

Matej Valúšek

1. ročník magisterského štúdia

Dátum spracovania: 5.12.2013

URL: <http://books.google.com/>

Google Books

Projekt digitálnej knižnice Google Books dostupný na adrese <http://books.google.com/> nie je prvý ani posledný z podobných projektov, prečo je teda taký zaujímavý? Z pohľadu knižníc a univerzít je to prvý projekt, ktorý pre ne zdigitalizuje knihy zdarma. V prípade zapojenia sa do programu môžu dané inštitúcie získať reklamu a takisto pomôcť uchovať cenné knihy. Z pohľadu autorov sa však jedná o službu, ktorá im kradne peniaze tým, že ponúka ich diela verejnosti zadarmo, čo vyvrcholilo až súdnym sporom. Názory verejnosti sa preto líšia, zatiaľ čo jedna skupina súhlasí s politikou Googlu, druhá skupina sa obáva monopolného postavenia tejto spoločnosti a vidia za tým snahu kontrolovať digitálny svet.

Inšpirovaní projektmi digitalizácie v Amerických knižniciach, spoločnosť Google začala v roku 2002 pracovať na vlastnom knižnom systéme. V roku 2004 bol projekt oficiálne predstavený pod názvom „Google Print“.

V Decembri 2004 Google predstavil rozšírenie pod názvom „Google Print Library Project“. Cieľom „Library Project“ je naskenovať a zdigitalizovať čo najviac kníh a vytvoriť tak kolekciu, ktorá by bola



Logo Google Books

dostupná každému. Preto sa do projektu zapojili najmä knižnice a univerzity, ktoré poskytli Googlu knihy, medzi najznámejšími boli napríklad Univerzita Michigan, Harvard, Oxford a Stanford.

Nie všetky univerzity a knižnice sú však s daným projektom spokojné. Verejná knižnica v Dallase sa napríklad odmieta do programu zapojiť kým Google nevyrieši správu elektronických práv. Obávajú sa najmä negatívnej reklamy v prípade ďalších súdnych sporov. V súčasnosti však niektoré inštitúcie z projektu odstupujú pretože podľa nich ich Google podviedol a naskenované dokumenty nepoužíva na výučbu, správy alebo výskum, teda na aktivity, ktoré spadajú pod spravodlivé používanie ale používa ich na komerčné účely a takisto preto aby získal monopolné postavenie na digitálnom trhu kníh. Podľa najnovších výsledkov z roku 2013 Google databáza obsahuje približne 30 miliónov zdigitalizovaných kníh.

Keďže spoločnosť Google nerešpektovala autorské práva autorov kníh ktoré zdigitalizovala v „Library project“, bola v roku 2005 žalovaná Americkým cechom autorov a piatimi najväčšími vydavateľstvami v Amerike. Súdny proces skončil až po dvoch rokoch mimosúdnym vyrovnaním. Google vytvoril „Partnerský program“ zameraný najmä na autorov a vydavateľov kníh, v ktorom im dáva možnosť poskytnúť knihy ku skenovaniu alebo už priamo v elektronickej podobe. V prípade poskytnutia kníh im Google ponúka lepší marketing v podobe reklám, odkazov na kníhkupectvá a v poslednej dobe aj finančný podiel z príjmov na kontextovej reklame. Google ďalej vytvoril program „Google Play“ v ktorom umožňuje autorom poskytnúť svoje knihy na predaj priamo a to vo formátoch PDF alebo ePub (formát používaný v čítačkách). Autorom zapojeným do Partnerského programu Google garantuje zachovanie ich autorských práv a tvrdí, že obsah je uložený na zabezpečených serveroch. Na stránkach, ktoré zobrazujú dané knihy Google zakázal funkcie tlačenie, kopírovania a ukladania. Momentálne je v programe zapojených približne 20000 partnerov.

Pôvodným cieľom služby Google Books podľa jej autorov Sergeya Brina a Larryho Pagea bolo vytvoriť vyhľadávací systém, ktorý by: „indexoval obsah kníh, analyzoval spojenie medzi nimi a určoval tak jejich dôležitosť a užitečnosť na základe sledovania počtu a kvality citácií z jiných kníh“ [1]. V súčasnosti služba Google Books umožňuje vyhľadávať knihy pomocou fulltextového vyhľadávania nielen v názve a bibliografii ale aj v obsahu knihy. Jedným z ďalších cieľov je aj digitalizácia kníh pretože v roku 2010 spoločnosť Google vyhlásila, že do konca roku 2020 má v úmysle naskenovať všetky existujúce knihy, ktorých približný počet podľa Googlu je 130 miliónov. Postupne sa k projektu Google Books pridáva čoraz viac univerzít a knižníc, pretože v ňom vidia možnosť ako zachrániť vzácne knihy pred prírodnými katastrofami, ktoré sú v podobe hurikánov v USA dosť časté. Aby splnili svoj

zámer, je Googlu dodávaných len od Kalifornskej Univerzity približne 3000 kníh denne.

Vznikajú však ohlasy, že Google preferuje kvantitu pred kvalitou, podporujú to najmä tvrdenia expertov, ktorí sa venujú digitalizácii kníh. Obviňujú Google, že proces digitalizácie zjednodušuje a namiesto komplexného procesu, ktorý trvá pár hodín na naskenovaných stránkach len zvýši belosť a sýtosť čiernej farby, aby ušetril čo najviac času. Google dokonca pri ukladaní naskenovaných kníh nepoužíva OCR (Optical Character Recognition) a neukladá text ale



Naskenovaná ruka z Google Books

obrázky strán, čo im na jednej strane šetrí čas, no na strane druhej ľudia, ktorý majú problémy so zrakom a používajú hlasový prehrávač nemajú potom žiadnu šancu si knihy prečítať. Niektorí ľudia dokonca tvrdia, že na skenovanie Google nepoužíva drahé skenovacie prístroje, ktoré pracujú samostatne. Namiesto toho sa strany otáčajú manuálne a podporujú svoje tvrdenia obrázkami z Google Books na ktorých sú naskenované prsty ľudí. Samotný Google sa k tejto otázke a ani k tomu, aké stroje na skenovanie používajú nevyjadril. Jediné vyjadrenie Googlu bolo, že na naskenovanie

[Sale Catalogues - Vydanie 792 - Strana 84](#)



[books.google.com/books?id... - Preložiť túto stránku](#)
American Art Association, Anderson Galleries (Firm) - 1590 - Plné znenie -
Ďalšie vydania
The theatre on this occasion was licensed to Dickens, and the whole was produced, under his direction. The original production ... 1852 * With original unpublished poem of four lines in the autograph of Charles Dickens, signed by him in full ...

[Catalogue / Pickering and Chatto, firm, booksellers, London](#)



[books.google.com/books?id... - Preložiť túto stránku](#)
1619 - Plné znenie
DICKENS (CHARLES). 12 1 82 The Adventures of Oliver Twist; or, THE PARISH BOY'S PROGRESS. By CHARLES DICKENS. With twenty-four illustrations on Steel by GEORGE CRUIKSHANK. A New Edition, revised and corrected. Bradbury ...

[Catalogue - Vydanie 245 - Strana 1844](#)





[books.google.com/books?id... - Preložiť túto stránku](#)
Pickering & Chatto - 1659 - Plné znenie - Ďalšie vydania
DICKENS (CHARLES). 12 1 82 The Adventures of Oliver Twist; or, THE PARISH BOY'S PROGRESS. By Charles Dickens. With twenty-four illustrations on Steel by GEORGE CRUIKSHANK. A New Edition, revised and corrected. Bradbury and ...

Vyhľadanie Charlesa Dickensa pred rokom 1812, ukážka zlého datovania

takého obrovského objemu kníh je potrebných množstvo ľudí. Ďalším nedostatkom Google Books je veľký počet chybných metadát. Medzi časté chyby patrí napríklad náhodné dátumy priradené ku knihe, alebo nekonzistentnosť záznamov, keď je k jednej knihe priradených viacero žánrov. Chybné datovanie však neplatí len pre knihy, ale takisto aj pre známe osobnosti alebo frázy. Podľa

Univerzity v Berkeley pri vyhľadávaní pred rokom 1920 a po zadaní frázy „candy bar“ Google Books ponúka 66 záznamov, z ktorých má 46 zlý dátum. Google Books má takisto problém s klasifikáciou jednotlivých hnutí alebo historických osôb. Priraduje ich ku knihám, ktoré s nimi nemajú nič spoločné. Chyby sa však týkajú aj samotných autorov textov. Za autora textu môže byť označený napríklad len autor predslovu alebo autor ilustrácií v knihe. Google má však na tieto tvrdenia svojský názor a prehlasuje, že za chyby môžu jednotlivé knižnice, ktoré im dáta poskytli. Otázkou však zostáva, kedy a či vôbec dôjde k náprave týchto chýb keďže momentálne má Google iné priority.



<p>Hľadáte informácie na konkrétnu tému?</p> <p>Vyhľadávajte v najnovšom indexe kníh zo sveta. Nájdite milióny úžasných kníh a zobrazte si ich ukážku alebo si ich bezplatne prečítajte.</p> <input type="text"/> <input type="button" value="Vyhľadávanie kníh"/>	
<p>Novinka! Nakupujte knihy v službe Google Play</p> <p>Nájdite si knihu v najväčšom e-knihkupectve na svete a pustite sa do čítania na webe, tablete, telefóne alebo čítačke e-kníh ešte dnes.</p> <input type="button" value="Prejsť do služby Google Play »"/>	 Google play

Moja knižnica

Úvodná stránka vyhľadávania v Google Books

Služba Google Books sa v podstate skladá z dvoch programov - Projekt knižnica a Partnerský program. Projekt knižnica je zameraný na spoluprácu s univerzitami a knižnicami, zatiaľ čo partnerský program dovoľuje súčasným autorom vytvoriť si účet a spravovať knihy, ktoré autor poskytol do Google Books. Vyhľadávanie v Google Books je veľmi podobné ako v klasickom Google vyhľadávači, teda obsahuje jedno okno, kde užívateľ zadá hľadaný výraz. Ako už bolo spomenuté, služba Google Books umožňuje fulltextové vyhľadávanie nie len v knihách ale aj v časopisoch a článkoch. Google Books nám samozrejme umožňuje rozšírené vyhľadávanie použitím napríklad mena autora, názvu knihy, jazyka, ISBN a ďalších parametrov.

Knihy môžeme rozdeliť do štyroch skupín:

- V prvej skupine text knihy nie je k dispozícii, Google Books však ponúka aspoň základné bibliografické údaje.
- V druhej skupine sú knihy, ktoré autori odmietli zapojiť do partnerského programu. V prípade, že sa takúto knihu alebo článok snažíte vyhľadať Google Books vám ponúkne takzvaný "snippet" čiže zobrazenie fragmentu a pridá odkazy, kde daný dokument možno zakúpiť. Bohužiaľ toto nemusí byť postačujúce, obzvlášť pri odborných prácach, kde snippet tvorí maximálne jednu vetu, z ktorej ťažko posúdiť či článok obsahuje hľadanú problematiku.
- Tretia skupina umožňuje limitované zobrazenie, teda povolí zobrazit' určitý počet strán. Pri nájdení konkrétnej knihy je pre prehliadač vygenerovaný tzv.



Buď ste sa dostali na stranu, ktorú nemožno zobraziť, alebo ste dosiahli limit zobrazenia pre túto knihu.

Obrázok znázorňujúci limitované zobrazenie

cookie, ktorý obsahuje informácie o čase a počte strán , ktoré môže užívateľ vidieť. Istý študent informatiky však dokázal napísať program, ktorý zbieral tieto cookies a pomocou nich dokázal stiahnuť celú knihu za týždeň. Momentálne sú však chyby, ktoré jeho program využíval už odstránené.

- V poslednej, štvrtej, skupine sa nachádzajú voľne dostupné celé texty kníh, ktorých autorské právo už vypršalo poprípade knihy, ktoré autor ponúkol sám.

Každý knižný záznam však obsahuje mimo textu samotného aj ďalšie údaje ako napríklad: „obálka, obsah, obľíbené pasáže, ďalší vydání, recenze, odkazy z webových stránek, odkazy z vědeckých prací, odkazy z knih, místa zmíněná v knize s vyznačením pozic na mapě Google-maps“ [1]. Registrovaní užívatelia majú dokonca možnosť v rámci svojej „knižnice“ písať recenzie a pridávať odkazy na web. Týmto spôsobom sa do Google Books pridávajú nové metadáta, ktoré sa pri ďalšom vyhľadávaní používajú a vzťahy medzi knihami tak vytvárajú akýsi knižný web.

Na jednej strane tento prístup uľahčuje Google technikom prácu, no na strane druhej to má za následok výskyt veľkého množstva chýb v metadátach, keďže takto vytvorené dáta nikto nekontroluje. Otázkou preto zostáva, či je lepšie mať menej prístupných dokumentov s korektnými dátami ako je to vo väčšine knižníc alebo obrovské množstvo článkov a kníh ktorých metadáta sú chybné a zavádzajúce.

Google Books je podľa môjho názoru veľkým prínosom hlavne čo sa týka dostupnosti niektorých ťažko získateľných kníh. Takisto v prípade cudzojazyčnej odbornej literatúry, ktorá je v knižnici ťažko prístupná. Jednou z malých nevýhod je dominancia anglickej literatúry, a následný nedostatok slovenskej a českej literatúry. Zaujímavosťou je aj to, že aktuálna verzia Google Books je stále prezentovaná ako beta verzia a v budúcnosti sa snáď dočkáme finálnej verzie. Celý projekt sa však ešte môže skomplikovať pretože nie všetky súdne procesy s autormi sú vyriešené a môže dôjsť k obmedzeniu momentálne dostupnej literatúry. Odhliadnuc od toho ako spory dopadnú, podstatné podľa mňa je, že nesmierne množstvo starých kníh je zdigitalizovaných a tak v prípade prírodnej alebo inej katastrofy budú uchované.

Literatúra

[1] M. Bartošek. Nástroje Google. 3. Google Book Search. Zpravodaj ÚVT MU. ISSN 1212-0901, 2009, roč. XIX, č. 3, s. 16-20.

<http://books.google.com/intl/cs/googlebooks/about.html>

<http://people.ischool.berkeley.edu/~nunberg/GBook/GoogBookMetadataSh.pdf>

http://en.wikipedia.org/wiki/Google_Book_Search

<http://campustechnology.com/Articles/2008/01/Google-Book-Search-The-Good-the-Bad-and-the-Ugly.aspx>

<http://www.kuro5hin.org/story/2005/3/7/95844/59875>

<http://www.aallnet.org/main-menu/Publications/spectrum/Archives/vol-17/No-7/google-books.pdf>

<https://support.google.com/books/partner/answer/1079107>

Dublin Core

```
<?xml version="1.0" ?>
<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcq="http://purl.org/dc/documents/rec/dcmes-qualifiers-20000711.htm">
<rdf:Description rdf:about="Google Books">
<dc:title>Google Books</dc:title>
<dc:description>Digitálna knižnica, ktorá obsahuje knihy, články a
časopisy</dc:description>
<dc:date>2013-12-08</dc:date>
<dc:creator>Matej Valúšek</dc:creator>
<dc:format>application</dc:format>
<dc:rights>policy:public</dc:rights>
<dc:type>text</dc:type>
<dc:language>sk</dc:language>
</rdf:Description>
</rdf:RDF>
```