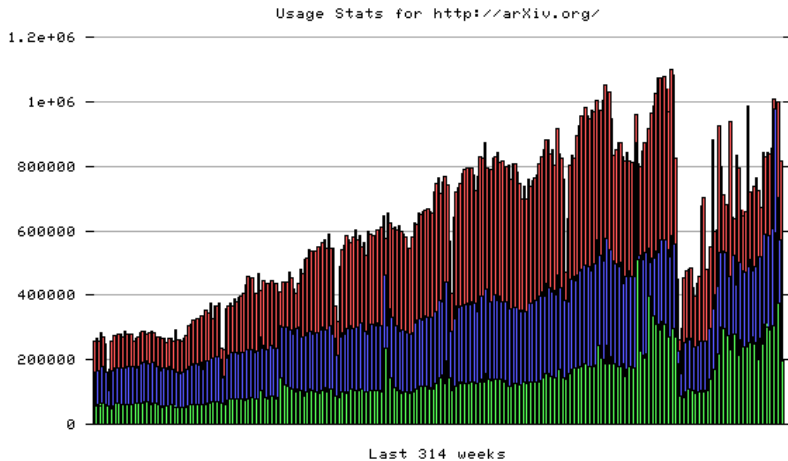
Automatické zasílání nových příspěvků. Další službou dostupnou pro každého uživatele je možnost automatického zasílání e-mailů obsahujících seznam nových článků (včetně aktuálních nahrazených článků a křížových referencí, jak jsme uvedli v kapitole 3). Struktura tohoto e-mailu je stejná jako v běžném grafickém rozhraní, prohlížíme-li nové články, tedy se zobrazují i abstrakty těchto příspěvků.

Zrcadla arXivu. Vzhledem k velkému množství dat na arXivu je přirozené, že čas nutný na připojení k databázi arXiv a hlavně čas nutný ke stažení vybraných příspěvků může znamenat v praxi i desítky sekund, u vzdálených serverů popřípadě i několik minut. Proto fyzikální společnosti některých zemí podporují tvorbu zrcadel arXivu v zájmu zrychlení přenosu dat do těchto a blízkých zemí. V současnosti existuje 17 zrcadel arXivu umístěných v zemích celého světa.

5 Návštěvnost arXivu

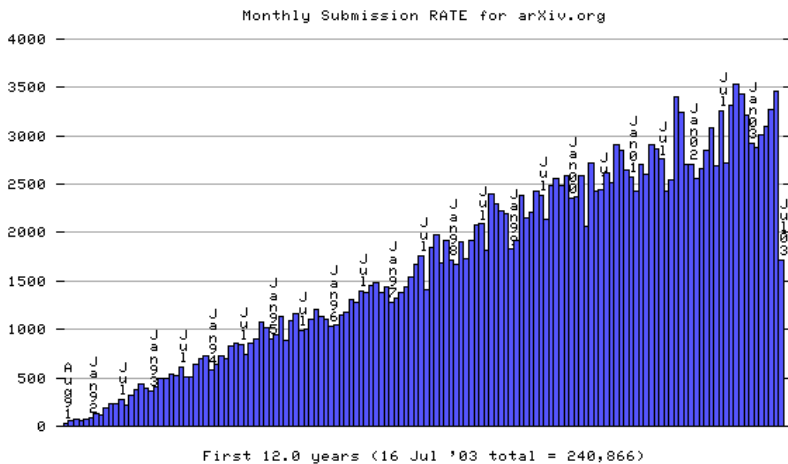
Abychom nemluvili pouze na abstraktní úrovni, v tomto odstavci použijeme řeč čísel, abychom kvantitativně vyjádřili jednak nárůst přístupů k serverům arXivu, a jednak abychom ukázali poměrně strmou závislost nárůstu objemu zaslaných článků na arXivu.

Na prvním z grafů, obr. 3, je vyneseno počet přístupů, počet připojených počítačů a také počty nových uživatelů na servery arXivu během posledních 314 týdnů. Jedná se tedy o histogram, kdy jednotkami na vodorovné ose jsou týdny, jednotkami na svislé ose jsou absolutní počty. Barevné odstíny odpovídají množství přístupů, připojených počítačů a nových hostitelů za týden (poslední dva údaje je nutno dělit 10). Skok blízko pravému konci grafu odpovídá přelomu roku a vánočnímu období.



Obrázek 3: Histogram počtu přístupů na arXiv.

Na druhém grafu, obr. 4, je opět znázorněn histogram, přičemž tentokrát se jedná o množství nových článků, které byly přijaty od srpna 1991 do července 2003.



Obrázek 4: Histogram počtu nových příspěvků, zaslaných na arXiv od srpna 1991 do července 2003.

Z obou grafů je patrný výrazný nárůst návštěvnosti stejně jako množství aktivních příspěvků do arXivu. To svědčí jednak o prudkém rozvoji vědních disciplín (například dramatický vývoj v teoretické fyzice spojený se strunovou teorií), nicméně také o stále rostoucí popularitě internetu a zejména elektronických informačních zdrojů.

6 Malé zamyšlení závěrem

Seznámili jsme se s plně automatizovaným elektronickým informačním zdrojem - databází nerecenzovaných fyzikálních a matematických článků arXiv. Uvedli jsme řadu jejích vlastností, které ji staví na čelní místo mezi databázemi odborných vědeckých publikací. Viděli jsme, že arXiv je velmi pružně a interaktivně stvořený pro účely maximálně efektivní práce se soudobými poznatky moderní teoretické fyziky a matematiky. Je zajímavé sledovat strmou křivku vývoje arXivu a jeho prosazení mezi ostatními databázemi. Původně zamýšlený jako „interní“ internetová databáze, určená pro asi 200 odborníků na poli částicové fyziky, arXiv rychle získal popularitu i mezi ostatními fyzikálními a později i matematickými disciplínami. Zamyslíme-li se nad důvodem tohoto úspěchu arXivu, zjistíme, že byl podmíněn několika příčinami. Za prvé je zde intenzivní snaha o překlenutí nedostatků klasických tištěných forem vědeckých publikací (půlroční prodleva mezi zasláním článku do konkrétního časopisu a jeho publikací není ojedinělá). Tento problém není ani zdaleka pouze akademický, vzhledem ke stále zrychlujícímu se tempu získávání vědeckých poznatků a jejich rozvoji se dnes články staré jeden nebo dva roky již stávají klasikou. Znovu zde poukazujeme na příklad teorie strun a superstrun v teoretické fyzice. Ve světle této skutečnosti je proto onen půlrok potřebný k publikaci v papírové formě plýtváním drahocenného času.

Druhá příčina, těsně související s první, je možnost radikálního prohloubení komunikace prostřednictvím internetu mezi vědci bádajícími v konkrétní oblasti. Nejenom, že elektronická pošta umožňuje snadnou korespondenci, ale navíc volně přístupný archiv výzkumných článků přirozeným způsobem nabízí možnost velmi otevřené diskuse dosažených výsledků.

Třetí příčinu lze také spatřovat v podstatně snadnějším vyhledávání požadovaných článků a seznámení se s jejich problematikou díky rozšiřujícímu se přístupu na internet. Tuto výhodu oceňují zejména mladí a začínající vědci, kteří ještě nemají mnoho osobních vědeckých kontaktů v oblasti. Je proto oprávněné očekávat další prohloubení zájmu o databáze podobného charakteru.

Už nyní je zjevné, že se databáze bude rozšiřovat, o čemž svědčí i další, nově vznikající zrcadla arXivu a finanční podpora, která se této myšlence dostává. Existuje perspektivní šance, že databáze s podobnou filozofií budou vznikat i v budoucnu a v dalších exaktních vědních oborech. Lze říci, že arXiv je prvním krokem k jakési vícedimenzionální vědecké kolaboraci.