

Masarykova univerzita

Fakulta informatiky

Digitální knihovny PV070 – závěrečná esej

9. prosince 2009

Karel Svačina

Aplikovaná informatika (3. ročník Bc. studia)

Sociologie (3. ročník Mgr. studia)

Některé otazníky kolem služby Google Book Search

Co je Google Book Search

Google Book Search¹ je služba společnosti Google pro fulltextové vyhledávání v obsahu knih. Google Book Search (dále GBS) pracuje s rozsáhlou databází digitalizovaných knih (v současnosti přes deset milionů svazků), jejichž obsah je indexován podobně jako je obsah webových stránek indexován službou Google Search². Uživatel tedy vyhledává v obsahu knih podobným způsobem jako u běžného vyhledávání Google. Po vyhledání patřičné knihy si může uživatel prohlédnout bližší údaje o knize, v závislosti na autorských právech konkrétního díla si může prohlédnout krátké úryvky nebo delší pasáže, nebo dokonce stáhnout celé dílo do svého počítače. V neposlední řadě je uživatel odkázán na internetové obchody, kde lze prohlíženou knihu zakoupit. GBS je velmi rozsáhlý a ambiciózní projekt, jenž si klade za cíl mimo jiné zpřístupnit mnoho obtížně dostupných knih. Do projektu se zapojilo mnoho prestižních a velkých knihoven. Avšak GBS také provází mnoho problémů a nejasností a jako každý velký a ambiciózní projekt je GBS častým terčem kritiky. Cílem této eseje je upozornit na některé z těchto problémů a kritik a nabídnout letný interdisciplinární pohled na službu Google Book Search. Nejdříve však stručně z historie služby a její současný stav.

Stručně z historie služby, aktuální stav

Služba firmy Google pro vyhledávání v obsahu knih byla poprvé představena v říjnu 2004 na knižním veletrhu ve Frankfurtu, pod jménem Google Print. Tato služba byla nejprve zaměřena spíše na nakladatele a vydavatele knih. Na konci téhož roku ale Google oznámil, že bude v rámci projektu Google Print spolupracovat také s některými špičkovými akademickými knihovnami – v té době to byla knihovna Michiganské univerzity, Harvardské univerzity a Stanfordské univerzity ve Spojených státech, dále Veřejná knihovna v New Yorku (New York Public Library) a také knihovna Oxfordské univerzity ve Velké Británii. Na podzim roku 2005 byla služba přejmenována na „Google Book Search“. Tento název je používán dodnes. (Wikipedia 2009)

Google Book Search se tedy skládá ze dvou částí – zatímco v rámci Google Books Partner Program mohou nakladatelé a autoři knih zařazovat svoje díla do databáze Google, v rámci Google Books Library Project dávají akademické a veřejné knihovny části svých fondů k dispozici firmě Google, která knihy skenuje, rozeznává pomocí OCR a indexuje. Z hlediska uživatele služby se rozdíl mezi těmito dvěma částmi stírá – GBS nabízí jen jedno rozhraní pro vyhledávání knih z obou částí projektu současně. Po nalezení a výběru knihy se však uživatel může setkat s různým množstvím informací o daném díle, v závislosti na autorských právech a na tom, kde se uživatel nachází. U některých knih může uživatel vidět jen abstrakt a základní bibliografické údaje³, u jiných lze shlédnout fragmenty textu (v angličtině „snippets“)⁴, u dalších omezený počet stran⁵ a u ostatních pak text v plném znění⁶.

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

2 <http://www.google.com>

3 Jako jeden příklad za všechny uvedme záznam o knize Oudshoorn, N., & Pinch, T. J. (2005). *How users matter: The co-construction of users and technology*. Inside technology. Cambridge, Mass: MIT. Viz <http://books.google.com/books?id=rXrHAAACAAJ&dq=how+users+matter&hl=cs&cd=1>.

4 Zde např. u knihy Dante Alighieri, White, L. G., & Doré, G. (1948). *The divine comedy: The Inferno, Purgatorio, and Paradiso*. New York: Pantheon Books. Viz http://books.google.com/books?id=HVhdAAAAMAAJ&q=alighieri&dq=alighieri&lr=&as_brr=0&hl=cs&cd=9

5 Např. Hine, C. (2000). *Virtual ethnography*. London: SAGE. Viz http://books.google.com/books?id=X5w1P2_iMNYC&printsec=frontcover&dq=christine+hine&hl=cs&cd=1#v=onepage&q=&f=false.

6 Např. Shakespeare, W. (1865). *Shakespeare's sonnets*. Boston: Ticknor and Fields. Viz

Objem databáze služby GBS roste ohromující rychlostí, a také počet institucí zapojených do projektu GBS se významně rozrůstá. K původním pěti knihovnám zapojeným do projektu GBS se přidalo mnoho dalších, například knihovny Kolumbijské a Cornellovy univerzity, Bavorská státní knihovna nebo Katalánská národní knihovna. Zatímco na jaře roku 2007 byl zpracován milion autorských děl (Hafner 2007), na podzim roku 2008 to bylo již sedm milionů (Wikipedia 2009) a na podzim roku 2009 více než deset milionů (Brin 2009). Na druhou stranu, tento masivní nárůst je v posledních letech provázen mnohými námitkami vůči službě GBS ze strany nakladatelů, knihoven i veřejnosti.

Cíle projektu Google Book Search a problémy a nejasnosti kolem něj

Proklamované cíle služby Google Book Search jsou v souladu s obecným cílem společnosti Google, jež zní: „uspořádat informace z celého světa tak, aby byly všeobecně přístupné a užitečné“ (Google 2009). Jak říká Sergey Brin, jeden ze zakladatelů firmy Google, cílem služby GBS je zpřístupnit světové kulturní dědictví co nejširšímu publiku. Toho má být dosaženo právě prostřednictvím Internetu. Brin argumentuje, že díla, která se nedostanou na Internet, budou ve svém důsledku ztracena, protože je nikdo nenajde (Brin 2009).

Uvedené cíle společnosti Google i jejího vyhledávání knih zní šlechetně, přesto se ale začalo objevovat množství námitek a kritiky proti GBS. Již roku 2005 podalo americké Sdružení autorů a také Americká asociace nakladatelů žaloby proti společnosti Google pro masivní porušování autorských práv (The Authors Guild 2009, Gilbert 2005). Američtí nakladatelé a autoři namítali, že Google porušuje autorská práva, když bez předchozího souhlasu nakladatelů či autorů skenuje knihy podléhající autorskému zákonu a dává být jen jejich úryvky číst svým uživatelům. Google se bránil klauzulí o „fair use“ neboli rozumném užívání v autorském zákoně a tvrdil, že svým počínáním nakladatele a autory nepoškozuje, ale naopak jejich díla přináší širšímu publiku, a tedy zvyšuje jejich prodej.

Na podzim roku 2008 došlo k dohodě o mimosoudním vyrovnání mezi firmou Google na jedné straně a nakladateli a autory na straně druhé. Důsledkem této dohody je, že vznikne nezávislá autorita pro dohlížení nad autorskými právy (Book Rights Registry). Google dále zaplatí vlastníkům autorských práv sumu 125 milionů amerických dolarů. Nejdůležitější je ale zřejmě to, že v závislosti na dohodách mezi vlastníky autorských práv a společností Google bude přístup k celým textům literárních děl podléhajících autorským právům možný, ale bude zpoplatněn (United States District Court 2009). Na jedné straně Google charakterizuje tuto dohodu jako „převratnou“, ale na druhé straně se proti ní zvedla další vlna kritiky ze strany bloggerů (Vaidhyanathan 2009), některých zúčastněných knihoven (Lyden 2009) a dokonce i ze strany národních vlád (Wiesmann a Waters 2009). Tento vývoj naznačuje, že z na první pohled nevinného ba dokonce šlechetného technologického projektu se stal předmět sporu s širokými právními, politickými a společenskými důsledky. Abychom mohli těmto širším souvislostem lépe porozumět, bylo by vhodné krátce se zastavit u nejdůležitějších aktérů či skupin, které hrají v projektu Google Book Search nějakou roli.

Významní aktéři okolo Google Book Search a jejich vztahy k této službě

Knihovny

V současné době je do projektu GBS zapojeno několik desítek prestižních akademických a veřejných knihoven. Jaké jsou jejich motivace pro zapojení do tohoto projektu? Cílem veřejných

<http://books.google.com/books?id=xSgGMZuLG4EC&printsec=frontcover&dq=shakespeare&lr=&as_drrb_is=b&as_minm_is=0&as_miny_is=&as_maxm_is=0&as_maxy_is=1920&as_brr=0&hl=cs&cd=10#v=onepage&q=&f=false>

knihoven v demokratických společnostech je chránit kulturní dědictví a zpřístupnit jej veřejnosti (srov. např. Lyden 2009). V době Internetu se mnoho knihoven obrací také k tomuto médiu a snaží se v rámci své veřejné služby nabízet své služby i jeho prostřednictvím⁷. Některé knihovny digitalizují své archivy, avšak často narážejí na technické či ekonomické problémy. Nabídka firmy Google digitalizovat jejich archiv jim proto může přijít velmi vhod⁸. Pro knihovny ale zůstává klíčové jejich poslání, tedy otevřený přístup čtenářů k jejich fondu, a proto jakékoliv omezení služby čtenářům, které nepramení z provozu knihovny či jejich úkolu chránit knihy, mohou nést s nevolí. Případem takového omezení může být právě zpoplatnění přístupu ke službě GBS. Není proto překvapením, že některé knihovny zareagovaly na dohodu o vyrovnání mezi firmou Google a autory a vydavateli ostrou kritikou. Například knihovna Harvardské univerzity v reakci na tuto dohodu přestala firmě Google poskytovat knihy podléhající autorským právům. Tento krok je z pohledu knihoven pochopitelný – knihovny poskytují firmě Google knihy zdarma, ale nyní se po nich bude chtít, aby za elektronický přístup do databáze Google platily. Navíc, dohodu o vyrovnání mezi firmou Google a vydavateli a autory mohou vnímat doslova jako dohodu „o nás bez nás“.

Zde je však zajímavé také to, že ne všechny knihovny se k dohodě staví tak jako ta harvardská. Například knihovna Michiganské univerzity v návaznosti na uzavření výše zmíněné dohody o vyrovnání upravila svůj smluvní vztah s firmou Google a i nadále se aktivně účastní projektu GBS (University of Michigan Library 2009). Dá se tedy říci, že než došlo k dohodě o vyrovnání mezi firmou Google a vydavateli resp. autory, knihovny téměř jednohlasně GBS iniciativu vítaly. Po uzavření dohody však začaly být některé knihovny mnohem skeptičtější.

Autoři a vydavatelé

Role a zájmy vydavatelů a potažmo autorů v souvislosti se službou GBS jsou v mnohém diametrálně odlišné než role a zájmy knihoven. Zatímco v zájmu knihoven je co nejširší dostupnost vědění a tedy knih, vydavatelé si svá autorská práva lidově řečeno střeží a hájí svůj ekonomický zájem v nakládání s literaturou. Ať už byla iniciativa GBS ve své původní podobě právně v pořádku nebo ne, vydavatelům se patrně příliš nelíbila. Přesněji řečeno, některé její části se vydavatelům nelíbily: nelíbilo se jim zejména to, že knihovny dávaly k dispozici díla podléhající autorskému zákonu a Google je digitalizoval a nechal v nich uživatele vyhledávat bez předchozího svolení vydavatelů. V praxi se tedy někteří vydavatelé aktivně účastnili té části projektu GBS, která zaštiťuje spolupráci mezi Googlem a vydavateli, zatímco ve stejnou dobu podávali trestní oznámení na druhou část projektu GBS, tu o spolupráci firmy Google s knihovnami.

Tak jako tak, díky dohodě o vyrovnání získávají vydavatelé mnohem větší kontrolu nad nakládáním s digitalizovanými knihami, jež spadají do autorského práva. Kromě toho jim dohoda zajišťuje finanční příjem z přístupu ke chráněným dílům. Dá se tedy říci, že za současné situace vydavatelé dosáhli svého.

Čtenáři

Pro individuálního čtenáře se služba GBS zdá být jednoznačným přínosem. Nejen že nabízí plně texty starých knih, často nabízí také náhledy nových knih. Je ale důležitou otázkou, jak se služba GBS změní po vstoupení dohody o vyrovnání v platnost. Zaprvé, dohoda umožní za úplaty zpřístupnit velké množství dosud nepřístupných knih. Ovšem zadruhé, součástí dohody je také možnost připojit k samotné prezentaci knihy inzerci (o jejíž příjem se dělí Google s vydavatelem

7 Viz např. digitální kolekce a služby Americké Kongresové knihovny, dostupné na adrese <http://www.loc.gov/library/libarch-digital.html>

8 Například digitalizace archivu knihovny Michiganské univerzity měla podle propočtů trvat asi tisíc let. Když přišla firma Google s nabídkou digitalizovat celý sedmimilionový archiv knihovny, odhadla dobu digitalizace na pouhých šest let (Google 2009b).

resp. autorem). Je tedy otázkou, jak se služba GBS s „komercializací“ z hlediska uživatele změní, zda bude i nadále tak atraktivní a zdánlivě užitečná.

Existují ale ještě další kritiky, jež přínos služby GBS pro uživatele jako čtenáře výrazně zpochybňují. Například Waller (2009) ukazuje, že význam slova „informace“ se ve světě knihoven a ve světě obchodních zájmů firmy Google diametrálně liší. Důsledkem toho je, že to, co by se dalo nazvat jako „potřeba informací“ ve světě knihoven, je na stránkách Google přeloženo do potřeb konzumních. Jinými slovy, informace ztrácí hodnotu vědění a místo toho funguje jako prostředek ke spotřebě. V této perspektivě zcela ztrácí služba GBS jako náhrada knihovnických katalogů svou legitimitu.

Google

Jak bylo nastíněno výše, oficiální rétorika firmy Google říká, že GBS je důležitou součástí firemní mise, která má za cíl „uspořádat informace z celého světa tak, aby byly všeobecně přístupné a užitečné“ (Google 2009). Z tohoto hlediska to vypadá, že GBS je ideologickou záležitostí, že ji Google dělá pro dobro lidstva. Avšak dohoda o vyrovnání mezi firmou Google a vydavateli nasvědčuje, že to není tak černobílé. Někteří komentátoři a autoři argumentují, že fakt, že si Google nestál za svým a nehájil legalitu svého počínání u soudu ukazuje, že v první řadě nejde o ideály, ale o obchod (srov. Perez 2008) – i v případě, že by soud rozhodl ve prospěch Google, bylo by velmi nepravděpodobné, že by Google mohl dosáhnout podobně výhodné situace jako nyní s dohodou o vyrovnání.

Jak z toho ven?

Předchozí část ukázala, že zájmy skupin, jež se účastní projektu GBS, jsou velmi různé a často protichůdné. Zatímco knihovny by chtěly vědění co nejvíce otevřít veřejnosti, nakladatelé musí hájit své obchodní zájmy. Google službu spustil s humanistickou rétorikou, avšak svým jednáním se přiklonil spíše svým obchodním zájmům. Je tedy zřejmé, že původně technologický projekt vyhledávání v obsahu knih byl opředen hustou a neprůhlednou sítí ekonomických, právních a společenských zájmů. Zatímco technologii digitalizace se firmě Google podařilo posunout o mnoho dál, ekonomické, právní a společenské vlivy dodávají této pozoruhodné technologické inovaci na neprůhlednosti a nejednoznačnosti. V tomto smyslu je individuální uživatel vydán protichůdným zájmům a jen stěží si může být jistý například relevancí informací vrácených vyhledávačem knih. Avšak situace přece jen není tak kritická. Google Book Search přece jen zůstává pozoruhodným počinem a co do počtu digitalizovaných knih daleko předčí svou konkurenci⁹. GBS se ostatně nepřímo podílel na vzniku dalších pozoruhodných projektů, například digitálního archivu HathiTrust¹⁰. Závěrem lze tedy říci, že Google Book Search v žádném případě nemůže být chápán jako univerzální či konečný způsob archivace knih a kulturního dědictví, avšak představuje v tomto směru významný a smělý krok.

9 Mezi konkurenční projekty lze počítat např. Open Library (openlibrary.org), Project Gutenberg (www.gutenberg.org), Europeana (europeana.eu) a jiné.

10 <http://hathitrust.org/>

Použitá literatura a zdroje

- The Authors Guild. 2009. „Authors Guild v. Google Settlement Resources Page.“ *The Authors Guild* [online]. Dostupné z: <<http://www.authorsguild.org/advocacy/articles/settlement-resources.html>>
- Brin, Sergey. 2009. „A Library to Last Forever.“ *The New York Times*, October 9, 2009 [online]. Dostupné z: <<http://www.nytimes.com/2009/10/09/opinion/09brin.html>>
- Gilbert, Alorie. 2005. „Publishers sue Google over book search project.“ *CNET News* [online]. Dostupné z <http://news.com.com/Publishers+sue+Google+over+book+search+project/2100-1030_3-5902115.html>.
- Google, Inc. 2009. „Informace o korporaci.“ [online]. Dostupné z: <<http://www.google.cz/intl/cs/corporate/>>
- Google, Inc. 2009a. „O službě Vyhledávání knih Google.“ [online]. Dostupné z: <<http://books.google.cz/intl/cs/googlebooks/about.html>>
- Google, Inc. 2009b. „About Google Books – History of Google Books“ [online]. Dostupné z: <<http://books.google.cz/intl/cs/googlebooks/history.html>>
- Hafner, Katie. 2007. „History, Digitized (and Abridged).“ *The New York Times* [online]. Dostupné z: <http://www.nytimes.com/2007/03/10/business/yourmoney/11archive.html>
- Lyden, Jacki. 2009. „Librarian Opposes Google's Library Fees,“ *National Public Radio* [online]. Dostupné z: <<http://www.npr.org/templates/story/story.php?storyId=100969810&ft=1&f=1003>>
- Perez, Juan Carlos. 2008. „In Google Book Settlement, Business Trumps Ideals.“ *PC World* [online]. Dostupné z: <http://www.pcworld.com/businesscenter/article/153085/in_google_book_settlement_business_trumps_ideals.html>
- Rohle, T. 2007. „Desperately seeking the consumer: Personalized search engines and the commercial exploitation of user data,“ *First Monday; Volume 12, Number 9 - 3 September 2007*.
- United States District Court, Southern District of New York. 2009. „Case No. 05 CV 8136-JES (The Authors Guild, Inc., Association of American Publishers, Inc., et al., v. Google Inc.) SETTLEMENT AGREEMENT“ [online]. Dostupné z: <<http://www.googlebooksettlement.com/intl/en/Settlement-Agreement.pdf>>
- University of Michigan Library. 2009. „Michigan Digitization Project“ [online]. Dostupné z: <<http://www.lib.umich.edu/michigan-digitization-project>>
- Vaidhyanathan, Siva. 2009. „The Googlization of Everything.“ [online]. Dostupné z: <<http://www.googlizationofeverything.com/>>
- Waller, V. 2009. „The relationship between public libraries and Google: Too much information,“ *First Monday; Volume 14, Number 9 - 7 September 2009*.
- Wiesmann, Gerrit a Richard Waters. 2009. „Germany seeks to block Google book deal,“ *Financial Times* [online]. Dostupné z <<http://www.ft.com/cms/s/0/fa90f2ce-972c-11de-83c5-00144feabdc0.html?SID=google>>
- Wikipedia. 2009. „Google Book Search“ [online]. Dostupné z <http://en.wikipedia.org/wiki/Google_book_search>

Metadata v DC

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Některé otázky kolem služby
Google Book Search" />
<meta name="DC.Creator" content="Karel Svačina" />
<meta name="DC.Subject" content="Google Book Search" />
<meta name="DC.Description" content="Google Book Search je služba
pro vyhledávání v obsahu knih" />
<meta name="DC.Date" content="09.12.2009" />
<meta name="DC.Type" content="Text" />
<meta name="DC.Type" content="Esej" />
<meta name="DC.Format" content="application/pdf" />
<meta name="DC.Format" content="computerFile" />
<meta name="DC.Source" content="http://books.google.com/" />
<meta name="DC.Language" content="cs" />
```