Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra informatiky a výpočetní techniky

Diplomová práce

Tvorba datových zdrojů pro bibliometrická měření

Plzeň 2014 Bc. Tomáš Hanke

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 1. června 2014.

Tomáš Hanke

Abstract

Creation of data sources for bibliometric measurements

This thesis deals with creating of data sources for bibliometric measuring. For this purpose it analyses especially SciVerse Scopus database and possibilities of its use for bibliometric data downloading. It describes the Scopus APIs and modifications of program for more efficient data downloading.

As next there are introduced various measurements realized on the created data sources and some scales as the results of the measurements. The goal of this thesis is to create as large data source for bibliometric measuring as possible.

Abstrakt

Tvorba datových zdrojů pro bibliometrická měření

Tato diplomová práce se zabývá vytvářením datových zdrojů pro bibliometrická měření. Za tímto účelem zkoumá především službu SciVerse Scopus a možnosti jejího použití pro získávání bibliometrických dat. Popisuje její jednotlivá aplikační rozhraní (API) a úpravy odpovídajícího modulu existujícího meta-vyhledávače pro efektivnější stahování dat.

Dále představuje různá měření provedená nad vytvořenými datovými zdroji a jako výsledky těchto měření předkládá několik žebříčků. Cílem práce je vytvořit co největší datový zdroj pro bibliometrická měření.

Poděkování

Děkuji Ing. Daliboru Fialovi, Ph.D. za vedení mé diplomové práce, za jeho podnětné rady, připomínky, pomoc a čas, který mi věnoval.

Obsah

1 Úvod

Věda, výzkum a poznatky z těchto disciplín plynoucí jsou v současnosti základním hnacím motorem veškerého lidského poznání. Věda se rozvíjí, objevují se nové otázky, které je třeba prozkoumat a ověřit. Tím pádem vychází velké množství vědecké literatury a odborných článků. Z informací o těchto dokumentech je možné vyčíst zajímavá fakta. Například počet publikovaných článků lze vnímat jako ukazatel výkonnosti (produktivity) jednotlivých vědeckých pracovníků nebo celých týmů. Počet citací určitého díla jinými autory zase ukazuje jeho kvalitu. Je jistě zřejmé, že čím více je určitý článek citován, tím bude kvalitnější, převratnější či jiným způsobem přínosný. Sledováním vývoje citací přes více generací lze také do jisté míry odhadnout směr, kterým se může věda dále ubírat.

Takovýmto měřením a analýzou vědeckých dokumentů se zabývají vědní obory bibliometrie a scientometrie. Aby však bylo vůbec možné bibliometrická měření provádět, je nejprve potřeba mít k dispozici zdroj bibliometrických dat.

Cílem této diplomové práce je vytvořit co nejrozsáhlejší lokální úložiště dat umožňující bibliometrická měření. Data budou uložena v databázi, takže bude možné k nim jednoduše přistupovat příslušnými SQL dotazy. Práce navazuje na předchozí bakalářské práce ([Aug12], [Han12], [Kru12] a [Bou13]), zabývající se problematikou získávání dat z bibliografických databází. Všechny aplikace jsou detailně popsány ve zmíněných pracích, proto zde bude zmíněn pouze jejich aktuální stav a případné opravy. V této práci byl použit meta-vyhledávač, popsaný ve zdroji [Bou13].

V rámci práce bude nad získanými daty provedeno několik měření a výpočtů a dojde ke zhodnocení výsledků, což ukáže, zda je zvolený přístup vhodný k provádění bibliometrických měření.

2 Vysvětlení pojmů

Tato kapitola byla sepsána na základě studia zdrojů [Kri97], [Kat98], [Vaš80], [Vaš93], [Vin10], [Moe05], [Tho14] a [His14].

2.1 Bibliografie

Bibliografie se zabývá akademickým studiem knihy, nicméně toto označení bylo používáno již na počátku našeho letopočtu a v průběhu staletí se jeho význam mírně měnil. Označovalo tak nejprve psaní, přepisování či opisování knih (knihopísařské práce), později jejich skladování dle jasně daných pravidel a pořizování seznamů literatury. Dnes se termínem bibliografie označuje bibliografická činnost a nauka, která se bibliografickou činností a jejími projevy zabývá. Cílem bibliografie je dát čtenáři co nejúplnější představu o knize. Nezahrnuje už jen knihy samotné, ale i časopisy, audionahrávky, filmy, obrazy a internetové stránky.

Produktem bibliografie jsou například:

- seznam knih, ze kterých autor čerpal při psaní vlastního díla, tzn. zdroje (uvádí se obvykle na konci knihy),
- katalogy knihoven (papírové i elektronické),
- samostatné publikace, věnující se seznamu knih.

Rozlišují se dva základní druhy bibliografie:

- **Enumerativní bibliografie**

Enumerativní neboli systematická bibliografie má společný nějaký faktor – například jazyk, téma nebo období. Může se tedy jednat třeba o klasický přehled zdrojů na konci vlastního díla. Každý takový přehled by měl obsahovat jméno autora, název díla, místo vydání, jméno nakladatelství, rok vydání, ISBN, počet stran dokumentu, pořadí vydání a v případě citace i číslo stránky.

- **Analytická bibliografie**

Analytická neboli kritická bibliografie zkoumá vzhled knihy (vazbu, formát, velikost), historické souvislosti nebo se zaměřuje na textovou kritiku.

2.2 Bibliometrie

Bibliometrie se zabývá měřením a kvantitativní analýzou dokumentů. Dříve se používal termín statistická bibliografie. Zakladateli jsou F. J. Cole a N. B. Eales, kteří v roce 1917 vydali dílo „*Statistická analýza literatury*“. V tomto díle provedli statistickou analýzu literatury z oboru anatomie, která vycházela v letech 1850 až 1860. Touto prací chtěli ukázat, jak se v čase měnil zájem o anatomickou literaturu, a dále tuto literaturu rozdělili podle zemí, ve kterých byla publikována.

Postupně se použití statistické analýzy začalo rozšiřovat. V roce 1923 provedl E. Wyndham Hulme statistickou analýzu historie vědy. Využil při tom záznamy z časopisů v 17 sekcích Mezinárodního katalogu vědecké literatury.

Podobně zásadní studii vytvořili v roce 1927 P. L. K. Gross a E. M. Gross. Základem práce byly celkové počty a analýzy citací k článkům v chemických časopisech. Byla to první práce, která stála na citacích, a metoda citací se poté stala velmi rozšířenou.

Na výše uvedené průkopníky postupně navazovali další autoři. Postupem času se již bibliometrie dala charakterizovat jako kvantifikace bibliografických informací pro různé typy analýzy.

Bibliometrie využívá matematicko-statistické metody, jakými jsou např. statistický odhad, analýza statistických jevů, ověřování statistických hypotéz a další. Vytváří také bibliometrické zákony, které zkoumají zákonitosti růstu, rozptylu a stárnutí publikací.

Nejdůležitějším objektem bibliometrických výzkumů jsou ale bez pochyby citace, na základě kterých následně vznikají citační analýzy.

Informační základnou bibliometrických údajů jsou:

- citační registry,
- rejstříky (např. mezinárodní rejstřík Who is Publishing in Science),
- různé seznamy a katalogy.

Citační registr (index) je soupisem publikovaných materiálů citovaných ve sledovaných pramenech v určitém roce. Soupis je seřazen abecedně podle citovaných autorů a práce u jednotlivých autorů jsou uvedeny chronologicky podle roku vydání.

Pomocí citačních registrů je možné zjistit citovanost publikovaných dokumentů a také jejich informační hodnotu. Mimo jiné v něm lze najít odpovědi na otázky typu:

- Byla tato práce někde citována, a pokud ano, kým?
- Byly využity poznatky z této teorie v praxi, a pokud ano, s jakým výsledkem?
- Rozvíjel tento návrh někdo další, a pokud ano, zdokonalil ho?
- Je myšlenka této teorie opravdu původní?
- Bylo toto téma použito v nějaké nové oblasti?
- V kolika pracích je tento autor prvním autorem, eventuálně spoluautorem?
- Jaké další práce tento autor ještě napsal?

Za nejvýznamnější citační registr je považován *Science Citation Index* (SCI), který v současné době zahrnuje časopisecké články ze zhruba 6 900 časopisů, které se zabývají 150 vědními disciplínami. Byl vytvořen v *Institut of Science Information* (ISI) v roce 1964 a jeho zakladatelem je Eugen Garfield.

Jednou z nejrozsáhlejších služeb svého druhu je Web of Science. Je určena především pro univerzity a vědecká pracoviště a poskytuje přístup do těchto sedmi databází: *Science Citation Index* (SCI), *Social Sciences Citation Index* (SSCI), *Arts & Humanities Citation Index* (A&HCI), *Index Chemicus*, *Current Chemical Reactions*, *Conference Proceedings Citation Index: Science* a *Conference Proceedings Citation Index: Social Science and Humanities*. Zahrnuje nejvýznamnější odborné časopisy a konference z více než 200 vědních disciplín.

2.3 Scientometrie

Scientometrie se v mnoha oblastech s bibliometrií překrývá a někdy jsou oba pojmy zaměňovány. Zásadní rozdíl mezi nimi je ten, že bibliometrie zkoumá spíše parametry literatury, dokumentů a ostatních komunikačních médií, zatímco scientometrie si všímá hlavně vědecké produktivity a její prospěšnosti. Scientometrie je v podstatě nadstavbou bibliometrie, někdy je pro ni také používán termín věda o vědě. Proto je za zakladatele scientometrie považován již výše zmíněný Eugen Garfield, který zároveň používá metody bibliometrie.

Mnohé studie o scientometrii se týkají i bibliometrie (jsou i bibliometrickými studiemi) a to z toho důvodu, že nejdůležitějším objektem v obou těchto disciplínách jsou publikace a v obou disciplínách dochází ke kvantitativnímu zkoumání těchto objektů.

Zde je třeba zdůraznit fakt, že obě vědní disciplíny vypovídají především o kvantitě, ale nutně nemusí vypovídat o kvalitě. Věda je totiž složitá tvořivá záležitost a nedá se mechanicky posuzovat. Scientometrické indexy by proto měly být posuzovány jako užitečná pomůcka v rukou vědce, ale nemělo by být zapomínáno na možnost poměrně lehkého zneužití.

Problematice scientometrie se věnuje mezinárodní časopis *Scientometrics*, který vychází od roku 1978. Časopis pravidelně publikuje studie, reporty, krátké zprávy, recenze a podobný materiál z oblasti scientometrie.

3 Bibliografické databáze

3.1 Google Scholar

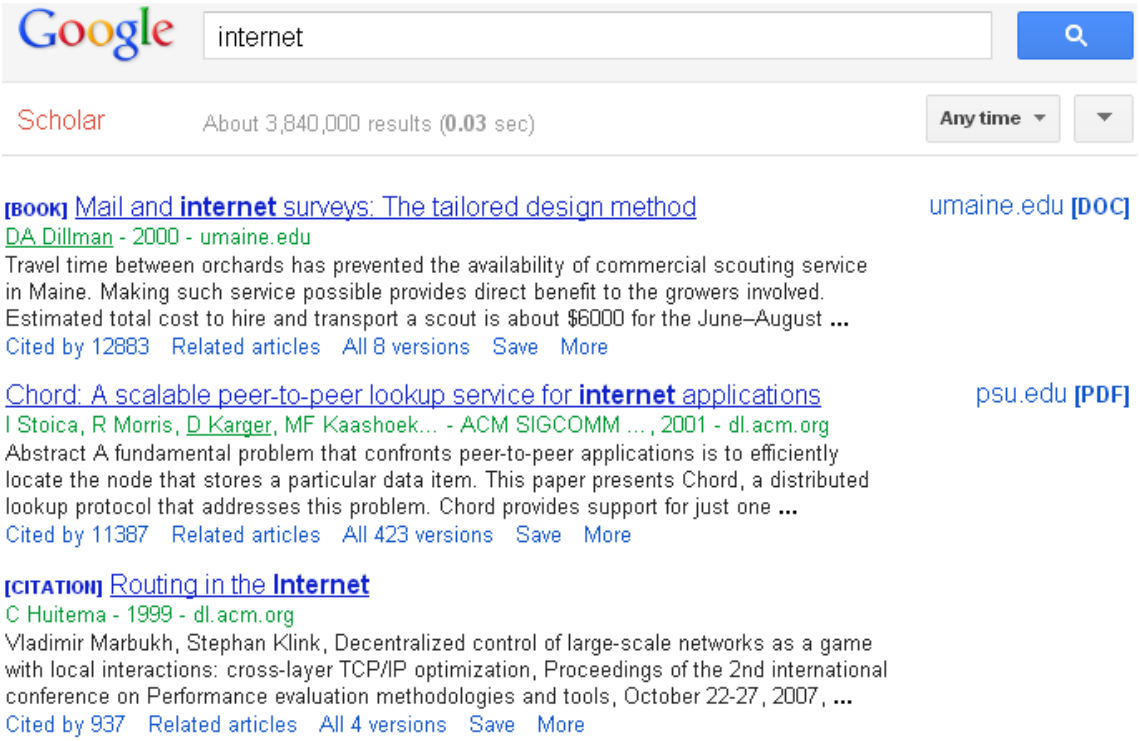
Scholar je služba firmy Google určená k vyhledávání odborných článků a informací o nich. Na rozdíl od většiny podobných služeb se neomezuje pouze na čistě vědecké práce, ale obsahuje například i bakalářské či diplomové práce. Zaměřuje se však především na odborné knihy, recenzované články, abstrakty a podobné publikace od akademických nakladatelství, z vědeckých konferencí a dalších odborných organizací. Kromě základních informací lze u vybraných publikací nalézt i informace o knihovně, v níž jsou k dispozici k půjčení, nebo dokonce stáhnout plný text práce. Mezi silné stránky patří rovněž možnost sledování citací jednotlivých textů [Han12].

Použití této služby je zdarma pro každého a vyhledávání v ní je v podstatě stejné jako klasické vyhledávání na Googlu. Scholar poskytuje uživatelské rozhraní v českém jazyce. Po přihlášení se k účtu Google umožňuje spravovat vlastní knihovnu článků a citací.

Vyhledávat lze podle klíčových slov, přičemž je možné určit, zda se ve vyhledaných článcích musí vyskytovat všechna klíčová slova, nejméně jedno či celá fráze. Některá slova je možné z výsledků i vyloučit. Standardně se prohledává celý text článku, ale je možné omezit vyhledávání pouze na titulek článku, což vede k přesnějším výsledkům hledání. Články lze hledat i podle jména autora nebo názvu časopisu, v němž byly publikovány. Dále je možné omezit hledání pouze na dokumenty vydané v určitých letech.

Po vyplnění kritérií pro vyhledávání je uživatel přesměrován na stránku s výsledky hledání. Ukázka prvních tří výsledků na dotaz „internet“ je na obrázku 1. Zde jsou o každém článku uvedeny základní informace jako název, hlavní autoři, rok publikování, vydavatel a počet dokumentů, které daný článek citují. Po kliknutí na název článku dojde k přesměrování na jeho plný text. Potřebuje-li uživatel zobrazit citující dokumenty, stačí kliknout na text „*Cited by*“ u příslušného článku.

Získávání dat ze služby Google Scholar pomocí meta-vyhledávače bylo bezproblémové. Program pracoval správně a dával požadovaná data. Jediným problémem bylo zřejmé zpřísnění pravidel Googlu pro přístup k jeho serverům. Limity popsané v pracích [Han12] a [Bou13], tedy hodinová přestávka ve stahování po každých 150 přístupech, již nestačí. Servery Googlu odmítly další dotazy již po 100 přístupech.



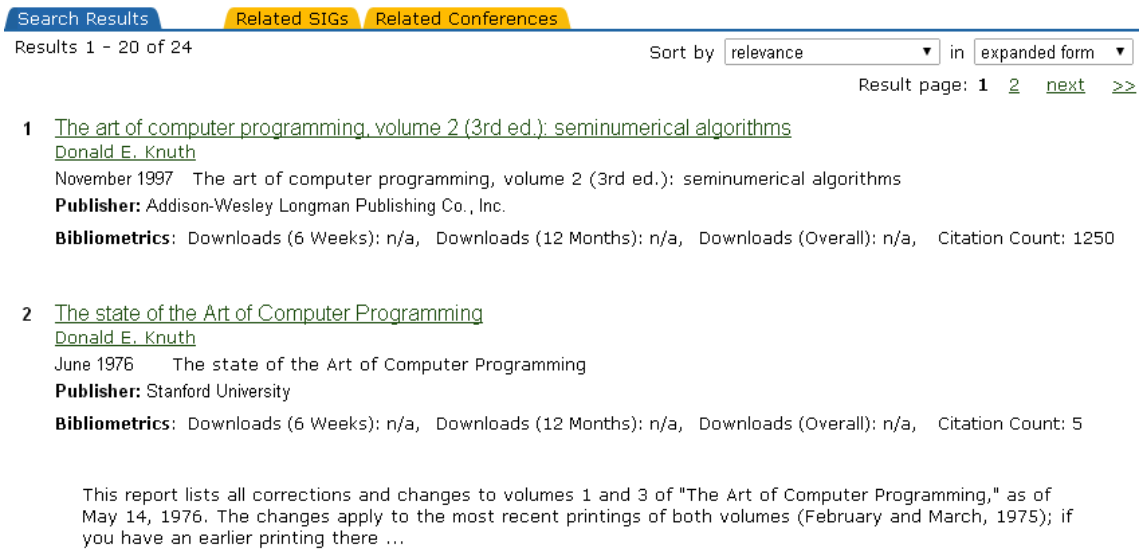
Obrázek 1: Ukázka výsledků hledání ve službě Google Scholar

3.2 ACM Digital Library

ACM Digital Library (DL) je rozsáhlá databáze sdružující odborné články a publikace z oboru výpočetní techniky a informačních technologií. Obsahuje více než 390 000 úplných textů odborných článků, přes 5 800 video souborů, osm odborných časopisů a 37 technických časopisů, vydávaných přímo skupinou ACM [ACM14].

Tato služba nabízí pokročilé vyhledávání, které uživateli umožňuje vyhledávat podle mnoha kritérií. Kromě standardních klíčových slov lze vyhledávat i podle afiliací, konferencí, identifikačních kódů jako ISBN nebo ISSN. Hledání lze rovněž omezit pouze na publikace vydané v určitém časovém rozmezí nebo publikované v konkrétním časopisu (či pouze jeho typu).

Stránka s výsledky hledání obsahuje přehled nalezených publikací se základními informacemi, mezi něž patří název článku s případným názvem časopisu či konference, seznam autorů, počet citací, datum vydání a, je-li k dispozici, krátký popis díla. Ukázka prvních dvou výsledků hledání na dotaz „The Art of Computer Programming“, nacházejícího se v titulku článku, je na obrázku 2.



Obrázek 2: Ukázka výsledků hledání ve službě ACM DL

Získávání dat z této služby nebylo možné, protože ihned po spuštění meta-vyhledávače došlo k chybové hlášce a příslušný parser byl ukončen, aniž by stáhnul jakákoli data.

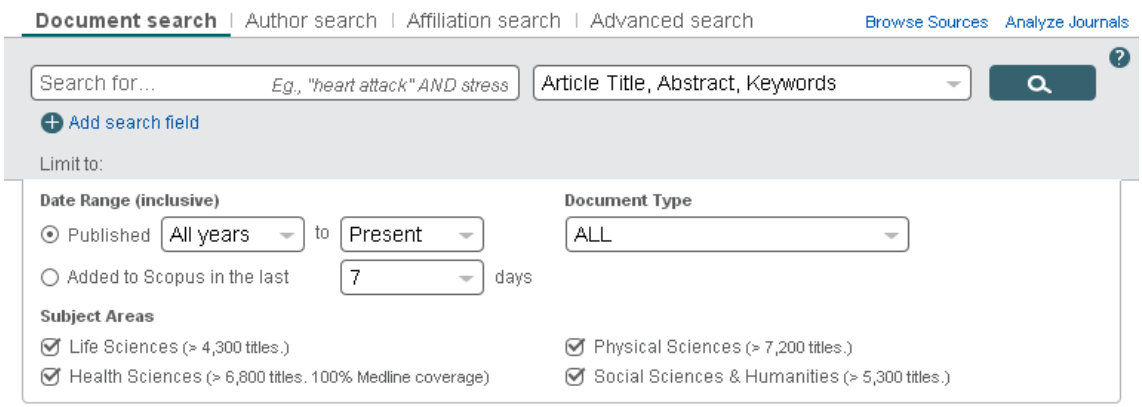
3.3 SciVerse Scopus

Scopus firmy Elsevier je placená služba pro vyhledávání odborné literatury. Jde o nejrozsáhlejší databázi svého druhu. Obsahuje přibližně 50 milionů záznamů, 21 000 titulů a sdružuje 5 000 vydavatelů [Els13].

Scopus podporuje čtyři druhy vyhledávání podle toho, co chce uživatel hledat. Jsou to *Document search*, *Author search*, *Affilliation search* a *Advanced search*.

3.3.1 Document search

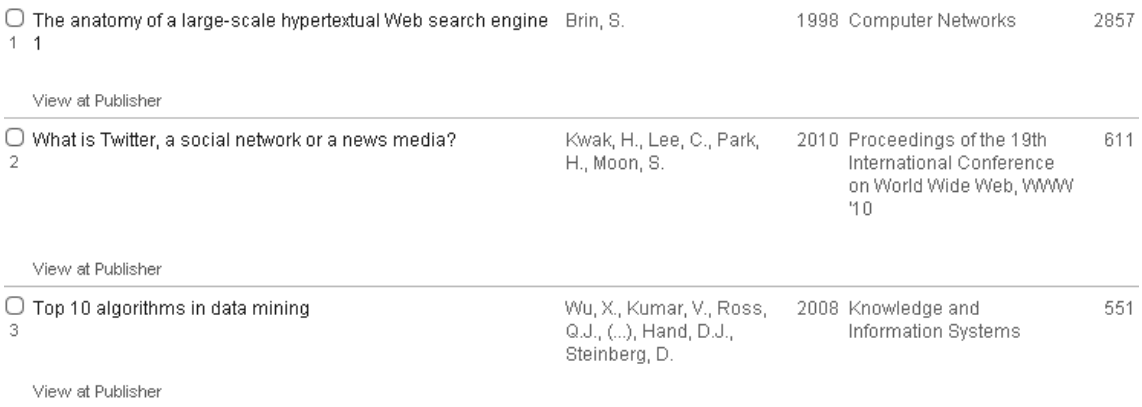
Document search umožňuje podrobné vyhledávání článků podle názvu, autora, afiliace, konference apod. Vyhledávat lze podle několika klíčových slov, přičemž jsou podporovány i logické operátory *AND* a *OR*. Samozřejmostí je možnost nastavit, kde se mají daná klíčová slova vyskytovat. Možnosti volby jsou široké, lze vyhledávat například v titulku dokumentu, klíčových slovech publikace, autorech, abstraktu či podle ISSN. Výsledky lze omezit rokem publikování, typem publikace (například článek, kniha, sborník atd.), ale i datem přidání do databáze Scopusu či oborem, kterému se daná publikace věnuje. Vyhledávací formulář je zobrazen na obrázku 3.



Obrázek 3: Scopus Document search

Po vyplnění kritérií pro vyhledávání je uživatel přesměrován na stránku s výsledky hledání. Ukázka prvních tří výsledků na dotaz „PageRank“ je na obrázku 4. Zde jsou o každé publikaci uvedeny základní informace jako název, hlavní autoři, rok publikování, název časopisu či konference a počet dokumentů, které danou publikaci citují. Po kliknutí na název článku se zobrazí jeho detailní popis, odkaz se jménem autora vede na autorův profil s podrobnějšími informacemi. Potřebuje-li uživatel zobrazit citující dokumenty, stačí kliknout na číslo s jejich počtem (na příslušném řádku úplně vpravo).

Na této stránce je možné výsledky hledání dále filtrovat podle mnoha kritérií, například podle roku vydání, autora, oboru, typu dokumentu, klíčových slov, země nebo jazyku. K tomu slouží rozsáhlé menu v levé části stránky.



☐ 1	The anatomy of a large-scale hypertextual Web search engine 1 View at Publisher	Brin, S.	1998	Computer Networks	2857
☐ 2	What is Twitter, a social network or a news media? View at Publisher	Kwak, H., Lee, C., Park, H., Moon, S.	2010	Proceedings of the 19th International Conference on World Wide Web, WWW '10	611
☐ 3	Top 10 algorithms in data mining View at Publisher	Wu, X., Kumar, V., Ross, Q.J., (...), Hand, D.J., Steinberg, D.	2008	Knowledge and Information Systems	551

Obrázek 4: Ukázka výsledků hledání ve službě Scopus

3.3.2 Author search

Tento mód slouží k vyhledávání autorů podle jména, příjmení, iniciál nebo afiliace. Stejně jako u dokumentů lze vyhledávání omezit oborem, kterému se daný autor věnuje.

Na stránce s výsledky hledání je u každého autora zobrazeno jeho jméno v několika variantách (odkaz vede na autorův podrobný profil), počet jeho publikací (ty lze

okamžitě zobrazit), obor, kterému se věnuje, afiliace, město a stát, kde publikuje. Výsledky hledání lze filtrovat podobně jako u dokumentů.

3.3.3 Affiliation search

Zde lze vyhledávat afiliace podle jejich názvu. Žádná další kritéria zadat nelze, vyhledávání je tedy velmi přímočaré.

Na stránce s výsledky hledání je u každé afiliace uveden její název v několika variantách (odkaz vede na její podrobný profil), počet publikací, které byly pod touto afiliací publikovány (lze je okamžitě zobrazit) a město a stát, do kterého afiliace patří. Výsledky hledání lze filtrovat podobně jako u autorů.

3.3.4 Advanced search

Pokročilé vyhledávání umožňuje uživateli sestavit vlastní vyhledávácí řetězec s použitím mnoha logických operátorů a interních kódů Scopusu. Stránka s výsledky hledání je totožná se stránkou s výsledky *Document search* a nabízí i stejné možnosti filtrování.

3.3.5 Omezení

Nevýhodou této databáze je její omezování počtu zobrazených položek při vyhledávání na 2 000. Pokusí-li se uživatel zobrazit více položek, je zobrazena hláška informující ho o celkovém počtu nalezených položek a faktu, že může zobrazit jen 2 000 prvních.

3.3.6 Program

Parser pro službu Scopus byl funkční jen částečně. Stahoval základní informace jako název publikace, jména autorů, rok vydání, nakladatelství, číslo svazku, rozsah stran a doplňující informaci. Ovšem téměř nikdy nestahoval afiliace (výjimkou byla pouze situace, kdy byl u článku uveden jen jeden autor s jedinou afiliací) a vůbec nikdy nestahoval citující záznamy ani jejich počet.

3.4 Zvolené řešení

Po analýze problému a dostupných prostředků bylo rozhodnuto, že se tato práce bude zabývat pouze bibliografickou databází SciVerse Scopus, protože je ze všech zmíněných nejkomplexnější a pro naše potřeby nejvhodnější. Obsahuje nejvíce ověřených záznamů a poskytuje pro práci s daty aplikační rozhraní.

4 Možnosti získávání dat

4.1 Parsování HTML

Jedním z možných způsobů získávání dat je parsování HTML kódu. Na tomto principu pracují všechny aplikace z předešlých bakalářských prací (viz [Aug12], [Han12], [Kru12] a [Bou13]). Tento přístup má výhodu v tom, že ho lze použít téměř na jakoukoli službu. Na druhou stranu má několik velmi nepříjemných omezení.

V našem případě aplikace k serveru přistupuje jako webový prohlížeč, stáhne zdrojový kód požadované stránky a ten poté zpracovává (parsuje). Tímto způsobem z kódu „vytáhne“ potřebná data a vše ostatní zahodí. Tím však vzrůstá objem zbytečně přenesených dat. Větším problémem je ale fakt, že takovýto parser je závislý na HTML kódu stránky. Jakmile dojde k jeho úpravě, ve většině případů přestává příslušný parser fungovat a je potřeba ho upravit tak, aby správně rozeznával nový HTML kód.

Toto omezení odstraňuje použití API (Application Programming Interface), které ale nemusí být vždy dostupné. Ze tří výše zmíněných bibliografických databází poskytuje aplikační rozhraní pouze služba Scopus.

4.2 API

Scopus pro přístup k datům poskytuje dvě aplikační rozhraní, *Javascript API* a *RESTful API*. Obě dvě jsou určena k použití výhradně ve webových aplikacích pro získání základních informací o vybraných titulech, případně ke zjištění počtu citujících prací. Obě jsou také dostupná pouze po registraci uživatele na stránkách firmy Elsevier [Dev13]. Po registraci je nutné svoji webovou stránku, z níž bude k API přistupováno, vložit do systému, čímž dojde k vygenerování jedinečného klíče. Ten musí být součástí každého dotazu, jinak na něj API nevrátí žádný výsledek.

4.2.1 Scopus Javascript API

Toto API je možné začít používat ihned po vygenerování klíče. Jde o v celku jednoduché javascriptové API, rozdělené do několika částí podle poskytované funkčnosti. Jde o *Document Search*, *Author Search*, *Affiliation Search* a *CitedBy Count Search* [API13]. Tato práce se bude zabývat pouze částí *Document Search* a *CitedBy Count Search*. Ostatní části staví na stejných principech, jsou však určeny k vyhledávání jiného obsahu. Nevýhodou všech je to, že modifikují přímo kód stránky, ze které byl do API odeslán dotaz. Stránka tak musí vždy obsahovat prvek DIV s konkrétním ID (pro *Document Search* je to „sciverse“, pro *Cited-By Count* „citedBy“). Tento DIV je poté měněn tím, že do něj API vloží výsledky dotazu. Z toho vyplývá, že API nelze použít v jiné než webové aplikaci.

Document Search API

Document Search API umožňuje vývojáři vyhledávat dokumenty podle zadaného klíčového slova, např. „pagerank“. Vrací seznam vyhovujících prací a základní informace o nich, jako například název práce, název časopisu nebo konference, kde byl článek publikován, typ dokumentu, ISSN, číslo části, rozsah stran, rok publikování, jméno autora, EID, Scopus ID, DOI a odkaz na detail článku přímo na stránkách Scopusu [Sco13].

Pro použití tohoto API je zapotřebí vytvořit dvě věci, webové rozhraní a skript. Webová stránka slouží pouze k zadání informací, podle nichž se má vyhledávat, a poté k zobrazení výsledků hledání. Skript obsahuje javascriptovou funkci, která informace ze stránky převezme, „zabalí“ do dotazu, přidá API klíč a vše odešle na server.

Příklad záhlaví (musí obsahovat odkaz na externí styly a API) a těla stránky pro *Document Search API*:

```
<head>
  <!-- Odkaz na styly -->
  <link REL="stylesheet" TYPE="text/css"
href="http://searchapi.scopus.com/stylesheets/css_sciverse_list_hil
ight.css"/>

  <!-- Vložení externího skriptu (odkaz na API) -->
  <script type="text/javascript"
src="http://api.elsevier.com/javascript/scopussearch.jsp"></script>
</head>

<body>
  <h2>Search Form:</h2>
  <form name="sciverseForm" onsubmit="return false">
    <input type="text" name="searchString"/>
    <button onClick="runSearch()"
name="searchButton"/>SEARCH</button>
  </form>
  <h2>Returned sciverse Content:</h2>
  <div id="sciverse">
    <!-- Sem budou vloženy výsledky dotazu. -->
    None.
  </div>
</body>
```

Ukázka kódu 1: Scopus Document Search API - HTML

Tělo stránky obsahuje pouze pole pro zadání hledaného textu (prvek *<input>*), tlačítko pro odeslání dotazu (prvek *<button>* - spustí javascriptovou funkci *runSearch* ukázanou níže) a HTML prvek * *, do něhož budou vloženy výsledky hledání vrácené serverem.

Následuje ukázka skriptu, který nejdříve zablokuje tlačítko pro odeslání formuláře a poté vytvoří objekt, obsahující nastavení vyhledávání. V tomto konkrétním případě převezme zadaný text pro vyhledávání, nastaví počet vrácených záznamů na deset a určí, že výsledky budou seřazeny sestupně podle počtu citací. Na závěr přidá API klíč a nastaví callback funkci, která je zavolána po obdržení odpovědi ze serveru.

Příklad skriptu pro *Document Search API*:

```
runSearch = function() {
        document.sciverseForm.searchButton.disabled = true;
        var varSearchObj = new searchObj();
        varSearchObj.setSearch(document.sciverseForm.searchString.value);
        varSearchObj.setNumResults(10);
        varSearchObj.setSort("CitedByCount");
        varSearchObj.setSortDirection("Descending");

        sciverse.setApiKey("API klíč");
        sciverse.setCallback(callback);

        sciverse.search(varSearchObj);
};

callback = function() {
        document.sciverseForm.searchButton.disabled = false;
};
```

Ukázka kódu 2: Scopus Document Search API - skript

Cited-By Count API

Cited-By Count API vrací počet dokumentů citujících námi hledaný článek. Vyhledávat lze na základě názvu článku, EID, DOI, SCP, PII, ISSN, ISBN, čísla části apod. Výsledek je vrácen ve formě obrázku, který je opět vložen do příslušného DIVu.

I zde je potřeba vytvořit HTML stránku a příslušný skript, obdobně jako u *Document Search API*. Obě tyto části jsou téměř totožné s verzemi z předchozího API, akorát obsahují více prvků pro vstup textu, protože zde lze vyhledávat podle více kritérií.

Příklad záhlaví (musí obsahovat odkaz na externí styly a API) a těla stránky pro *Cited-By Count API*:

```
<head>
  <!-- Odkaz na styly -->
  <link REL="stylesheet" TYPE="text/css"
href="http://searchapi.scopus.com/stylesheets/css_sciverse_list_hil
ight.css"/>

  <!-- Vložení externího skriptu (odkaz na API) -->
  <script type="text/javascript"
src="http://api.elsevier.com/javascript/citedby_image.jsp">
</script>
</head>

<body>
  <h2>Search Form:</h2>
  <form name="sciverseForm" onsubmit="return false">
    eid:<input type="text" name="eid"/><br>
    doi:<input type="text" name="doi"/><br>
    scp:<input type="text" name="scp"/><br>
    pii:<input type="text" name="pii"/><br>
    issn:<input type="text" name="issn"/><br>
    isbn:<input type="text" name="isbn"/><br>
    vol:<input type="text" name="vol"/><br>
    issue:<input type="text" name="issue"/><br>
    title:<input type="text" name="title"/><br>
    firstpage:<input type="text" name="firstpg"/><br>
    artno:<input type="text" name="artno"/><br>
    <button onClick="runSearch()"
name="searchButton"/>SEARCH</button>
  </form>
  <h2>Returned Image</h2>
  <div id="citedBy">
    <!-- Sem bude vložen obrázek s počtem citujících dokumentů. -->
    None.
  </div>
</body>
```

Ukázka kódu 3: Scopus Cited-By Count API - HTML

Příklad skriptu pro *Cited-By Count API*:

```
runSearch = function() {
        document.sciverseForm.searchButton.disabled = true;
        var varSearchObj = new searchObj();
        varSearchObj.setEid(document.sciverseForm.eid.value);
        varSearchObj.setDoi(document.sciverseForm.doi.value);
        varSearchObj.setScp(document.sciverseForm.scp.value);
        varSearchObj.setPii(document.sciverseForm.pii.value);
        varSearchObj.setIssn(document.sciverseForm.issn.value);
        varSearchObj.setIsbn(document.sciverseForm.isbn.value);
        varSearchObj.setVol(document.sciverseForm.vol.value);
        varSearchObj.setIssue(document.sciverseForm.issue.value);
        varSearchObj.setTitle(document.sciverseForm.title.value);
        varSearchObj.setFirstPg(document.sciverseForm.firstpg.value);
        varSearchObj.setArtNo(document.sciverseForm.artno.value);

        sciverse.setApiKey("API klíč");
        sciverse.setCallback(callback);
        sciverse.search(varSearchObj);
};

callback = function() {
        document.sciverseForm.searchButton.disabled = false;
};
```

Ukázka kódu 4: Scopus Cited-By Count API - skript

Některé dostupné metody pro Javascript API (kompletní seznam je k dispozici ve zdroji [Sco13]):

Metoda	Popis
setSearch(„hledaný text“)	Nastavuje hledaný text.
setNumResults(počet)	Nastavuje počet výsledků vyhledávání. Maximálně 2000 výsledků.
setOffset(počet)	Nastavuje, kolik výsledků bude přeskočeno. Např. je-li hodnota nastavena na 20, budou zobrazeny výsledky od 21. dále.
setSort(sort)	Nastavuje pole, podle něhož budou výsledky

	seřazeny. Povolené hodnoty jsou: Date (datum publikace), Relevancy (relevance), Authors (autoři), SourceTitle (název časopisu/konference), CitedByCount (počet citujících), LoadDate (datum přidání do Scopusu).
setSortDirection(sortDirection)	Nastavuje směr řazení. Povolené hodnoty jsou: Ascending (vzestupně) a Descending (sestupně).
setApiKey(„API klíč“)	Nastavuje API klíč, použitelný pro vyhledávání. Klíč musí být zadán vždy a musí být validní.
setCallback(Javascript funkce)	Nastavuje metodu, jež bude zavolána po dokončení vyhledávání.
search(objekt)	Spustí vyhledávání a zobrazí výsledky. Poté zavolá callback metodu.
areSearchResultsValid()	Bylo-li vyhledávání úspěšně dokončeno, vrací true, jinak vrací false.
getNumResults()	Vrací aktuálně nastavený počet výsledků. Použitelné pouze pokud areSearchResultsValid() vrací true.
getTotalHits()	Vrací celkový počet vyhovujících záznamů. Použitelné pouze pokud areSearchResultsValid() vrací true.
getSearchResults()	Vrací objekt s výsledky hledání. Použitelné pouze pokud areSearchResultsValid() vrací true.
getField(pozice, pole)	Vrací obsah pole na zadané pozici. Použitelné pouze pokud areSearchResultsValid() vrací true.

Tabulka 1: Metody Scopus Javascript API

4.2.2 Scopus RESTful API

Použití RESTful API je o něco komplikovanější než Javascript API. Po vygenerování klíče je nutné ho ještě aktivovat, do té doby je klíč nefunkční. To spočívá v zařazení projektu do jedné ze čtyř kategorií a jeho písemné představení společnosti Elsevier

[Pol13]. Její zástupci následně rozhodnou o tom, zda projekt splňuje veškerá pravidla a klíč může být aktivován.

Kategorie projektů podle používání dat ze Scopusu:

- zobrazování publikací ze Scopusu na webu,
- zobrazování počtu citací na webu,
- repositáře institucí, výzkumné systémy,
- federativní vyhledávání (vyhledávání ve více databázích za účelem srovnání výsledků).

REST (Representational State Transfer) je architektura pro webové rozhraní. Používá se k jednotnému přístupu ke zdrojům a jejich modifikaci. Každý zdroj musí mít vlastní URI[1] identifikátor, aby k němu bylo možné přistupovat. Zdroj mohou přestavovat data nebo nějaký stav aplikace, který lze těmito daty popsat.

REST poskytuje pro manipulaci se zdroji čtyři základní metody: *Create* pro vytvoření zdroje, *Retrieve* pro jeho získání, *Update* pro změnu a *Delete* pro smazání zdroje. Všechny tyto metody jsou implementovány pomocí metod protokolu HTTP[2]. K získání zdroje slouží HTTP metoda GET, pro jeho vytvoření metoda POST, pro aktualizaci zdroje metoda PUT a pro smazání zdroje metoda DELETE. Data mohou být klientovi doručena ve formátech XML, ATOM, JSON nebo RSS.

Samotné prohledávání Scopusu pomocí tohoto API je vcelku snadné a intuitivní. Vše spočívá pouze ve vytvoření správné URL[3] adresy, určující přístup ke zdroji, a nastavení příslušných filtrů. Pro vyhledávání dokumentů, autorů a afiliací je adresa vždy ve tvaru:

```
http://api.elsevier.com/content/search/index:{KDE_HLEDAT}?query={PODMÍNKY}
```

Řetězec „KDE_HLEDAT“ se mění v závislosti na typu hledaného obsahu. Určuje totiž, jaký typ vyhledávání bude použit (*Document search* apod., viz kapitola 3.3).

[1] Uniform Resource Identifier
[2] Hypertext Transfer Protocol
[3] Uniform Resource Locator

V případě hledání dokumentů bude nahrazen řetězcem „SCOPUS“, pro hledání autorů slouží řetězec „AUTHOR“ a při hledání afiliací bude použito slovo „AFFILIATION“.

„PODMÍNKY“ představují klíčová slova a další informace, sloužící k omezení výsledků hledání. Lze použít libovolné pole dostupné na Scopusu ve vyhledávacím módu *Advanced search* (viz kapitola 3.3), včetně logických operátorů.

Příklady adres pro různé druhy vyhledávání:

- **Hledání publikací autora s příjmením Novák:**

http://api.elsevier.com/content/search/index:SCOPUS?query=AUTHLASTNAME(novak)

- **Hledání publikací z oblasti chemie od autora s příjmením Novák:**

http://api.elsevier.com/content/search/index:SCOPUS?query=AUTHLASTNAME(novak)%20AND%20SUBJAREA(CHEM)

- **Hledání autorů s příjmením Novák a afiliací s identifikátorem „60032114“:**

http://api.elsevier.com/content/search/index:AUTHOR?query=af-id(60032114)%20AND%20authlast(novak)

- **Hledání afiliací z Plzně (vyhledává se v názvu afiliace a jejím popisu):**

http://api.elsevier.com/content/search/index:AFFILIATION?query=affil(plzen)

Díky REST API je možné na Scopusu nejen vyhledávat, ale dokonce si lze vyžádat plné znění článku, abstrakt, profil autora či afiliace nebo informace o časopisu. Následuje několik příkladů užitečných adres:

- **Plné znění článku s identifikátorem DOI „10.1016/0092-8674(93)90500-P“:**

http://api.elsevier.com/content/article/DOI:10.1016/0092-8674(93)90500-P?view=FULL

- **Abstrakt článku s identifikátorem „0027359827“:**

http://api.elsevier.com/content/abstract/SCOPUS_ID:0027359827

- **Profil autora s identifikátorem „44372231200“:**

http://api.elsevier.com/content/author/AUTHOR_ID:44372231200?view=STANDARD

- **Profil afiliace s identifikátorem „60016849“:**

http://api.elsevier.com/content/affiliation/AFFILIATION_ID:60016849?view=COMPLETE

- **Informace o časopisu s ISSN „07400551“:**

http://api.elsevier.com/content/serial/title?ISSN=07400551

Scopus RESTful API je velice robustní, kombinující technologie XML, ATOM a JSON [Con13]. Umožňuje vyhledávat téměř libovolný obsah a velkou část z něj dokonce přenášet k uživateli. Vše je navíc realizováno protokolem HTTP, takže pro práci s ním není zapotřebí žádného speciálního programového vybavení. Bohužel ho nelze vyzkoušet bez aktivovaného API klíče (viz začátek této kapitoly).

Omezení

Aby nedocházelo ke zneužívání přístupu k citlivým datům Scopusu, implementuje toto API (kromě již zmíněné nutnosti aktivovat API klíč) několik omezení v podobě limitovaného počtu hledaných položek či omezeného přístupu k různým částem API. Tato omezení jsou popsána v tabulce 2. Mít aktivovaný API klíč je nutné vždy, proto tento fakt už není v tabulce uveden.

API	Omezení
ScienceDirect Search	Počet výsledků hledání je omezen na maximálně 200 záznamů.
Scopus (documents) search	Bez Scopus licence maximálně 25 záznamů se základními metadaty ke každému záznamu. S licencí

	maximálně 200 záznamů se všemi dostupnými údaji.
Author search	Uživatel se Scopus licencí maximálně 200 záznamů. Bez licence nelze použít.
Affiliation search	Uživatel se Scopus licencí maximálně 200 záznamů. Bez licence nelze použít.
Full-text retrieval (ScienceDirect)	Uživatel s oprávněním k tomuto dokumentu na ScienceDirect dostane plný text dokumentu. Ostatní dostanou pouze metadata a abstrakt.
Abstract retrieval (Scopus)	Uživatel se Scopus licencí dostane celý záznam včetně referencí a rozšířených metadat. Ostatní dostanou pouze základní metadata.
Author retrieval	Uživatel se Scopus licencí dostane celý záznam o autorovi. Bez licence nelze použít.
Affiliation retrieval	Uživatel se Scopus licencí dostane celý záznam o afiliaci. Bez licence nelze použít.

Tabulka 2: Omezení Scopus RESTful API

4.3 Použité řešení

V průběhu práce na projektu bylo bohužel zjištěno, že ani jedno aplikační rozhraní není pro tento projekt použitelné. Žádné z nich totiž neumožňuje získat seznam citujících dokumentů, vždy vrátí pouze jejich počet a odkaz na příslušnou webovou stránku Scopusu.

Z tohoto důvodu byla s helpdeskem firmy Elsevier vedena emailové korespondence s žádostí o udělení výjimky a povolení stahování přes API běžně nedostupných dat. Avšak komunikace se zástupci firmy je v tomto ohledu dosti zdlouhavá. Poslední email, upřesňující informace o tomto projektu byl na helpdesk Scopusu odeslán 10. prosince 2013. Tři dny poté přišla odpověď, že žádost byla předána jinému oddělení a čeká se na její posouzení. Ještě 5. ledna 2014 nebyla věc vyřízena, tudíž byla možnost použití tohoto API definitivně vyloučena.

5 Příprava datových zdrojů

5.1 Úpravy meta-vyhledávače

V této podkapitole jsou rozebírány úpravy meta-vyhledávače, tedy programu pro stahování dat z bibliografických databází Google Scholar, ACM Digital Library a SciVerse Scopus. Tato aplikace byla vytvořena v rámci bakalářské práce Radka Boudy, kde je také blíže popsána [Bou13]. Jak bylo zmíněno v kapitole 3.4, tato práce se zabývá pouze databází Scopus a z tohoto důvodu byly prováděny úpravy pouze v těch částech meta-vyhledávače, které se touto službou zabývají.

Aplikace je napsána v programovacím jazyku Java ve verzi 1.6.0_24. K vývoji bylo použito vývojové prostředí *eclipse Helios* verze 3.6.2. Program byl vyvíjen pod operačním systémem Microsoft Windows 7 Professional 64bit a testován i na operačním systému Microsoft Windows XP Professional 32bit.

Jak bylo napsáno v kapitole 3.3, na začátku program nestahoval afiliace a citující položky. V případě citujících položek to bylo způsobeno změnou HTML kódu na stránkách Scopusu. Příslušná úprava v kódu aplikace je ve třídě *DocumentParser.java*, v metodě *getNodeList*. Ukázka je v následujícím rámečku. Na prvním řádku je původní kód, na druhém nový.

```
filter = new
CssSelectorNodeFilter("a[onclick~=\"javascript:submitRecord\"]");
filter = new CssSelectorNodeFilter(
"a[href~=\"http://www.scopus.com/record/display.url?eid\"]");
```

Ukázka kódu 5: Úprava meta-vyhledávače

S tímto problémem souvisela ještě změna v metodě *addCitations* třídy *Document.java*, která se stará o správné rozparsování stažených citací a vložení do příslušných kolekcí.

Pro zprovoznění stahování afiliací bylo zapotřebí udělat úpravy kódu na více místech. Program je sice z webu stahoval, ale špatně s nimi pracoval a nevypisoval je do výsledného XML souboru. Kvůli tomu byly prováděny úpravy v metodách *parseAuthors* třídy *Document.java* a *writeAuthors* třídy *XmlFileStax.java*.

Dále byla upravena metoda *parseSource* třídy *Document.java* tak, aby v případě, že od serveru nedostane požadovanou stránku, pokračoval dál. V předchozí verzi přestal program při takovémto problému pracovat.

V metodě *setCitationUrl* třídy *Document.java* je do URL adresy přidáván parametr pro řazení výsledků vyhledávání podle počtu citujících dokumentů.

Metoda *getNextPage* třídy *DocumentParser.java* byla upravena tak, aby v případě, kdy není možné vypočítat další stránku s výsledky hledání, program nepřestal pracovat. Dočasně nedostupné výsledky prostě ignoruje a pokračuje v činnosti.

V metodě *printHeader* třídy *DocumentParser.java* byla do úvodního textu přidána kontrolní informace o hledaném textu, počtu generací a počtu záznamů v první generaci.

Metoda *getResponse* třídy *DocumentParser.java* uspí aplikaci na pět minut v případě, že server vrátí chybový HTTP kód 5xx (chyba na straně serveru) nebo 4xx (chyba na straně klienta). Poté se činnost programu obnoví. V předchozí verzi aplikace v tomto případě skončila chybou.

Aplikace měla původně pro komunikaci s uživatelem grafické rozhraní. To muselo být odstraněno, protože byla provozována vzdáleně na linuxovém serveru. Kód grafického rozhraní nebyl úplně odstraněn, ale pouze zakomentován pro případ, že by byl v budoucnu opět potřebný.

5.2 Data

Bylo rozhodnuto, že úložiště dat bude sestaveno z nejobsáhlejší množiny výsledků hledání na tři různá klíčová slova. Celkově se podařilo nashromáždit 55 578 záznamů s následujícím rozložením:

- 14 453 na téma *internet*,
- 29 599 na téma *pagerank*,
- 11 526 na téma *telecommunication*.

Data jsou uložena v XML souboru, jehož struktura je popsána v kapitole 5.3. V této fázi nebyla data nijak čištěna či zjišťovány duplicity. Vzhledem k tomu, že byl datový základ složen ze tří různých dotazů, je možné, že budou některé položky duplicitní. Také se může stát, že budou tvořit tři nezávislé množiny.

Duplicita dat a jejich další čištění proběhne až po exportu dat do databáze, protože se tam s nimi bude lépe manipulovat.

5.3 Zdrojový XML soubor

V rámci oborového projektu bylo rozhodnuto, že XML soubor se zdrojovými daty bude sestaven z nejobsáhlejší množiny výsledků hledání na tři různá klíčová slova. Podařilo se nashromáždit 55 578 záznamů. Z toho 14 453 na téma *internet*, 29 599 na téma *pagerank* a 11 526 na téma *telecommunication*. Data jsou uložena v XML souboru, jehož struktura byla převzata z práce [Bou13].

Ukázka jednoho záznamu ve vygenerovaném XML souboru se čtyřmi autory a dvěma tisíci citujícími položkami (zobrazeno je pouze prvních pět citujících položek, další jsou pouze naznačeny tečkami):

```
<?xml version="1.0" encoding="UTF-8"?>
<publications>
 <publication id="2-s2.0-34547781750" generation="0" number="1">
  <title>MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software
version 4.0</title>
  <authors count="4">
   <author>
    <fullname>Tamura, K.</fullname>
    <firstname></firstname>
    <middlename></middlename>
    <lastname></lastname>
    <affiliations count="2">
     <affiliation>Center for Evolutionary Functional Genomics, Biodesign Institute,
Arizona State University</affiliation>
     <affiliation>Department of Biological Sciences, Tokyo Metropolitan University,
Tokyo, Japan</affiliation>
    </affiliations>
   </author>
   <author>
    <fullname>Dudley, J.</fullname>
    <firstname></firstname>
    <middlename></middlename>
    <lastname></lastname>
    <affiliations count="1">
     <affiliation>Center for Evolutionary Functional Genomics, Biodesign Institute,
Arizona State University</affiliation>
    </affiliations>
   </author>
   <author>
    <fullname>Nei, M.</fullname>
    <firstname></firstname>
    <middlename></middlename>
    <lastname></lastname>
    <affiliations count="1">
     <affiliation>Department of Biology, Institute of Molecular Evolutionary
Genetics, Pennsylvania State University</affiliation>
    </affiliations>
   </author>
```

Ukázka kódu 6: Ukázka XML souboru - začátek

```
  <author>
   <fullname>Kumar, S.</fullname>
   <firstname></firstname>
   <middlename></middlename>
   <lastname></lastname>
   <affiliations count="2">
    <affiliation>Center for Evolutionary Functional Genomics, Biodesign Institute,
Arizona State University</affiliation>
    <affiliation>School of Life Sciences, Arizona State University</affiliation>
   </affiliations>
  </author>
 </authors>
 <citedBy count="2000">
  <id>2-s2.0-74549125386</id>
  <id>2-s2.0-84863230041</id>
  <id>2-s2.0-52949099119</id>
  <id>2-s2.0-77957357555</id>
  <id>2-s2.0-58149090404</id>
  .....
 </citedBy>
 <sourceType>unknown</sourceType>
 <journal/>
 <conferenceName/>
 <year>2007</year>
 <publisher>Oxford University Press</publisher>
 <volume>24</volume>
 <number>8</number>
 <pages>1596-1599</pages>
 <country/>
 <city/>
 <month/>
 <isbn/>
 <issn/>
 <misc>Molecular Biology and Evolution</misc>
 </publication>
</publications>
```

Ukázka kódu 7: Ukázka XML souboru - pokračování

5.4 Databáze

Za účelem lepší práce s daty byla vytvořena relační databáze, do níž byla data z původního XML souboru exportována. Struktura databáze byla navržena tak, aby co nejlépe vystihovala zdrojová data a bylo možné do ní uložit i citační vazby.

Databáze sestává z následujících jedenácti tabulek:

- **Tabulka *publication***

Tabulka *publication* je nejobsáhlejší tabulkou v databázi. Obsahuje veškeré informace o publikaci, mezi nejdůležitější z nich patří identifikátor (sloupec *id*), číslo generace (*generation*), pořadové číslo (*pubnumber*), název publikace (*title*), název časopisu (*journal*) nebo konference (*conferencename*), rok vydání (*year*), nakladatelství (*publisher*), číslo svazku (*volume*), rozsah stran (*pages*), ISBN (*isbn*) a další. Velká část těchto informací však u většiny položek chybí, protože je při přípravě dat nebylo možné získat. Tabulka však tyto sloupce obsahuje a to z důvodu zachování struktury původního XML souboru.

- **Tabulka *author***

Tabulka *author* uchovává informace o autorech. Kromě jednoznačného identifikátoru (sloupec *id*) obsahuje sloupce pro křestní jméno (*firstname*), prostřední jméno (*middlename*), příjmení (*lastname*) a celé jméno (*fullname*). V aktuálních datech je používáno pouze celé jméno, ostatní položky jsou prázdné.

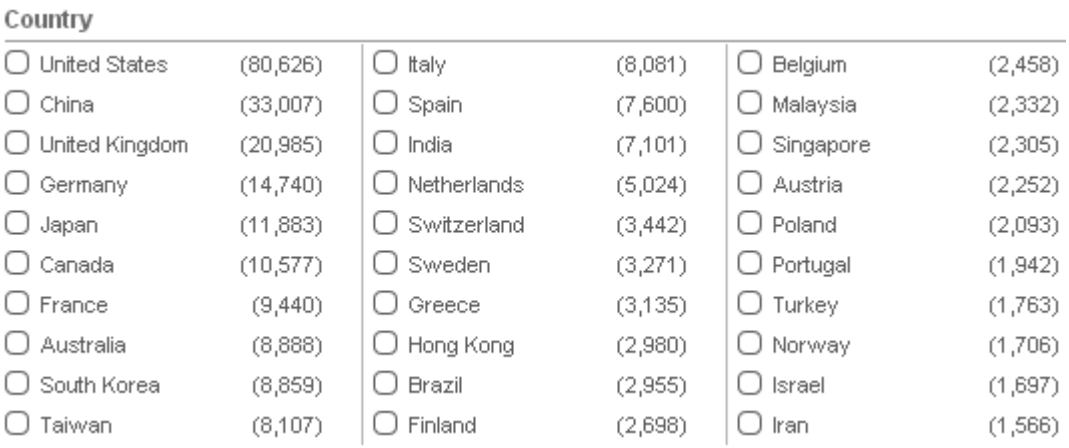
- **Tabulka *affiliation***

Tabulka *affiliation* obsahuje data o afiliacích (institucích). Obsahuje pouze položku pro název afiliace (sloupec *name*). Ten může zahrnovat název instituce (u univerzit ho může tvořit název katedry a/nebo fakulty a/nebo univerzity), její adresu (ulice a/nebo město) a stát, v němž se nachází. Z původního XML souboru se sem ukládá celý obsah atributu <*affiliation*>.

- **Tabulka *country***

Tabulku *country* tvoří opět pouze jedna položka, kterou je název země (sloupec *name*). Obsahuje názvy všech zemí získaných z databáze Scopusu. Zdrojová data u publikací neobsahují explicitně vyjádřenou příslušnost ke konkrétní zemi, ale afiliace autorů státní příslušnost obsahují. Při konverzi dat z XML souboru do databáze bylo proto postupováno následujícím způsobem. V každém afiliačním záznamu každého autora dané publikace byl hledán název nějakého státu z tabulky *country*. Pokud byl příslušný stát v afiliaci uveden, byl přiřazen k příslušné publikaci (vazba v tabulce *publicationcountry*). Nebyla-li nalezena žádná shoda mezi afiliací a kterýmkoli státem v tabulce, nebyla publikaci přiřazena žádná země.

Seznam zemí pro tuto tabulku byl získán z filtru zemí ve vyhledávání na Scopusu. Byly odeslány tři dotazy na stejná klíčová slova, která byla použita pro vytvoření datového podkladu (viz kapitola 5.2), a z filtru zemí na stránce s výsledky hledání (viz obrázek 5) byl vytvořen seznam, který byl následně importován do tabulky *country*.



Obrázek 5: Filtr zemí na stránce s výsledky hledání (www.scopus.com)

- **Tabulka *publicationauthor***

Jde o rozkladovou tabulku mezi tabulkami *publication* a *author*. Obsahuje pouze identifikátory položek z obou zmíněných tabulek.

- **Tabulka *publicationaffiliation***

Jde o rozkladovou tabulku mezi tabulkami *publication* a *affiliation*. Obsahuje pouze identifikátory položek z obou zmíněných tabulek.

- **Tabulka *publicationcountry***

Jde o rozkladovou tabulku mezi tabulkami *publication* a *country*. Obsahuje pouze identifikátory položek z obou zmíněných tabulek.

- **Tabulka *citationpublication***

Každý záznam v této tabulce znázorňuje citační vazbu mezi dvěma publikacemi. Obsahuje identifikátor citované (cited) a citující (citing) položky.

- **Tabulka *citationauthor***

Každý záznam v této tabulce znázorňuje citační vazbu mezi dvěma autory. Obsahuje identifikátor citované (cited) a citující (citing) položky.

- **Tabulka *citationaffiliation***

Každý záznam v této tabulce znázorňuje citační vazbu mezi dvěma afiliacemi. Obsahuje identifikátor citované (cited) a citující (citing) položky.

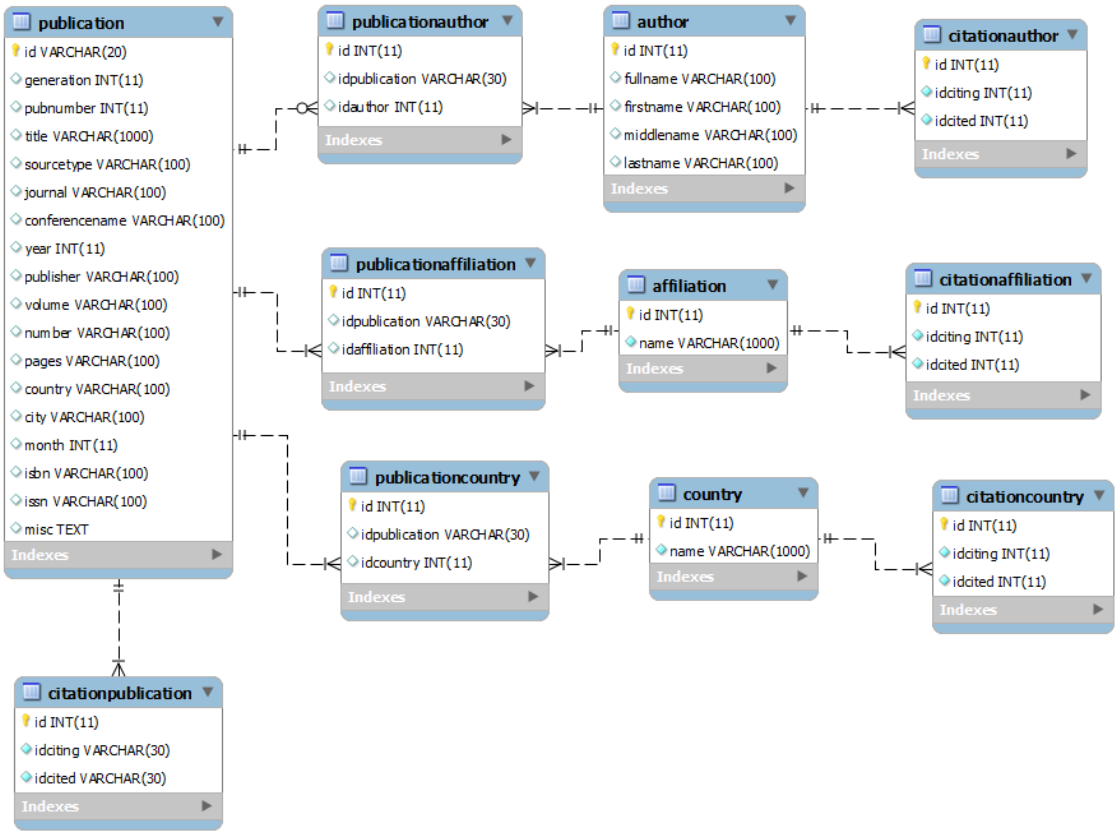
- **Tabulka *citationcountry***

Každý záznam v této tabulce znázorňuje citační vazbu mezi dvěma zeměmi. Obsahuje identifikátor citované (cited) a citující (citing) položky.

Poznámka: Je důležité upozornit, že tabulka *publication* obsahuje pouze ty záznamy, které byly ve zdrojovém XML souboru reprezentovány tagem <*publication*>. U těchto záznamů je k dispozici různé množství informací, pro jejichž uchování byla struktura XML souboru navržena (název, rok vydání, nakladatelství apod.). Oproti tomu tabulka *citationpublication* obsahuje pouze identifikátory publikací, a to všech publikací takových, mezi nimiž existuje citační vazba. Nezáleží na tom, zda mají tyto publikace záznam i v tabulce *publication*. Aby mohla být prováděná měření (viz

kapitola 6) uskutečněna na co nejobsáhlejší množině dat, bude výsledný graf pro tato měření sestaven právě na základě dat z tabulky *citationpublication* (neplatí pro žebříček počtu citací). To znamená, že do měření budou zahrnuty i ty publikace, k nimž v databázi neexistují ani základní informace, jako je například název. Tyto publikace budou ve výsledcích měření vystupovat pouze pod svým identifikátorem, na rozdíl od ostatních, které budou mít uveden kromě identifikátoru i název.

Model databáze je znázorněn na obrázku 6.



Obrázek 6: Model databáze

Vazby mezi tabulkami nejsou v databázi přímo zaneseny, důvod je popsán v poznámce výše v této kapitole (mnohdy neexistující reference mezi tabulkami *publication* a *citationpublication*). Každá tabulka však samozřejmě obsahuje primární klíč a podle potřeby několik indexů pro rychlejší vyhledávání. Export struktury databáze je k dispozici na přiloženém CD.

5.4.1 Import dat

Import dat z XML souboru do databáze probíhá ve dvou fázích. V první fázi jsou naplněny tabulky *publication*, *author*, *affiliation*, *country* a příslušné rozkladové tabulky (*publicationauthor*, *publicationaffiliation*, *publicationcountry*). V druhé fázi se plní všechny ostatní tabulky. V první fázi importu jsou totiž používána pouze data z XML souboru, kdežto ve druhé fázi se kombinují informace ze souboru s informacemi z již naplněných tabulek v databázi a plní se ostatní, dosud prázdné tabulky (*citationpublication*, *citationauthor*, *citationaffiliation* a *citationcountry*).

Průběh první fáze importu:

1) START
2) Přečti jeden záznam <*publication*> z XML souboru.
3) Vyparsuj informace ukládané do tabulky *publication*, tzn. název článku, generaci, identifikátor, rok vydání, stránkování apod. Ulož vše do tabulky *publication*.
4) Přečti jeden záznam <*author*>.
5) Ulož jméno autora do tabulky *author* a vlož příslušné vazby do tabulky *publicationauthor*.
6) Přečti jeden záznam <*affiliation*>.
7) Ulož název afiliace do tabulky *affiliation* a vlož příslušné vazby do tabulky *publicationaffiliation*.
8) Vyparsuj zemi z afiliace. Je-li informace o zemi nalezena, vlož příslušnou vazbu do tabulky *publicationcountry*.
9) Existuje-li další afiliace, jdi na bod 6.
10) Existuje-li další autor, jdi na bod 4.
11) Existuje-li další publikace, jdi na bod 2.
12) KONEC

Pseudoalgoritmus 1: První fáze importu dat do databáze

Průběh druhé fáze importu:

1) START
2) Přečti jeden záznam <*publication*> z XML souboru.
3) Získej z databáze identifikátory všech autorů, zemí a afiliací přidružených k citovanému záznamu (z bodu 2).
4) Získej identifikátor jedné citující publikace z elementu *citedBy* z XML souboru.
5) Jde-li o samocitaci, jdi na bod 4, jinak vlož citační vazbu do tabulky *citationpublication*.
6) Získej identifikátory všech zemí přidružených k citující publikaci (z bodu 4) a vlož citační vazby do tabulky *citationcountry* tak, aby existovala vazba mezi každou citující a každou citovanou (z bodu 3) zemí.
7) Získej identifikátory všech afiliací přidružených k citující publikaci (z bodu 4) a vlož citační vazby do tabulky *citationaffiliation* tak, aby existovala vazba mezi každou citující a každou citovanou (z bodu 3) afiliací.
8) Získej identifikátory všech autorů přidružených k citující publikaci (z bodu 4) a vlož citační vazby do tabulky *citationauthor* tak, aby existovala vazba mezi každým citujícím a každým citovaným (z bodu 3) autorem.
9) Existuje-li další citující publikace, jdi na bod 4.
10) Existuje-li další publikace, jdi na bod 2.
11) KONEC

Pseudoalgoritmus 2: Druhá fáze importu dat do databáze

5.5 Program pro import dat do databáze

Aplikace je napsána v programovacím jazyku Java ve verzi 1.6.0_24. K vývoji bylo použito vývojové prostředí *eclipse Juno*. Program byl vyvíjen pod operačním systémem Microsoft Windows 7 Professional 64bit a testován v tomtéž prostředí.

Komentované zdrojové kódy jsou k dispozici na přiloženém CD, zde tedy bude jen stručně vysvětlena architektura aplikace a popsány její základní vlastnosti.

Struktura balíků a tříd projektu:

- app – Main.java, Import.java
- data – ImportStAX.java
- db – Database.java, DbAffiliation.java, DbAuthor.java, DbCountry.java, DbPublication.java
- types
    - dbtypes – DbTypeAffiliation.java, DbTypeAuthor.java, DbTypePublication.java
    - xmltypes – XmlTypeAuthor.java, XmlTypePublication.java

Třída Main.java

Hlavní třída programu. Spouští import a řídí jeho průběh.

Třída Import.java

Třída pro import dat do databáze. Obsahuje tři metody: *prepareImport()*, *round1()* a *round2()*. Metoda *prepareImport* provádí přípravné práce před vlastním importem. Připraví spojení s databází a z tabulky *country* vytáhne seznam dostupných zemí.

Metody *round1* a *round2* provádí vlastní import dat. Každá metoda je určena pro jednu fázi importu (jednotlivé fáze jsou popsány v kapitole 5.4.1).

Třída ImportStAX.java

Třída datové vrstvy pro čtení dat z XML souboru technologií StAX. Obsahuje metodu getNextPublication(), která ze souboru vždy přečte jeden záznam typu <publication>. Vrací objekt typu *XmlTypePublication* (obsahující objekt *XmlTypeAuthor*).

Třída Database.java

Třída pro uchování údajů pro spojení s databází. Udržuje informace o databázi a poskytuje ostatním třídám přístup k ní.

Třída DbAffiliation.java

Poskytuje metody pro manipulaci s daty v databázové tabulce *affiliation* a v tabulkách na ní navázaných (např. *citationaffiliation*).

Třída DbAuthor.java

Poskytuje metody pro manipulaci s daty v databázové tabulce *author* a v tabulkách na ní navázaných (např. *citationauthor*).

Třída DbCountry.java

Poskytuje metody pro manipulaci s daty v databázové tabulce *country* a v tabulkách na ní navázaných (např. *citationcountry*).

Třída DbPublication.java

Poskytuje metody pro manipulaci s daty v databázové tabulce *publication* a v tabulkách na ní navázaných (např. *citationpublication*).

Třída XmlTypeAuthor.java

Třída reprezentující záznam typu <author> v XML souboru. Používá se pro přenos údajů o autorech načtených z XML mezi jednotlivými objekty programu.

Třída XmlTypePublication.java

Třída reprezentující záznam typu <publication> v XML souboru. Používá se pro přenos informací spojených s publikací a načtených z XML mezi jednotlivými objekty programu.

Třídy balíku dbtypes

Balík *dbtypes* obsahuje tři třídy ekvivalentní k třídám v balíku *xmltypes*, ale určené k uchování příslušné informace získané z databáze, nikoli z XML souboru. Třídy tohoto balíku byly vytvořeny, ale nakonec nebyly v programu použity. Přesto zde byly ponechány pro případ, že by byly v budoucnu potřeba.

5.6 JUNG

JUNG (Java Universal Network/Graph Framework) je knihovna pro Javu, poskytující prostředky pro modelování, analýzu a vizualizaci dat, která mohou být vyjádřena grafem nebo sítí. Knihovna je napsána v jazyce Java, což umožňuje aplikacím založeným na JUNG využívat jak rozsáhlé možnosti Javy, tak i již existující knihovny třetích stran.

JUNG podporuje různé entity a relace mezi nimi, takže dovoluje modelovat například orientované a neorientované grafy, multigrafy, grafy s paralelními hranami, hypergrafy a obecné grafy. Další metrikou pro výběr vhodné implementace může být hustota grafu. Vrcholy i hrany mohou být standardních datových typů nebo pro ně lze definovat vlastní typy.

Balík *algorithms* obsahuje rovněž velké množství implementovaných grafových algoritmů, vhodných pro analýzy sociálních sítí, optimalizaci, statistickou analýzu, výpočet vzdáleností v grafu, toky v síti a různé důležité grafové míry. Lze zde najít algoritmy jako například *Betweenness Centrality*, *PageRank*, *HITS*, *Shortest Path*, *Dijkstra* a další.

Tato knihovna byla v práci použita pro vytvoření grafové reprezentace a následné provedení výpočtů, popsaných v kapitole 6.

Následuje ukázka kódu s využitím knihovny JUNG pro vytvoření jednoduchého orientovaného grafu se třemi vrcholy (datového typu String) a čtyřmi hranami (datového typu Integer) mezi nimi.

```
// Vytvoření orientovaného grafu
DirectedGraph<String, Integer> g = new
DirectedSparseMultigraph<String, Integer>();

// Přidání vrcholů do grafu
g.addVertex("Vrchol1");
g.addVertex("Vrchol2");
g.addVertex("Vrchol3");

// Vložení hran mezi vrcholy
g.addEdge(1, "Vrchol1", "Vrchol2");
g.addEdge(2, "Vrchol1", "Vrchol3");
g.addEdge(3, "Vrchol2", "Vrchol3");
g.addEdge(4, "Vrchol3", "Vrchol1");
```

Ukázka kódu 8: Vytvoření grafu pomocí knihovny JUNG

Následující kód ukazuje, jak na grafu z předchozí ukázky spustit algoritmus PageRank, jehož implementace je také součástí knihovny JUNG.

```
// Vytvoření objektu typu PageRank nad původním grafem g
PageRank<String, Integer> pageRank = new
PageRank<String, Integer>(g, 0.15);

// Spuštění výpočtu na grafu
pageRank.evaluate();

// Zjištění ohodnocení vrcholu pro Vrchol2
double score = pageRank.getVertexScore("Vrchol2");
```

Ukázka kódu 9: Spuštění algoritmu PageRank nad grafem

6 Měření

Poznámka: V následujícím textu budou termíny „afiliace“ a „instituce“ brány jako synonyma. Oba budou v našem pojetí představovat katedru, univerzitu či jinou organizaci, pod kterou autor příslušný článek vydal. Také je třeba připomenout, že veškeré domněnky a závěry v tomto textu jsou vyvozeny z omezených datových podkladů a proto jsou platné jen a pouze pro tuto konkrétní množinu dat.

Po vyčištění dat od všech duplicit, samocitací mezi publikacemi apod. zbylo v databázi:

- 47 936 publikací (záznamy v tabulce *publication*),
- 97 559 autorů (záznamy v tabulce *author*),
- 272 afiliací/institucí (záznamy v tabulce *affiliation*),
- 238 zemí (záznamy v tabulce *country*).

V citačních tabulkách byly počty záznamů rozloženy následovně:

- 645 527 citací mezi publikacemi (tabulka *citationpublication*),
- 6 783 304 citací mezi autory (tabulka *citationauthor*),
- 173 322 citací mezi afiliacemi/institucemi (tabulka *citationaffiliation*),
- 842 976 citací mezi zeměmi (tabulka *citationcountry*).

Jak bylo uvedeno v poznámce v kapitole 5.4, graf pro měření nad publikacemi byl sestaven na základě dat z tabulky *citationpublication*. Z tohoto důvodu neobsahoval pouze 47 936 vrcholů (publikací), ale plných 291 337. Graf pro měření nad autory obsahoval 97 559 vrcholů a graf pro země 172 vrcholů (což znamená, že ve vstupních datech pochopitelně nebyly obsaženy všechny země světa).

Z výše uvedených čísel stojí za povšimnutí zejména dvě zvláštnosti. První z nich je velice malý počet záznamů v tabulce *affiliation*, což je důsledek malého rozsahu dat skrz různé organizace. Tento fakt vede na domněnku, že při vytváření datové základny sice došlo ke stažení informací o mnoha publikacích, ale jejich autoři pravděpodobně působili ve stejných institucích.

Problém nastíněný v předchozím odstavci dává tušit, že spolupráce mezi autory v rámci jedné instituce by mohla být vysoká, ale spolupráce mezi autory napříč různými organizacemi nebude pravděpodobně příliš obvyklá. Nicméně vzorek těchto dat je tak malý, že by nemělo smysl provádět na něm bibliometrická měření, protože by tato neměla prakticky žádnou vypovídací hodnotu. Z tohoto důvodu bylo rozhodnuto, že pro instituce nebudou žádná další měření provedena.

Druhou zvláštností je velký počet citačních vazeb mezi autory vzhledem k citačním vazbám mezi ostatními subjekty (desetinásobek oproti počtu citačních vazeb mezi publikacemi). Při studiu datových podkladů a programu pro import dat do databáze se došlo k závěru, že je toto číslo správné, protože v XML souboru nejsou výjimkou publikace, které mají desítky autorů, někdy dokonce hodně přes sto. Absolutním extrémem je publikace „*Guidelines for the use and interpretation of assays for monitoring autophagy*“, která má 1 269 autorů. Vzhledem k tomu, že u citace dvou publikací, které mají X a Y autorů, vzniká mezi autory X * Y citačních vazeb, lze uvedený celkový počet citací mezi autory považovat za reálný.

6.1 Problém s autory

V průběhu měření se vyskytl problém v datech o autorech. Program pro stahování dat z bibliografické databáze Scopus totiž o autorovi stahuje pouze jméno a afiliaci. Nestahuje tedy žádný identifikátor ani nezavádí žádný vlastní. Jméno je navíc složeno pouze z příjmení a prvního písmene křestního jména. Pod jedním jménem autora se tak může skrývat vícero skutečných osob. Tedy například záznam „Wang, J.“ může zahrnovat autory se jmény „Wang, Jing“, „Wang, Jianping“, „Wang, Jianbo“, „Wang, Jia“, „Wang, Jianmin“, „Wang, Juntao“, „Wang, Jin“ a dalšími podobnými.

6.1.1 Možná řešení

Jak bylo řečeno výše, program pro autory nezavádí žádné jednoznačné identifikátory, ty jsou autorům přiřazovány až v průběhu importu z XML souboru do databáze. Při tomto procesu by tedy bylo možné pokusit se autory se stejným jménem rozlišit.

Rozlišení autorů podle afiliace

Jedním z možných způsobů odlišení autorů se stejným jménem v XML souboru je analýza jejich afiliace. Tato varianta dělí osoby podle organizace, k níž jsou přiřazeni. Při použití této filtrace by byli rozlišeni například autoři „Wang, J.“ s afiliací „Tsinghua University, Graduate School at Shenzhen, Beijing, China“ a autor stejného jména s afiliací „Hebei Agricultural University, College of Food Science and Technology, Baoding, China“. Pokud by ale měli oba stejnou afiliaci, byli by bráni jako jedna osoba. Další osobu by tvořili autoři se stejným jménem, ale bez afiliace.

Překážkou tomuto postupu je nejednoznačnost informace, která se pod afiliací skrývá, a její formát (viz kapitola 5.4 – „Tabulka affiliation“). Není totiž jasně definováno, jaké informace má tento parametr obsahovat a v jakém pořadí. Navíc mají někteří autoři více afiliací, které nemusí být u různých publikací stejné.

Rozlišení autorů podle země

Dalším možným řešením je rozlišovat osoby pouze na základě země. Tato informace se v XML souboru sice vyskytuje taktéž v atributu *affiliation*, ale na rozdíl od jiných údajů je zde uvedena téměř vždy (a navíc pokaždé na konci záznamu, což usnadňuje identifikaci).

Ve výsledku je však tato metoda jen zjednodušením metody předchozí. Autoři se stejným jménem a ze stejné země, byť s působností v rozdílné organizaci, by opět byli považováni za tutéž osobu.

Poznámka: Metody určování autorství jsou na tento problém nepoužitelné, jelikož jsou k jejich aplikaci potřeba články, u nichž se má rozhodnout o autorovi. Tento problém je však odlišný od problému popisovaného v této kapitole, jelikož ten se žádnými články nezabývá.

6.1.2 Zvolené řešení

Vzhledem k rozsáhlosti této problematiky bylo rozhodnuto, že autoři se stejným jménem budou považováni za jednu osobu. To pravděpodobně zásadně ovlivní výsledné žebříčky v oblasti autorů, především ty, které se týkají produktivity. Prostředky vynaložené na řešení této situace by však byly vyšší než výsledný efekt. Autoři by s velkou pravděpodobností stejně nebyli rozlišeni správně, protože by toto rozlišování probíhalo pouze na základě domněnek, nikoli něčím podložených faktů.

6.2 Žebříčky

Celkem bylo sestaveno 21 žebříčků z různých měření. Jejich přehled zachycuje tabulka 3. Mezi žebříčky HITS a PageRank byl počítán také korelační koeficient (viz kapitola 6.3).

	publikace	**autoři**	**instituce**	**země**
počet publikací	✘	✓	✘	✓
relativní počet publikací	✘	✘	✘	✓

počet citací	✓	✓	✘	✓
relativní počet citací	✘	✓	✘	✓
in-deg	✓	✓	✘	✓
relativní in-deg	✘	✓	✘	✓
PageRank	✓	✓	✘	✓
HITS	✓	✓	✘	✓
spolupráce	✘	✓	✘	✓

Tabulka 3: Provedená měření

Dále budou jednotlivé žebříčky vyhodnoceny. Z důvodu velkého počtu položek bude vždy zmíněno jen několik prvních míst každého žebříčku. Kompletní žebříčky jsou k dispozici na přiloženém CD.

6.2.1 Spolupráce

Žebříček spolupráce byl sestaven pro autory a země. Prvních pět míst pro autory zachycuje tabulka 4.

Pozice	Autor 1	Autor 2	Počet společných publikací
1.	Wang, B.-H.	Zhou, T.	64
2.	Zhang, Z.	Zhou, S.	50

3.	Zhou, S.	Guan, J.	41
4.	Zhang, Z.	Guan, J.	40
5.	Kim, D.	Kahng, B.	39

Tabulka 4: Žebříček spolupráce autorů

Tabulka 5 ukazuje žebříček spolupráce pro země. Při porovnání s žebříčkem pro počet publikací jednotlivých zemí (tabulka 7) je sice zřejmé, že nejvíce spolupracují země s velkým počtem publikací, ale pravděpodobně budou nezanedbatelnou roli hrát i jiné aspekty. Tato domněnka vychází ze spolupráce Číny s Hong Kongem, který se v žebříčku počtu publikací umístil až na šestnáctém místě.

Pozice	Země 1	Země 2	Počet společných publikací
1.	China	United States	1 163
2.	United States	United Kingdom	929
3.	United States	Germany	665
4.	China	Hong Kong	550
5.	United States	Canada	532

Tabulka 5: Žebříček spolupráce zemí

Zajímavostí u tohoto žebříčku je to, že mezi prvními deseti místy se pouze ve dvou případech nevyskytují USA.

6.2.2 Počet publikací

V tabulce 6 je zachycen žebříček počtu publikací jednotlivých autorů. Pod uvedenými jmény se však může skrývat více osob (viz kapitola 6.1).

Pozice	Jméno autora	Počet publikací
1.	Wang, J.	449
2.	Wang, Y.	427

3.	Zhang, Y.	371
4.	Wang, X.	362
5.	Wang, L.	344

Tabulka 6: Žebříček počtu publikací pro autory

Tabulka 7 představuje žebříček počtu publikací jednotlivých zemí. Není tu nijak zohledněn počet autorů, kteří v dané zemi publikují.

Pozice	Název země	Počet publikací
1.	United States	16 182
2.	China	9 645
3.	United Kingdom	4 404
4.	Germany	3 770
5.	France	2 744

Tabulka 7: Žebříček počtu publikací pro země

Tabulka 8 zachycuje žebříček relativního počtu publikací na zemi. Na rozdíl od předchozího žebříčku tento zohledňuje počet autorů, kteří v dané zemi publikují.

Pozice	Název země	Relativní počet publikací
1.	Chad	1.00
2.	China	0.77
3.	Macao	0.57
4. – 5.	Barbados	0.50
4. – 5.	Liechtenstein	0.50

Tabulka 8: Žebříček relativního počtu publikací pro země

Čadu bylo přiřazeno devět publikací a stejný počet autorů, Číně 9 645 publikací a 12 476 autorů.

6.2.3 Počet citací

Žebříček, zachycený v tabulce 9, ukazuje jednu zajímavost. Maximální stahovaný počet citujících dokumentů z webu Scopusu byl dva tisíce položek. Jak je vidět, první tři dokumenty se ve zdrojovém XML souboru vyskytly vícekrát a pokaždé byla množina prvních dvou tisíc citujících dokumentů trochu jiná.

Pozice	Název publikace	Počet citací
1. – 2.	Complex networks: Structure and dynamics	2085
1. – 2.	Finding and evaluating community structure in networks	2085
3.	The structure and function of complex networks	2051
4. – 5.	Statistical mechanics of complex networks	2000
4. – 5.	The anatomy of a large-scale hypertextual Web search engine 1	2000

Tabulka 9: Žebříček počtu citací pro publikace

Tabulka 10 představuje žebříček počtu citací autorů. Počet citací je zde reprezentován citacemi na všech publikacích daného autora. Čím více publikací autor vydal, tím má vyšší šanci se v tomto žebříčku umístit na vedoucích pozicích. Počet publikací tu tedy není zohledněn.

Pozice	Jméno autora	Počet citací
1.	Newman, M.E.J.	86 704
2.	Barabási, A.-L.	44 208
3.	Sporns, O.	37 120
4.	Bullmore, E.	29 619
5.	Moreno, Y.	27 319

Tabulka 10: Žebříček počtu citací pro autory

V tabulce 11 je zachycen žebříček relativního počtu citací autorů. Na rozdíl od předchozího žebříčku tento zohledňuje počet publikací, které autor vydal. Umístění se na předních pozicích v tomto žebříčku je tedy pro autora prestižnější než u předchozího žebříčku, protože tento ukazuje pravděpodobnou vysokou úroveň všech jeho publikací (nebo alespoň většiny), zatímco u předchozího žebříčku stačilo mít vydánu jednu extrémně citovanou publikaci. Například pátý autor žebříčku, Nei, M., měl 14 069 citací na třech vydaných publikacích.

Pozice	Jméno autora	Relativní počet citací
1.	Dudley, J.	13 772
2.	Fax, J.A.	10 284
3.	Brin, S.	6 199
4.	Mongru, D.A.	5 755
5.	Nei, M.	4 690

Tabulka 11: Žebříček relativního počtu citací pro autory

Tabulka 12 ukazuje žebříček počtu citací zemí. Počet citací je zde reprezentován citacemi na všech publikacích přiřazených dané zemi a není tu zohledněn počet autorů z příslušné země. Opět zde tedy platí, že čím lidnatější země (více autorů), tím větší pravděpodobnost umístění se na předních pozicích žebříčku.

Pozice	Název země	Počet citací
1.	United States	259 532
2.	China	65 143
3.	United Kingdom	58 668
4.	Germany	51 340
5.	Spain	49 489

Tabulka 12: Žebříček počtu citací pro země

Žebříček v tabulce 13 představuje relativní počty citací jednotlivých zemí. Na rozdíl od předchozího žebříčku tento zohledňuje počet autorů, kteří publikují pod danou zemí.

Pozice	Název země	Relativní počet citací
1.	Chad	9.11
2.	Hungary	9.00
3.	United States	5.88
4.	Hong Kong	5.73
5.	Spain	5.60

Tabulka 13: Žebříček relativního počtu citací pro země

Čad měl 82 citací na publikacích od devíti autorů, Maďarsko 22 077 citací na publikacích od 2 454 autorů a Spojené státy dostaly 259 532 citací na publikacích od 44 144 autorů.

6.2.4 In-deg

In-degree neboli vstupní stupeň uzlu grafu udává, z kolika ostatních různých uzlů grafu vede hrana na tento uzel. Všechny násobné hrany mezi dvojicí uzlů se tedy počítají jako jedna hrana. V tom spočívá hlavní rozdíl oproti žebříčkům počtů citací, neboť ty počítají násobné hrany vícekrát.

Tabulka 14 ukazuje žebříček vstupních stupňů pro publikace. Tento žebříček je stejný jako žebříček počtu citací mezi publikacemi (tabulka 9). To je dáno tím, že jedna publikace může být citována jinou publikací maximálně jednou, ne vícekrát.

Pozice	Název publikace	Vstupní stupeň
1. – 2.	Complex networks: Structure and dynamics	2085
1. – 2.	Finding and evaluating community structure in networks	2085
3.	The structure and function of complex	2051

	networks	
4. – 5.	Statistical mechanics of complex networks	2000
4. – 5.	The anatomy of a large-scale hypertextual Web search engine 1	2000

Tabulka 14: Žebříček in-degree pro publikace

V tabulce 15 je zachycen žebříček vstupních stupňů pro autory. Tento žebříček nezohledňuje počet publikací jednotlivých autorů.

Pozice	Jméno autora	Vstupní stupeň
1.	Barabási, A.-L.	13 322
2.	Newman, M.E.J.	12 931
3.	Tamura, K.	11 078
4.	Nei, M.	11 071
5.	Kumar, S.	10 964

Tabulka 15: Žebříček in-degree pro autory

Tabulka 16 představuje žebříček relativních vstupních stupňů autorů. Na rozdíl od předchozího žebříčku tento zohledňuje počet publikací, které autor vydal. Například čtvrtý autor žebříčku, Nei, M., byl citován 11 071 autory na třech vydaných publikacích.

Pozice	Jméno autora	Relativní vstupní stupeň
1.	Dudley, J.	10 855
2.	Brin, S.	4 057
3.	Mongru, D.A.	3 766
4.	Nei, M.	3 690
5.	Fax, J.A.	2 011

Tabulka 16: Žebříček relativního in-degree pro autory

Tabulka 17 ukazuje žebříček vstupních stupňů zemí. Asi nepřekvapí, že se na prvních dvou místech opět umístily Spojené státy a Spojené království.

Pozice	Název země	Vstupní stupeň
1.	United States	162
2.	United Kingdom	148
3. – 5.	France	136
3. – 5.	Germany	136
3. – 5.	Japan	136

Tabulka 17: Žebříček in-degree pro země

Žebříček v tabulce 18 představuje relativní počty vstupních stupňů zemí. Na rozdíl od předchozího žebříčku tento zohledňuje počet autorů, kteří publikují pod danou zemí.

Pozice	Název země	Relativní vstupní stupeň
1.	Chad	1.88
2.	Burundi	1.69
3.	New Caledonia	1.0
4.	North Korea	0.75
5.	Mauritania	0.71

Tabulka 18: Žebříček relativního in-degree pro země

Například Čad byl citován 17 různými zeměmi na publikacích od devíti autorů a Severní Korea byla citována šesti zeměmi na publikacích od osmi autorů.

6.2.5 PageRank

V tabulce 19 je zachycen žebříček publikací podle ohodnocení algoritmem PageRank.

Pozice	Název publikace	Ohodnocení
1.	Statistical mechanics of complex networks	0.071
2.	MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0	0.055
3.	The structure and function of complex networks	0.023
4.	Hierarchical organization of modularity in metabolic networks	0.019
5.	The anatomy of a large-scale hypertextual Web search engine 1	0.012

Tabulka 19: Žebříček PageRank pro publikace

Tabulka 20 ukazuje žebříček autorů podle ohodnocení algoritmem PageRank.

Pozice	Jméno autora	Ohodnocení
1.	Dudley, J.	0.012128
2.	Kumar, S.	0.012121
3.	Tamura, K.	0.012043
4.	Barabási, A.-L.	0.010110
5.	Nei, M.	0.010066

Tabulka 20: Žebříček PageRank pro autory

Tabulka 21 ukazuje žebříček zemí podle ohodnocení algoritmem PageRank.

Pozice	Název země	Ohodnocení
1.	United States	0.1031
2.	United Kingdom	0.0325
3.	Germany	0.0247
4.	Spain	0.0214
5.	Italy	0.0211

Tabulka 21: Žebříček PageRank pro země

6.2.6 HITS

Žebříček v tabulce 22 představuje ohodnocení publikací algoritmem HITS.

Pozice	Název publikace	Ohodnocení
1.	Finding and evaluating community structure in networks	0.422
2.	The structure and function of complex networks	0.259
3.	Modularity and community structure in networks	0.247
4.	Uncovering the overlapping community structure of complex networks in nature and society	0.243
5.	Fast algorithm for detecting community structure in networks	0.227

Tabulka 22: Žebříček HITS pro publikace

Žebříček v tabulce 23 představuje ohodnocení autorů algoritmem HITS.

Pozice	Jméno autora	Ohodnocení
1.	Barabási, A.-L.	0.108
2.	Newman, M.E.J.	0.105
3.	Wang, J.	0.100
4.	Li, Y.	0.092
5.	Wang, X.	0.080

Tabulka 23: Žebříček HITS pro autory

Žebříček v tabulce 24 představuje ohodnocení zemí algoritmem HITS.

Pozice	Název země	Ohodnocení
1.	United States	0.864
2.	Spain	0.206
3.	Germany	0.205
4.	Italy	0.199
5.	United Kingdom	0.185

Tabulka 24: Žebříček HITS pro země

6.3 Korelační koeficient

Není-li u konkrétního odstavce napsáno jinak, byly informace v této kapitole čerpány ze zdrojů [Chr07], [Nyk11] a [Rei04].

Mezi žebříčky PageRank a HITS byl pro publikace, autory a země vypočítán také Spearmanův koeficient pořadové korelace. Jde o bezrozměrné číslo, vyjadřující statistickou závislost mezi dvěma veličinami. Může nabývat libovolné hodnoty z intervalu <-1, 1>. Čím více se korelační koeficient blíží k hodnotě +1, tím více jsou na sobě obě veličiny závislé (oba algoritmy, PageRank i HITS, by vytvořily úplně stejné žebříčky). Je-li koeficient roven nule, pak mezi veličinami není žádný vztah, a pokud se blíží hodnotě -1, potom jsou na sobě veličiny závislé opačně (v tomto případě by jeden žebříček měl prvky uspořádané v přesně opačném pořadí než druhý).

Spearmanův korelační koeficient se počítá podle následujícího vzorce:

$$\rho = 1 - \frac{6 * \quad _i(p_i - q_i)^2}{n * (n^2 - 1)}$$

kde ρ je hodnota Spearmanova koeficientu, n je počet porovnávaných hodnot jednotlivých měření a p_i a q_i jsou pořadová čísla jednotlivých hodnot obou žebříčků.

Výše uvedený vzorec však předpokládá, že na jednotlivých pozicích žebříčku je vždy právě jeden prvek, tudíž se v žádném žebříčku nesmí vyskytnout více prvků se stejným ohodnocením. To ale pro námi analyzovaná data neplatí, jelikož se na spodních pozicích jednotlivých žebříčků vždy vyskytuje několik prvků na stejné pozici. Z tohoto

důvodu je nutné použít upravený vzorec Spearmanova korelačního koeficientu ve tvaru:

$$\rho = \frac{n * (\ {}_i p_i q_i) - \quad {}_i p_i \ * (\ {}_i q_i)}{\overline{(n * \ {}_i p_i^2) - (\ {}_i p_i)^2} * \overline{(n * \ {}_i q_i^2) - (\ {}_i q_i)^2}}$$

Význam všech proměnných je stejný jako u předchozího vzorce.

Vypočítané hodnoty Spearmanova koeficientu pořadové korelace pro algoritmy PageRank a HITS jsou:

- 0,729 pro publikace,
- 0,837 pro autory a
- 0,950 pro země.

Podle tabulky kritických hodnot Spearmanova koeficientu jsou vypočítané hodnoty pro všechna tři měření významná na hladině 0,01 [Cri14].

Z uvedených hodnot je zřejmé, že oba algoritmy, PageRank i HITS, poskytují pro naši množinu dat velice podobné výsledky a to zejména v případě žebříčků pro hodnocení zemí, které jsou téměř totožné.

7 Závěr

Cílem této diplomové práce bylo vytvořit co nejrozsáhlejší datový zdroj umožňující provádět bibliometrická měření. Práce navazuje na čtyři předchozí bakalářské práce ([Aug12], [Han12], [Kru12] a [Bou13]) a vychází z jejich výsledků. Tito předchůdci zkoumají možnosti stahování dat ze tří bibliometrických databází, Google Scholar, ACM Digital Library a SciVerse Scopus. Byly vytvořeny tři programy, které byly následně integrovány do jednoho meta-vyhledávače. Těmito prostředky se podařilo stáhnout maximálně 16 329 záznamů a uložit je do XML souboru, jehož struktura byla definována v práci [Bou13].

Po prozkoumání všech předchozích prací bylo rozhodnuto, že se dále bude pracovat pouze s databází Scopus. Ze všech tří je nejkvalitnější, nejobsáhlejší a pro přístup poskytuje aplikační rozhraní (API). Pozdějším zkoumáním tohoto API však bylo zjištěno, že je pro účely této práce nepoužitelné, protože pro jeho použití musí být projektu uděleno povolení od pracovníků Scopusu, které se bohužel nepodařilo získat. Z tohoto důvodu musel být pro vytvoření datových zdrojů použit stejný způsob jako v přechozích bakalářských pracích (parsováním HTML stránek). Této problematice se věnuje kapitola 4, která popisuje aplikační rozhraní služby Scopus a vysvětluje, proč ho není možné použít.

Další část práce spočívala v úpravě meta-vyhledávače tak, aby ze Scopusu dokázal stáhnout co největší množství dat. Nakonec se podařilo vytvořit zdrojový XML soubor s 55 578 záznamy, složený z výsledků hledání na tři různá klíčová slova: 14 453 záznamů na téma *internet*, 29 599 na téma *pagerank* a 11 526 na téma *telecommunication*.

Pro lepší manipulaci s daty při následných měřeních, byla vytvořena relační databáze s jedenácti tabulkami tak, aby v co největší míře respektovala strukturu zdrojového XML souboru. Pro účel importu dat ze souboru do databáze byl vytvořen program v jazyce Java. Import probíhá kvůli potřebě spárovat citující a citované

publikace ve dvou fázích. Celý proces přípravy dat pro následná měření je detailně popsán v kapitole 5.

Po vytvoření datové základny bylo pro ověření její funkčnosti provedeno několik měření, z nichž bylo sestaveno celkem 21 různých žebříčků. Tato měření, sledující produktivitu (počet publikací), významnost (počet citací, in-degree a hodnocení dle algoritmů PageRank a HITS) a spolupráci, byla provedena pro autory, publikace a země. Sledování těchto parametrů u institucí nebylo možné z důvodu malého rozsahu jejich výskytu ve vstupních datech. Tato měření by neměla téměř žádnou vypovídací hodnotu, a proto nebyla provedena. Na závěr byl mezi žebříčky, získanými algoritmy PageRank a HITS, vypočítán Spearmanův koeficient pořadové korelace. Veškeré provedené experimenty jsou popsány v kapitole 6. Přílohy B až V ukazují prvních čtyřicet míst z každého sestaveného žebříčku.

7.1 Návrhy pro další práci

Nezbytným základem pro skutečně přínosná měření jsou rozsáhlé a především kvalitní datové zdroje. Primárním cílem pro další podobné práce by proto mělo být zkvalitnění procesu získávání datových podkladů. Především by bylo vhodné upravit program pro stahování dat ze Scopusu tak, aby o autorech a afiliacích nestahoval pouze jména či názvy, ale především jejich jedinečné identifikátory, čímž by bylo zajištěno rozlišení autorů se stejnými jmény. Další výhodou by byl fakt, že by se v případě vícenásobného výskytu autora ve výsledcích hledání nemusely jeho údaje stahovat opakovaně. Studiem služby Scopus bylo zjištěno, že stejně jako má každá publikace svůj jedinečný identifikátor *eid*, tak i každý autor má své jednoznačné *authorId*. To lze získat ze stránky s detaily publikace, která je už v současné verzi programem parsována.

Dalším možným zlepšením by mohlo být odstranění mezikroku v podobě ukládání dat do XML souboru. V případě, že by program pro stahování dat ze Scopusu dokázal komunikovat přímo s databází (a ukládat do ní data), mohl by z ní získávat informace o tom, jaká data již obsahuje, což by umožňovalo sestavovat podobné dotazy bez toho,

aby se stahovala stejná data vícekrát, protože by program vždy v podstatě navázal tam, kde při předchozím pokusu skončil.

Samozřejmě úplně nejlepší variantou by bylo použití Scopus RESTful API, které pro tyto účely poskytuje perfektní podmínky. Avšak jedinou možností, jak získat právo k jeho použití, je vést časově náročný dialog s helpdeskem Scopusu, představit celý projekt a snažit se prokázat jeho užitečnost.

Literatura

[Aug12] Augusta Rudolf, *Import dat ze služby Scopus do formátu XML*, Západočeská univerzita v Plzni, 2012, Bakalářská práce

[Han12] Hanke Tomáš, *Import dat ze služby Google Scholar do XML*, Západočeská univerzita v Plzni, 2012, Bakalářská práce

[Kru12] Krupička Jan, *Import dat ze služby ACM DL do formátu XML*, Západočeská univerzita v Plzni, 2012, Bakalářská práce

[Bou13] Bouda Radek, *Vytváření citačních sítí z bibliografických dat*, Západočeská univerzita v Plzni, 2013, Bakalářská práce

[Kri97] Krištofičová Eva, *Prostriedky hodnotenia knižničných a vedeckoinformačných procesov*, Centrum vedecko-technických informácií SR v Bratislave, 1997, ISBN 80-85165-62-7

[Kat98] Katuščák Dušan, Matthaeidesová Marta, Nováková Marta, *Informačná výchova – Terminologický a výkladový slovník*, Slovenské pedagogické nakladatelstvo, 1998, ISBN 80-08-02818-1

[Vaš80] Vašák Pavel, *Metody určování autorství*, ACADEMIA Praha, 1980

[Vaš93] Vašák Pavel, *Textologie – Teorie a ediční praxe*, KAROLINUM Praha, 1993, ISBN 80-7066-638-2

[Vin10] Vinkler Péter, *The Evaluation of Research By Scientometric Indicators*, Chandos Publishing, 2010, ISBN 978-1-84334-572-5

[Moe05] Moed Henk, *Citation Analysis in Research Evaluation*, Springer, 2005, ISBN 978-1-4020-3713-9

[Chr07] Chráska Miroslav, *Metody pedagogického výzkumu – Základy kvantitativního výzkumu*, Grada, 2007, ISBN 978-80-247-1369-4

[Nyk11] Nykl Michal, *Vyhodnocování informačních sítí*, Západočeská univerzita v Plzni, 2011, Diplomová práce

[Rei04] Reif Jiří, *Metody matematické statistiky*, Západočeská univerzita v Plzni, 2004, ISBN 80-7043-302-7

[ACM14] The Digital Library | The ACM Digital Library, nahlíženo 4. 1. 2014, http://librarians.acm.org/digital-library

[Els13] Scopus | Elsevier, nahlíženo 27. 12. 2013, http://www.elsevier.com/online-tools/scopus

[Dev13] Elsevier | Developers, nahlíženo 27. 12. 2013, http://www.developers.elsevier.com/cms/scopusintegration#

[API13] Scopus API: Home, nahlíženo 27. 12. 2013, http://searchapidocs.scopus.com/

[Sco13] Scopus Search JavaScript API, nahlíženo 27. 12. 2013, http://www.developers.elsevier.com/devcms/scopus-search-javascript-api

[Pol13] Content Policies, nahlíženo 27. 12. 2013, http://www.developers.elsevier.com/devcms/content-policies

[Con13] Content APIs, nahlíženo 27. 12. 2013, http://www.developers.elsevier.com/cms/content-apis

[Cri14] Critical values of Spearman's rho (two-tailed), nahlíženo 30. 5. 2014, http://www.sussex.ac.uk/Users/grahamh/RM1web/Rhotable.htm

[Tho14] Web of Science | Thomson Reuters, nahlíženo 5. 5. 2014, http://thomsonreuters.com/thomson-reuters-web-of-science

[His14] History of Citation Indexing, nahlíženo 5. 5. 2014, http://wokinfo.com/essays/history-of-citation-indexing

Seznamy

Tabulky

Obrázky

Ukázky kódu

Pseudoalgoritmy

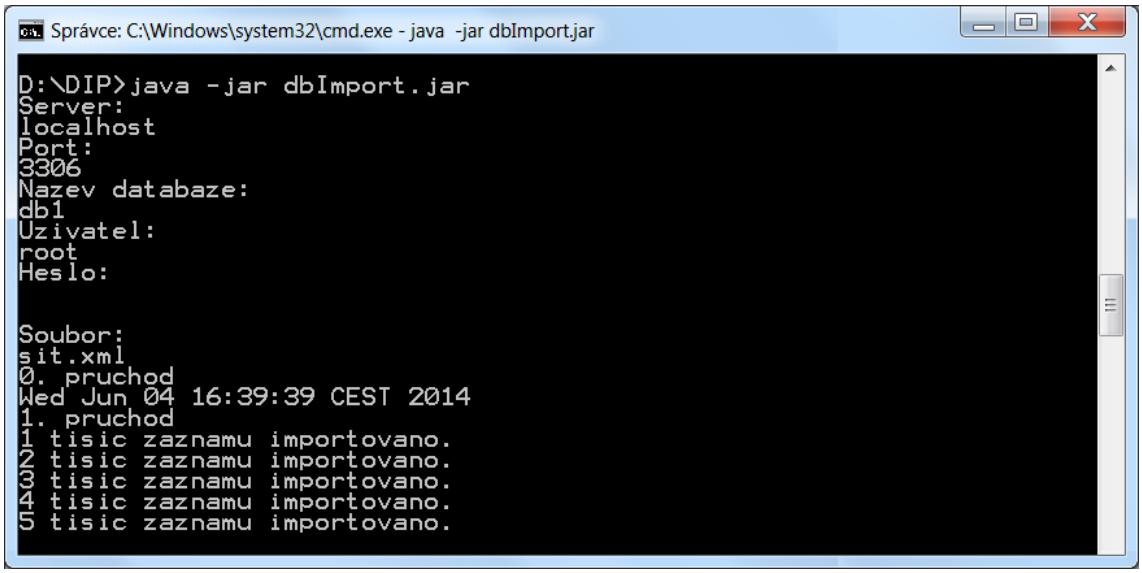
Přílohy

A Uživatelská dokumentace

Program pro import dat z XML souboru do databáze se nachází na přiloženém CD v adresáři *„bin“*. Lze ho spustit z příkazové řádky příkazem *„java –jar dbImport.jar“*. Pro správný běh aplikace je potřeba (kromě nainstalované Javy) mít ve stejné složce i podadresář *„lib“* včetně všech knihoven, které obsahuje.

Program má textové uživatelské rozhaní a ovládá se velice jednoduše. Po spuštění je uživatel vyzván k zadání informací pro připojení k databázi a cesty ke zdrojovému XML souboru. Poté je zahájen samotný import a uživatel je o jeho průběhu informován po každém tisícím importovaném záznamu.

Na obrázku 7 je náhled programu.



Obrázek 7: Program pro import dat z XML souboru do databáze

B Žebříček spolupráce autorů

Pozice	Autor 1	Autor 2	Počet společných publikací
1	Wang, B.-H.	Zhou, T.	64
2	Zhang, Z.	Zhou, S.	50
3	Zhou, S.	Guan, J.	41
4	Zhang, Z.	Guan, J.	40
5 – 6	Kim, D.	Kahng, B.	39
5 – 6	Wu, B.	Wang, B.	39
7	Kimura, M.	Saito, K.	38
8	Wang, J.	Li, M.	36
9 – 10	Havlin, S.	Stanley, H.E.	35
9 – 10	Duan, Z.	Chen, G.	35
11 – 12	Perc, M.	Szolnoki, A.	34
11 – 12	Gómez-Gardeñes, J.	Moreno, Y.	34
13	Saramäki, J.	Kaski, K.	33
14 – 17	Santos, F.C.	Pacheco, J.M.	32
14 – 17	Kimura, M.	Motoda, H.	32
14 – 17	Wang, B.-H.	Wang, W.-X.	32
14 – 17	Saito, K.	Motoda, H.	32
18	Fu, F.	Wang, L.	31
19	Wang, J.	Li, Y.	30
20	Chen, X.	Wang, L.	28
21 – 25	Di, Z.	Fan, Y.	27
21 – 25	Martin, D.P.	Varsani, A.	27
21 – 25	Szabó, G.	Szolnoki, A.	27
21 – 25	Wang, W.-X.	Lai, Y.-C.	27
21 – 25	Goh, K.-I.	Kahng, B.	27
26 – 28	Kajikawa, Y.	Sakata, I.	26
26 – 28	Li, J.	Wang, J.	26
26 – 28	Babiloni, F.	Mattia, D.	26
29 – 37	Marwan, N.	Kurths, J.	25
29 – 37	Hahn, B.H.	Shaw, G.M.	25
29 – 37	Wang, L.	Wang, J.	25
29 – 37	Wang, B.-H.	Yang, H.-X.	25
29 – 37	Dorogovtsev, S.N.	Mendes, J.F.F.	25
29 – 37	Wang, Y.	Liu, J.	25
29 – 37	Cao, X.-B.	Du, W.-B.	25
29 – 37	De Vico Fallani, F.	Babiloni, F.	25
29 – 37	Kertész, J.	Kaski, K.	25
38 – 40	Li, X.	Wang, X.	24
38 – 40	Cincotti, F.	Mattia, D.	24
38 – 40	Astolfi, L.	Babiloni, F.	24

C Žebříček spolupráce zemí

Pozice	Země 1	Země 2	Počet společných publikací
1	China	United States	1 163
2	United States	United Kingdom	929
3	United States	Germany	665
4	China	Hong Kong	550
5	United States	Canada	532
6	Italy	United States	492
7	France	United States	490
8	Germany	United Kingdom	482
9	United States	Australia	380
10	Spain	United States	373
11	United States	Japan	295
12	United Kingdom	France	285
13	Italy	France	284
14	Switzerland	United States	272
15	Italy	United Kingdom	267
16	United States	Netherlands	266
17	China	United Kingdom	263
18	United States	Israel	261
19	Germany	France	260
20	China	Australia	254
21	Italy	Spain	246
22	United States	Georgia	235
23	United Kingdom	Spain	233
24	United Kingdom	Australia	226
25	China	Germany	221
26	United States	Brazil	209
27	United States	South Korea	206
28	United Kingdom	Netherlands	205
29 – 30	China	Japan	194
29 – 30	Canada	United Kingdom	194
31	Canada	China	187
32	Mexico	United States	184
33	Netherlands	Germany	178
34	Spain	France	177
35	Germany	Italy	176
36	Spain	Germany	173
37	Sweden	United States	169
38	Switzerland	Germany	164
39	Switzerland	United Kingdom	157
40	India	United States	155

D Žebříček počtu publikací pro autory

Pozice	Jméno autora	Počet publikací
1	Wang, J.	449
2	Wang, Y.	427
3	Zhang, Y.	371
4	Wang, X.	362
5	Wang, L.	344
6	Li, Y.	334
7	Li, X.	320
8	Liu, Y.	319
9	Zhang, J.	286
10	Liu, J.	272
11	Zhang, Z.	247
12	Li, J.	244
13	Wang, H.	234
14	Wang, Z.	223
15	Chen, G.	222
16	Li, H.	206
17	Zhang, X.	203
18	Liu, X.	190
19	Chen, L.	187
20	Chen, Y.	185
21	Li, Z.	183
22 – 23	Zhang, L.	179
22 – 23	Zhang, H.	179
24	Chen, X.	178
25	Liu, Z.	173
26	Chen, J.	162
27 – 28	Li, L.	153
27 – 28	Chen, H.	153
29	Zhou, T.	150
30	Li, M.	148
31	Li, W.	147
32	Wu, J.	145
33	Yang, Y.	144
34	Wang, S.	143
35 – 36	Wang, B.-H.	140
35 – 36	Liu, H.	140
37 – 38	Wang, B.	136
37 – 38	Li, S.	136
39	Wang, W.	135
40	Chen, Z.	133

E Žebříček počtu publikací pro země

Pozice	Název země	Počet publikací
1	United States	16 182
2	China	9 645
3	United Kingdom	4 404
4	Germany	3 770
5	France	2 744
6	Italy	2 666
7	Japan	2 495
8	Spain	2 192
9	Canada	2 074
10	Australia	1 947
11	Netherlands	1 398
12	Switzerland	1 272
13	Brazil	1 173
14	South Korea	1 155
15	India	973
16	Hong Kong	932
17	Sweden	820
18	Belgium	723
19	Israel	647
20	Singapore	549
21	Denmark	539
22	Taiwan	531
23	Portugal	491
24	Finland	464
25	Mexico	464
26	Austria	460
27	Poland	451
28	Hungary	444
29	Georgia	435
30	Greece	372
31	Norway	343
32	Russian Federation	327
33	Ireland	323
34	New Zealand	314
35	Argentina	291
36	Iran	286
37	South Africa	286
38	Czech Republic	234
39	Slovenia	226
40	Thailand	215

F Žebříček relativního počtu publikací pro země

Pozice	Název země	Relativní počet publikací
1	Chad	1.000
2	China	0.773
3	Macao	0.571
4 – 5	Barbados	0.500
4 – 5	Liechtenstein	0.500
6	Qatar	0.440
7	Azerbaijan	0.428
8	Yemen	0.400
9	Cyprus	0.372
10	United States	0.366
11 – 13	Bahamas	0.333
11 – 13	Cape Verde	0.333
11 – 13	Saint Kitts and Nevis	0.333
14	Hong Kong	0.320
15	Macedonia	0.304
16	Japan	0.296
17	Armenia	0.285
18	Italy	0.276
19	Algeria	0.263
20	Sri Lanka	0.261
21	United Arab Emirates	0.257
22	United Kingdom	0.255
23	South Korea	0.253
24	Germany	0.251
25	Jamaica	0.250
26	North Korea	0.250
27	Spain	0.248
28	Romania	0.244
29	Canada	0.238
30	Venezuela	0.233
31	Australia	0.223
32	Syrian Arab Republic	0.218
33	Switzerland	0.217
34	Fiji	0.214
35	Oman	0.213
36	Cuba	0.211
37	Brazil	0.210
38 – 39	India	0.207
38 – 39	France	0.207
40	Tunisia	0.203

G Žebříček počtu citací pro publikace

Pozice	Název publikace	Počet citací
1 – 2	Complex networks: Structure and dynamics	2 085
1 – 2	Finding and evaluating community structure in networks	2 085
3	The structure and function of complex networks	2 051
4 – 7	Statistical mechanics of complex networks	2 000
4 – 7	The anatomy of a large-scale hypertextual Web search engine 1	2 000
4 – 7	Network biology: Understanding the cell's functional organization	2 000
4 – 7	MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0	2 000
8	Consensus and cooperation in networked multi-agent systems	1 906
9	Social network sites: Definition, history, and scholarship	1 695
10	Hierarchical organization of modularity in metabolic networks	1 535
11	GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox	1 467
12	Sequencing technologies the next generation	1 442
13	Assortative mixing in networks	1 364
14	A Protein Interaction Map of Drosophila melanogaster	1 357
15	Modularity and community structure in networks	1 347
16	Uncovering the overlapping community structure of complex networks in nature and society	1 337
17	Complex brain networks: Graph theoretical analysis of structural and functional systems	1 229
18	Fast algorithm for detecting community structure in networks	1 139
19	Finding community structure in very large networks	1 084
20	The spread of obesity in a large social network over 32 years	1 064
21	Community detection in graphs	1 048
22	A human protein-protein interaction network: A resource for annotating the proteome	977
23	A survey of trust and reputation systems for online service provision	949
24	Five rules for the evolution of cooperation	904

25	Understanding individual human mobility patterns	897
26	Hybrid recommender systems: Survey and experiments	889
27	Flocking for multi-agent dynamic systems: Algorithms and theory	886
28	Stochasticity in gene expression: From theories to phenotypes	879
29	Functional cartography of complex metabolic networks	870
30	The architecture of complex weighted networks	865
31	The human disease network	842
32	Mapping the structural core of human cerebral cortex	820
33	Opinion mining and sentiment analysis	794
34	Spread of epidemic disease on networks	757
35	Network motifs: Theory and experimental approaches	754
36	Hierarchical organization in complex networks	739
37	Finding community structure in networks using the eigenvectors of matrices	727
38	Biological robustness	707
39	Adaptive governance of social-ecological systems	689
40	Optimizing search engines using clickthrough data	683

H Žebříček počtu citací pro autory

Pozice	Jméno autora	Počet citací
1	Newman, M.E.J.	86 704
2	Barabási, A.-L.	44 208
3	Sporns, O.	37 120
4	Bullmore, E.	29 619
5	Moreno, Y.	27 319
6	He, Y.	23 009
7	Fortunato, S.	22 077
8	Latora, V.	21 060
9	Stam, C.J.	20 617
10	Arenas, A.	18 771
11	Oltvai, Z.N.	18 485
12	Boccaletti, S.	18 160
13	Wang, J.	17 775
14	Kurths, J.	16 686
15	Chavez, M.	16 608
16	Hwang, D.-U.	15 674
17	Chen, G.	15 516
18	Zhou, T.	14 940
19	Vicsek, T.	14 939
20	Albert, R.	14 681
21	Tamura, K.	14 077
22	Nei, M.	14 069
23	Bassett, D.S.	14 044
24	Palla, G.	13 977
25	Havlin, S.	13 949
26	GuimerÄ , R.	13 933
27	Kumar, S.	13 926
28	Dudley, J.	13 772
29	Wang, B.-H.	13 756
30	Girvan, M.	13 566
31	Christakis, N.A.	13 323
32	Szabó, G.	13 181
33	Vespignani, A.	13 142
34	Wang, L.	12 445
35	Olfati-Saber, R.	12 266
36	Fowler, J.H.	12 117
37	Amaral, L.A.N.	12 017
38	Díaz-Guilera, A.	11 905
39	Clauset, A.	11 866
40	Zhou, C.	11 465

I Žebříček relativního počtu citací pro autory

Pozice	Jméno autora	Relativní počet citací
1	Dudley, J.	13 772
2	Fax, J.A.	10 284
3	Brin, S.	6 199
4	Mongru, D.A.	5 755
5	Nei, M.	4 689
6	Whitcher, B.	4 686
7	Farkas, I.	4 539
8	Van Wedeen, J.	4 227
9	Fáth, G.	3 901
10	Paris, D.	3 798
11	Lefebvre, E.	3 435
12 – 14	Shen-Orr, S.	3 321
12 – 14	Ayzenshtat, I.	3 321
12 – 14	Sheffer, M.	3 321
15	Somera, A.L.	2 946
16 – 18	Yoon, S.-H.	2 923
16 – 18	Jeon, Y.S.	2 923
16 – 18	Won, S.	2 923
19	Turtschi, A.	2 788
20	Ravasz, E.	2 602
21	Concha, L.	2 548
22	Favera, R.D.	2 515
23	Childs, B.	2 366
24	Salvador, R.	2 357
25	Guindon, S.	2 320
26	Nakarado, G.L.	2 152
27	Baliki, M.	2 090
28 – 32	Polson, H.E.J.	1 779
28 – 32	De Lartigue, J.	1 779
28 – 32	Rigden, D.J.	1 779
28 – 32	Reedijk, M.	1 779
28 – 32	Urbé, S.	1 779
33	Tamura, K.	1 759
34	Handsley, M.M.	1 731
35	Wiggins, C.	1 698
36	Suckling, J.	1 631
37	Zwi, J.D.	1 568
38	Hwang, D.-U.	1 567
39	Gascuel, O.	1 546
40	Spirin, V.	1 542

J Žebříček počtu citací pro země

Pozice	Název země	Počet citací
1	United States	259 532
2	China	65 143
3	United Kingdom	58 668
4	Germany	51 340
5	Spain	49 489
6	Italy	47 548
7	France	35 558
8	Hungary	22 077
9	Japan	19 875
10	Canada	19 416
11	Australia	18 491
12	Switzerland	17 794
13	Netherlands	16 687
14	Hong Kong	16 681
15	Israel	14 511
16	South Korea	10 904
17	Portugal	9 778
18	Brazil	9 431
19	Belgium	9 355
20	Mexico	8 237
21	Sweden	7 502
22	Finland	6 404
23 – 24	Denmark	6 075
23 – 24	Slovenia	6 075
25	Russian Federation	4 682
26	India	3 981
27	Argentina	3 715
28	Austria	3 343
29	Georgia	3 255
30	New Zealand	3 231
31	Singapore	3 100
32	Greece	2 500
33	Poland	2 415
34	South Africa	2 220
35	Norway	2 094
36	Taiwan	1 864
37	Ireland	1 764
38	Czech Republic	1 140
39	Turkey	1 126
40	Cuba	1 050

K Žebříček relativního počtu citací pro země

Pozice	Název země	Relativní počet citací
1	Chad	9.111
2	Hungary	8.996
3	United States	5.879
4	Hong Kong	5.732
5	Spain	5.603
6	China	5.221
7	Cuba	4.929
8	Italy	4.926
9	North Korea	3.750
10	Israel	3.682
11	Slovenia	3.523
12	Germany	3.423
13	United Kingdom	3.403
14	Portugal	3.350
15	Switzerland	3.036
16	Burundi	2.875
17	Bahamas	2.722
18	France	2.684
19	Benin	2.678
20	Mexico	2.455
21	Ecuador	2.404
22	South Korea	2.395
23	Japan	2.358
24	Albania	2.266
25	Canada	2.235
26	Netherlands	2.183
27	Australia	2.122
28	Finland	1.996
29	Armenia	1.928
30	Belgium	1.901
31	Brazil	1.690
32	Russian Federation	1.669
33	Cyprus	1.655
34	Sweden	1.622
35	Central African Republic	1.515
36	Venezuela	1.483
37	Palestine	1.454
38	Malawi	1.443
39	Argentina	1.400
40	Denmark	1.395

L Žebříček in-degree pro publikace

Pozice	Název publikace	Vstupní stupeň
1 – 2	Complex networks: Structure and dynamics	2 085
1 – 2	Finding and evaluating community structure in networks	2 085
3	The structure and function of complex networks	2 051
4 – 7	Statistical mechanics of complex networks	2 000
4 – 7	The anatomy of a large-scale hypertextual Web search engine 1	2 000
4 – 7	Network biology: Understanding the cell's functional organization	2 000
4 – 7	MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0	2 000
8	Consensus and cooperation in networked multi-agent systems	1 906
9	Social network sites: Definition, history, and scholarship	1 695
10	Hierarchical organization of modularity in metabolic networks	1 535
11	GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox	1 467
12	Sequencing technologies the next generation	1 442
13	Assortative mixing in networks	1 364
14	A Protein Interaction Map of Drosophila melanogaster	1 357
15	Modularity and community structure in networks	1 347
16	Uncovering the overlapping community structure of complex networks in nature and society	1 337
17	Complex brain networks: Graph theoretical analysis of structural and functional systems	1 229
18	Fast algorithm for detecting community structure in networks	1 139
19	Finding community structure in very large networks	1 084
20	The spread of obesity in a large social network over 32 years	1 064
21	Community detection in graphs	1 048
22	A human protein-protein interaction network: A resource for annotating the proteome	977
23	A survey of trust and reputation systems for online service provision	949
24	Five rules for the evolution of cooperation	904

25	Understanding individual human mobility patterns	897
26	Hybrid recommender systems: Survey and experiments	889
27	Flocking for multi-agent dynamic systems: Algorithms and theory	886
28	Stochasticity in gene expression: From theories to phenotypes	879
29	Functional cartography of complex metabolic networks	870
30	The architecture of complex weighted networks	865
31	The human disease network	842
32	Mapping the structural core of human cerebral cortex	820
33	Opinion mining and sentiment analysis	794
34	Spread of epidemic disease on networks	757
35	Network motifs: Theory and experimental approaches	754
36	Hierarchical organization in complex networks	739
37	Finding community structure in networks using the eigenvectors of matrices	727
38	Biological robustness	707
39	Adaptive governance of social-ecological systems	689
40	Optimizing search engines using clickthrough data	683

M Žebříček in-degree pro autory

Pozice	Jméno autora	Vstupní stupeň
1	Barabási, A.-L.	13 322
2	Newman, M.E.J.	12 931
3	Tamura, K.	11 078
4	Nei, M.	11 071
5	Kumar, S.	10 964
6	Dudley, J.	10 855
7	Oltvai, Z.N.	8 834
8	Wang, J.	5 856
9	Albert, R.	5 224
10	Sporns, O.	5 218
11	Moreno, Y.	4 639
12	Latora, V.	4 589
13	Li, Y.	4 527
14	Ravasz, E.	4 454
15	Fortunato, S.	4 380
16	Bullmore, E.	4 302
17	Christakis, N.A.	4 281
18	Wang, Y.	4 125
19	Brin, S.	4 057
20	Fowler, J.H.	4 039
21	Girvan, M.	4 026
22	Boccaletti, S.	3 917
23	Guimerà, R.	3 902
24	Chavez, M.	3 842
25	Somera, A.L.	3 836
26 – 27	Guindon, S.	3 782
26 – 27	Gascuel, O.	3 782
28	Mongru, D.A.	3 766
29	Zhang, Y.	3 762
30	Li, X.	3 698
31	Hwang, D.-U.	3 694
32	Liu, Y.	3 641
33	Vespignani, A.	3 623
34	Amaral, L.A.N.	3 555
35	Wang, X.	3 535
36	Li, J.	3 531
37	Gouy, M.	3 474
38	Vicsek, T.	3 456
39	Wang, L.	3 451
40	Xu, X.	3 326

N Žebříček relativního in-degree pro autory

Pozice	Jméno autora	Relativní vstupní stupeň
1	Dudley, J.	10 855
2	Brin, S.	4 057
3	Mongru, D.A.	3 766
4	Nei, M.	3 690
5	Fax, J.A.	2 011
6	Somera, A.L.	1 918
7	Guindon, S.	1 891
8 – 12	Polson, H.E.J.	1 605
8 – 12	De Lartigue, J.	1 605
8 – 12	Rigden, D.J.	1 605
8 – 12	Reedijk, M.	1 605
8 – 12	Urbé, S.	1 605
13	Ravasz, E.	1 484
14 – 16	Yoon, S.-H.	1 428
14 – 16	Jeon, Y.S.	1 428
14 – 16	Won, S.	1 428
17	Handsley, M.M.	1 427
18	Whitcher, B.	1 408
19	Tamura, K.	1 384
20	Kovács, A.	1 327
21	Van Wedeen, J.	1 317
22 – 26	Mauthe, M.	1 275
22 – 26	Jacob, A.	1 275
22 – 26	Freiberger, S.	1 275
22 – 26	Hentschel, K.	1 275
22 – 26	Stierhof, Y.-D.	1 275
27	Robenek, H.	1 273
28	Farkas, I.	1 261
29 – 30	Gascuel, O.	1 260
29 – 30	Knodler, L.A.	1 260
31 – 32	Piggott, N.	1 256
31 – 32	Cook, M.A.	1 256
33 – 34	Fueyo-Margareto, J.	1 253
33 – 34	Gewirtz, D.	1 253
35 – 38	Øverbye, A.	1 252
35 – 38	Sætre, F.	1 252
35 – 38	Hagen, L.K.	1 252
35 – 38	Johansen, H.T.	1 252
39	Lefebvre, E.	1 214
40	Shen-Orr, S.	1 203

O Žebříček in-degree pro země

Pozice	Název země	Vstupní stupeň
1	United States	162
2	United Kingdom	148
3 – 5	France	136
3 – 5	Germany	136
3 – 5	Japan	136
6	China	123
7 – 8	Australia	120
7 – 8	Canada	120
9 – 10	Italy	116
9 – 10	Netherlands	116
11 – 12	Mexico	111
11 – 12	Spain	111
13	Brazil	106
14	Switzerland	105
15	New Zealand	102
16	Denmark	101
17 – 18	Belgium	96
17 – 18	Hungary	96
19	South Korea	93
20 – 21	South Africa	91
20 – 21	Sweden	91
22 – 23	Israel	90
22 – 23	Russian Federation	90
24	Finland	88
25	Hong Kong	86
26	India	81
27	Norway	78
28 – 29	Singapore	77
28 – 29	Thailand	77
30	Czech Republic	75
31	Portugal	74
32 – 33	Ireland	72
32 – 33	Taiwan	72
34	Niger	70
35 – 36	Argentina	66
35 – 36	Georgia	66
37	Austria	65
38	Poland	61
39 – 40	Greece	60
39 – 40	Slovenia	60

P Žebříček relativního in-degree pro země

Pozice	Název země	Relativní vstupní stupeň
1	Chad	1.888
2	Burundi	1.750
3	New Caledonia	1.000
4	North Korea	0.750
5	Mauritania	0.710
6	Cape Verde	0.666
7	Botswana	0.636
8	Bahamas	0.611
9	Nicaragua	0.578
10 – 11	Sudan	0.566
10 – 11	Togo	0.566
12	Albania	0.555
13	Ecuador	0.516
14	Palestine	0.515
15 – 16	Dominican Republic	0.500
15 – 16	Jamaica	0.500
17	Armenia	0.464
18	Central African Republic	0.453
19	Macao	0.428
20	Namibia	0.400
21	Seychelles	0.400
22	Benin	0.392
23 – 24	Mauritius	0.333
23 – 24	Suriname	0.333
25	Paraguay	0.326
26	Mozambique	0.312
27	Mali	0.310
28	Djibouti	0.307
29	Kuwait	0.303
30	French Guiana	0.296
31	Virgin Islands (U.S.)	0.285
32	Guinea	0.264
33	Nepal	0.263
34	Cuba	0.253
35 – 36	Laos	0.250
35 – 36	Liechtenstein	0.250
37	Sri Lanka	0.246
38	Panama	0.245
39	Malawi	0.210
40	Morocco	0.198

Q Žebříček PageRank pro publikace

Pozice	Název publikace	Ohodnocení
1	Statistical mechanics of complex networks	0.07166174
2	MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0	0.05572580
3	The structure and function of complex networks	0.02375036
4	Hierarchical organization of modularity in metabolic networks	0.01939938
5	The anatomy of a large-scale hypertextual Web search engine 1	0.01205546
6	Large-scale topological and dynamical properties of the Internet	0.01085634
7	Network biology: Understanding the cell's functional organization	0.00904166
8	Fluctuation-driven dynamics of the Internet topology	0.00602721
9	The spread of obesity in a large social network over 32 years	0.00414788
10	Assortative mixing in networks	0.00397359
11	Highly clustered scale-free networks	0.00390346
12	Hierarchical organization in complex networks	0.00365149
13	Tissue architecture: The ultimate regulator of breast epithelial function	0.00330967
14	Attack vulnerability of complex networks	0.00330862
15	Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences	0.00292303
16	The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group	0.00288576
17	Consensus and cooperation in networked multi-agent systems	0.00271608
18	Finding and evaluating community structure in networks	0.00235339
19	Ultrafast consensus in small-world networks	0.00224817
20	A Protein Interaction Map of Drosophila melanogaster	0.00223168
21	Organization, development and function of complex brain networks	0.00220718
22	The ADAM metalloproteinases	0.00220708
23	Optimizing search engines using clickthrough data	0.00219654

24	Network structure and biodiversity loss in food webs: Robustness increases with connectance	0.00216005
25	Scale-free topology of e-mail networks	0.00204751
26	Food-web structure and network theory: The role of connectance and size	0.00193991
27	Modular organization of cellular networks	0.00191267
28	Looking for inspiration: New perspectives on respiratory rhythm	0.00185546
29	Assessing experimentally derived interactions in a small world	0.00172901
30	Inferring genetic networks and identifying compound mode of action via expression profiling	0.00171698
31	Mixing patterns in networks	0.00164946
32	The EigenTrust algorithm for reputation management in P2P networks	0.00161675
33	Spread of epidemic disease on networks	0.00159487
34	Sociology: Team assembly mechanisms determine collaboration network structure and team performance	0.00158990
35	Classification of scale-free networks	0.00156614
36	Flocking for multi-agent dynamic systems: Algorithms and theory	0.00151657
37	Superfamilies of Evolved and Designed Networks	0.00147726
38	Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building	0.00143197
39	Dynamic network visualization	0.00140863
40	Pseudofractal scale-free web	0.00138545

R Žebříček PageRank pro autory

Pozice	Jméno autora	Ohodnocení
1	Dudley, J.	0.01212882
2	Kumar, S.	0.01212134
3	Tamura, K.	0.01204339
4	Barabási, A.-L.	0.01011045
5	Nei, M.	0.01006596
6	Newman, M.E.J.	0.00650166
7	Oltvai, Z.N.	0.00616644
8	Albert, R.	0.00533644
9	Ravasz, E.	0.00426807
10	Mongru, D.A.	0.00422248
11	Somera, A.L.	0.00395130
12	Brin, S.	0.00379672
13	Christakis, N.A.	0.00312825
14	Fowler, J.H.	0.00311647
15	Pastor-Satorras, R.	0.00250220
16	Vespignani, A.	0.00247849
17	Vázquez, A.	0.00203876
18	Roth, F.P.	0.00198836
19	Goldberg, D.S.	0.00197779
20	Gascuel, O.	0.00179360
21	Guindon, S.	0.00179360
22	Eguíluz, V.M.	0.00121701
23	Arenas, A.	0.00120947
24	Vicsek, T.	0.00118538
25	Sporns, O.	0.00117105
26	Girvan, M.	0.00116008
27	Amaral, L.A.N.	0.00115783
28	Guimerà, R.	0.00113855
29	Moreno, Y.	0.00111900
30	Gouy, M.	0.00111834
31	Joachims, T.	0.00109484
32	Fortunato, S.	0.00106429
33	Clauset, A.	0.00104348
34	Palla, G.	0.00102552
35	Watts, D.J.	0.00101843
36	Kahng, B.	0.00099786
37	Latora, V.	0.00097391
38	Moore, C.	0.00096984
39	Alon, U.	0.00092467
40	Olfati-Saber, R.	0.00091340

S Žebříček PageRank pro země

Pozice	Název země	Ohodnocení
1	United States	0.10308157
2	United Kingdom	0.03257135
3	Germany	0.02478903
4	Spain	0.02146737
5	Italy	0.02117557
6	France	0.01918751
7	China	0.01895957
8	Japan	0.01475885
9	Australia	0.01119634
10	Canada	0.01068479
11	Hungary	0.00970883
12	Netherlands	0.00903922
13	Switzerland	0.00856105
14	Israel	0.00724833
15	Belgium	0.00605430
16	Mexico	0.00602845
17	South Korea	0.00601443
18	Hong Kong	0.00569060
19	Brazil	0.00553130
20	Denmark	0.00524428
21	Finland	0.00491720
22	Portugal	0.00487227
23	Sweden	0.00457041
24	South Africa	0.00409655
25	New Zealand	0.00398088
26	Russian Federation	0.00361700
27	India	0.00310601
28	Argentina	0.00272076
29	Slovenia	0.00272028
30	Georgia	0.00251926
31	Austria	0.00231766
32	Singapore	0.00221741
33	Norway	0.00211531
34	Greece	0.00205641
35	Poland	0.00197558
36	Ireland	0.00194439
37	Taiwan	0.00186232
38	Cuba	0.00181519
39	Czech Republic	0.00177597
40	Niger	0.00171409

T Žebříček HITS pro publikace

Pozice	Název publikace	Ohodnocení
1	Finding and evaluating community structure in networks	0.42280319
2	The structure and function of complex networks	0.25917272
3	Modularity and community structure in networks	0.24755538
4	Uncovering the overlapping community structure of complex networks in nature and society	0.24273152
5	Fast algorithm for detecting community structure in networks	0.22662740
6	Complex networks: Structure and dynamics	0.22573924
7	Finding community structure in very large networks	0.22172057
8	Statistical mechanics of complex networks	0.21816200
9	Community detection in graphs	0.18524781
10	Functional cartography of complex metabolic networks	0.17543898
11	Resolution limit in community detection	0.15991472
12	Hierarchical organization of modularity in metabolic networks	0.15026028
13	Defining and identifying communities in networks	0.14216663
14	Finding community structure in networks using the eigenvectors of matrices	0.13924229
15	Assortative mixing in networks	0.12446822
16	Community detection in complex networks using extremal optimization	0.12347593
17	Detecting community structure in networks	0.11212515
18	Comparing community structure identification	0.11113681
19	Fast unfolding of communities in large networks	0.11091288
20	Network biology: Understanding the cell's functional organization	0.10876132
21	Complex brain networks: Graph theoretical analysis of structural and functional systems	0.08700211
22	Hierarchical organization in complex networks	0.07944278
23	The architecture of complex weighted networks	0.07574121
24	Benchmark graphs for testing community	0.07234369

	detection algorithms	
25	Maps of random walks on complex networks reveal community structure	0.07233004
26	Modularity from fluctuations in random graphs and complex networks	0.07010920
27	Statistical mechanics of community detection	0.06787442
28	Detecting fuzzy community structures in complex networks with a potts model	0.06609773
29	Mixing patterns in networks	0.06605253
30	Detecting the overlapping and hierarchical community structure in complex networks	0.06176602
31	Analysis of weighted networks	0.05678556
32	Community detection algorithms: A comparative analysis	0.05607504
33	A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs	0.05593319
34	Self-similar community structure in a network of human interactions	0.05271995
35	Synchronization reveals topological scales in complex networks	0.05266741
36	Protein complexes and functional modules in molecular networks	0.05230703
37	Near linear time algorithm to detect community structures in large-scale networks	0.04979377
38	Mapping the structural core of human cerebral cortex	0.04891533
39	Quantifying social group evolution	0.04787740
40	An information-theoretic framework for resolving community structure in complex networks	0.04696254

U Žebříček HITS pro autory

Pozice	Jméno autora	Ohodnocení
1	Barabási, A.-L.	0.10864902
2	Newman, M.E.J.	0.10591641
3	Wang, J.	0.10003369
4	Li, Y.	0.09272973
5	Wang, X.	0.08080426
6	Wang, Y.	0.07971531
7	Li, J.	0.07943527
8	Oltvai, Z.N.	0.07884347
9	Li, X.	0.07760623
10	Wang, L.	0.07440698
11	Zhang, Z.	0.07370787
12	Xu, X.	0.07370483
13	Wang, B.	0.06891008
14	Tang, H.	0.06771861
15	Liu, Y.	0.06534580
16	Wang, Z.	0.06462013
17	Moreno, Y.	0.06450631
18	Latora, V.	0.06426180
19	Albert, R.	0.06217509
20	Song, C.	0.06155805
21	Li, Z.	0.06019284
22	Boccaletti, S.	0.05890592
23	Chavez, M.	0.05861800
24	Huang, S.	0.05843620
25	Wang, H.	0.05818071
26	Zhang, S.	0.05769161
27	Ravasz, E.	0.05766462
28	Guimerà, R.	0.05755507
29	Zhang, J.	0.05741092
30	Fortunato, S.	0.05722371
31	Sporns, O.	0.05636722
32	Hwang, D.-U.	0.05633668
33	Girvan, M.	0.05543889
34	Kumar, S.	0.05536213
35	Wu, J.	0.05487551
36	Zhang, Y.	0.05480609
37	Liu, X.	0.05479232
38	Fang, L.	0.05436193
39	Amaral, L.A.N.	0.05433690
40	Arenas, A.	0.05428060

V Žebříček HITS pro země

Pozice	Název země	Ohodnocení
1	United States	0.86436690
2	Spain	0.20653399
3	Germany	0.20569834
4	Italy	0.19908870
5	United Kingdom	0.18500411
6	France	0.13605074
7	Hong Kong	0.12218229
8	Hungary	0.10654518
9	Switzerland	0.08043843
10	Canada	0.07912555
11	China	0.07867403
12	Japan	0.07190430
13	Australia	0.07154404
14	Israel	0.05784400
15	Netherlands	0.05596840
16	South Korea	0.04892461
17	Portugal	0.04333945
18	Brazil	0.03846508
19	Slovenia	0.03741154
20	Belgium	0.03639854
21	Mexico	0.03550494
22	Sweden	0.02863567
23	Finland	0.02343445
24	Denmark	0.02173477
25	Russian Federation	0.01814599
26	Singapore	0.01670336
27	India	0.01511137
28	Argentina	0.01478291
29	Austria	0.01331564
30	Georgia	0.01316424
31	Greece	0.00949241
32	New Zealand	0.00918548
33	Taiwan	0.00904337
34	Poland	0.00888994
35	Norway	0.00715396
36	Ireland	0.00624839
37	Turkey	0.00531652
38	South Africa	0.00501489
39	Cuba	0.00404206
40	Czech Republic	0.00325818