

Google Book Search

Fulltextové vyhledávání v knihách

Provozuje Google Inc. na adrese <http://books.google.com/> (v listopadu změněn název z původního Google Print).

Tomáš Obšívač, 13. 12. 2005

Obsah eseje

- Charakteristika Google Book Search
- Vývoj projektu
 - Kde vzít dost knih?
 - Z pohledu uživatele
 - Timeline
- Spor s autorským právem?
- Subjektivní zhodnocení projektu
 - "Veřejnoprávní alternativa"
- Zdroje

Charakteristika Google Book Search

Google, ve své snaze uspořádat světové informace a zpřístupnit je všem (a samozřejmě na tom neprodělat), se rozhodl po vyhledávání ve webových stránkách, dokumentech, obrázcích, diskuzních skupinách a dalších on-line dostupných zdrojích, konečně zpřístupnit všechny knihy světa.

„*We want to make all the world's books discoverable and searchable online.*“ – Google

Vývoj projektu

Digitalizace knih není idea posledního roku, ale zakladatelé Google ji mají na mysli už od doby, kdy o vlastní firmě vůbec neuvažovali. Ve škole totiž pracovali na digitalizaci knihovny...

Kde vzít dost knih?

Library Project je prvním zdrojem knih. Google uzavřel partnerství s pěti velkými knihovnami (při University of Michigan, Harvard University, Stanford University, Oxford University a New York Public Library). S každou zvlášť a za různých podmínek.

Ze Stanfordu a Michiganu dostal přístup k celým sbírkám (kolem 7 mil. knih každá). Z největší americké akademické knihovny v Harvardu si zajistil pouhých 40 tisíc náhodně vybraných knih (z 15 mil.) s příslibem další spolupráce, pokud nebudou problémy. Oxford dovolil naskenovat všechny knihy vydané do roku 1900 včetně. [2]

Michigan digitalizoval asi 5000 knih ročně a od Google si slibuje, že za 6 let bude mít k dispozici kompletní digitální kopii pro vlastní potřebu! Tento zázrak dovolí skenovací zařízení, která snímají stránky už během jejich otáčení.

Partner Program je pokus o dohodu s komerčními firmami a autory s podtitulem *online book marketing program*. Snaží se zaujmout držitele práv výhodami projektu a získat od nich povolení k zobrazování více knih (a více z jejich obsahu). Dává také přístup ke statistikám a dovoluje zakázat vyhledávání pro určitou publikaci (opt-out).

Nové knihy Google slíbil zařadit do indexu nejdříve rok po jejich vydání, aby dovolil vydavateli připojit je k

partnerskému programu sám (nebo naopak zakázat indexaci). Pokud by se měl Google spolehnout jen na tyto partnery, odhaduje, že by nezískal více než 15 % ze všech světových publikací.

Jediný údaj, který jsem našel k velikosti indexu, pochází z Wikipedie a udává stovky tisíc knih od vydavatelů a autorů a 10 000 z knihoven.

Z pohledu uživatele

Při vyhledání konkrétní knihy slouží je jako výsledek jeden ze tří režimů zobrazení:

1. Snippet View

– několik vět kolem hledaného výrazu a informace o knize chráněné autorským právem (metainformace, odkazy na recenze, prodejce)

The screenshot shows the Google Book Search interface for the book "The Role of GATT in Relation to Trade and Development" by The Contracting Parties to the General Agreement on Tariffs and Trade. The search term "treaty" is entered in the search box. The results show three references to "treaty" in the book, with snippets from pages 5, 54, and 55. Annotations on the left side of the screenshot explain various features: "Enter new keywords to conduct more searches within the book" points to the search box; "Total number of times this search term appears in this book. You can only view three instances for each term" points to the "3 references to treaty in this book" text; "The search term that led you to this book is highlighted" points to the word "treaty" in the snippets; "If a book is out of print, search for a used copy" points to the "Buy this Book" section with links to Abebooks, Alibris, and Froogle; "Find this book in a local library" points to the "Find this book in a library" link; "Find related information on the web" points to the "Related information" section with links to web searches for reviews and other web pages related to the book. The bibliographic information at the bottom lists the title, author(s), publisher, and page count.

Annotations on the left side of the screenshot:

- Enter new keywords to conduct more searches within the book
- Total number of times this search term appears in this book. You can only view three instances for each term
- The search term that led you to this book is highlighted
- If a book is out of print, search for a used copy
- Find this book in a local library
- Find related information on the web

Search results for "treaty" in "The Role of GATT in Relation to Trade and Development":

3 references to **treaty** in this book

Page 5

1. This introduction deals briefly with the nature and objectives of the General Agreement on Tariffs and Trade, the functions of the CONTRACTING PARTIES, and the membership of the Agreement.

2. GATT is a multilateral **treaty** embodying reciprocal rights and obli-

Page 54

examination by the GATT of the **treaty** of Rome establishing the European Economic Community and the association of the overseas States with the Community.

17. The Seventh Supplement to the Basic Instruments and Selected Docu-

Page 55

issues raised by the **Treaty** of Rome in its provisions relating to the association of the overseas territories. It also contains commodity reports on twelve primary products which would be affected by the EEC arrangements, namely: cocoa, coffee, tea, bananas, sugar, tobacco, oilseeds, cotton,

Where's the rest of this book?

Buy this Book

- Abebooks
- Alibris
- Froogle

Find this book in a local library

Find this book in a library

About Google Print

Related information

- Web search for reviews of *The Role of GATT in Relation to Trade and Development*
- Other web pages related to *The Role of GATT in Relation to Trade and Development*

Bibliographic information

Title The Role of GATT in Relation to Trade and Development

Author(s) The Contracting Parties to the General Agreement on Tariffs and Trade

Publisher Geneva

Pages 56

2. Sample Pages View

– omezený počet stran knihy, u které to držitel práv povolil (navíc dostupných po přihlášení ke Google Account)

Enter new keywords to conduct more searches within the book.

Learn more about the publisher and view other books they offer.

The search term that led you to the book is highlighted on the page.

Read book excerpts and view reference information about the book

"Buy this Book" links take you to online book stores where you can buy the book immediately.

Targeted, relevant ads provide you with links to related products and services.

Google Book Search BETA

close out the chapter with a few fi Search this book Search all books

Mastering Digital Photography: The Photographer's Guide to Professional-Quality
by David D. Busch Provided by Thomson Course Technology through the Google Print Publisher Program

Page 282

Back to Search results

Front Cover
Copyright
Table of Contents
Index
Back Cover

More results from this book
About this Book

MASTERING Digital Photography

THOMSON COURSE TECHNOLOGY

Buy this Book
Thomson Course Technology
Amazon.com
Barnes&Noble.com
BookSense.com
Froogle
About Google Print

282 Macro Photography

Some Final Tips

close out this chapter with a few final tips that might come in handy for specific types of close-up photography.

- Chill out. Some nature photographers looking to increase the patience of their insectoid subjects put the little creatures in an icebox for a few moments before posing them carefully back in more natural surroundings. Chilled butterflies, for example, will remain in one spot long enough for an interesting series of pictures. A brief visit to colder climes (a notable ice chest may work fine in the field) doesn't harm them in the least.
- Have a spritz bottle handy. Flowers, vegetables, fruit, and spiderwebs often look better when dusted with a light mist, as you can see in Figure 9.29. Don't over-dew (hehehe) the moisture and your photos can be enhanced.
- Make creative use of reflections. I spent a lot of time in this chapter telling you how to avoid reflections, but this is one rule that deserves to be shattered from time to time. Some interesting close-up photos have been produced when the photographer accidentally or intentionally included a reflection of something in the shiny surface of the subject being photographed. It might have been a photographic umbrella or even the photographer. When photographing spoons, chromium bumpers, or anything imprinted with an "Objects are closer than they appear" warning, see what you can do creatively with the reflection. Check out Figure 9.30 for an example.



Figure 9.29 A spritz of water can add interest to a fruit or vegetable picture.

ISBN: 1592001149

Page 282

See a problem with this page? Please tell us.

Close out Buy or sell closeout/surplus goods. Bulk lot auctions start at \$100!
www.Liquidation.com

Close Out Shop for Sports Equipment! Find, Compare and Buy
www.Shopping.com

Close Out Close Out Start your search here.
zimply.com

Copyrighted Material

Used with permission. Copyright © 2004. Thomson Course Technology

3. Full Book View

– dostupná celá kniha, protože copyright vypršel

Enter new keywords to conduct more searches within the book

View reference information about the book

Find more results from this keyword throughout this book

The search term that led you to this book is highlighted

If a book is out of print, search for a used copy

Find this book in a local library

Google Book Search

Books and Culture
by Hamilton Wright Mabie - Public domain book provided through the Google Print Library Project

Search this book Search all books

Back to Search results

Front Cover
Copyright
Table of Contents
Back Cover

More results from this book about this topic

Buy this Book
eBooks
eBooks
eBooks

Find this book in a library
About Google Print

By Way of Illustration.

evening when a play of Sophocles was to be rendered by the students of a certain university in which the tradition of culture has never wholly died out, and I led the talk along the lines of the play. I was rewarded by an hour of such delight as comes only from the best kind of talk, and I felt anew the peculiar charm and power of culture. For what I got that enriched me and prepared me for real comprehension of one of the greatest works of art in all literature was not information, but atmosphere. I saw rising about me the vanished life, which the dramatist knew so well that its secrets of conviction and temperament were all open to him; in architecture, poetry, religion, politics, and manners, it was quietly rebuilt for me in such wise that my own imagination was stirred to meet the

98

See a problem with this page? Please tell us.

Volně přístupné jsou obsahy a rejstříky. Systém přidává odkazy na knihovny, u prodávaných knih na internetové obchody, u vyprodaných na použitou literaturu. U chráněných knih se snaží zabránit tisku a kontroluje počet zobrazených stran.

Timeline

- 1998, Larry a Sergey opouští Stanford a zakládají "search engine company" :-)
- říjen 2004, Amazon představuje [Search Inside!™ \(the Book\)](#)
- prosinec 2004, oficiální spuštění projektu Google Print Beta
- květen 2005, AAUP veřejně pokládá 16 otázek, které ji na projektu znepokojují
- srpen 2005, Google zavádí opt-out a provize z reklamy pro vydavatele a pozastavuje skenování "chráněných" knih (do listopadu) [6]
- září 2005, žaloba AG proti Library Project [7]
- říjen 2005, žaloba pěti vydavatelů podporovaná AAP proti Library Project (McGraw-Hill v. Google) [5]
- říjen 2005, Google chce rozšířit službu na další evropské jazyky (Francie, Itálie, Německo, Nizozemí, Rakousko, Švýcarsko, Belgie a Španělsko mají mít vlastní verzi)
- 18. listopad 2005, přejmenování služby

Spor s autorským právem?

Ve Spojených státech se žaluje jako na běžícím pásu a oblast copyright plus Internet je úplný evergreen. Není tedy překvapením, že pár dolarů se rádobypostižení autoři a vydavatelé snaží vysoudit.

Copyright jasně říká: žádné kopie bez povolení od jeho držitele. Na této jednoduché tezi se snaží stavět obžaloba. "Zpaťte mi za kopie, i když je pořizujete jen jako prostředek k jiné činnosti, která je navíc v mém zájmu."

Google se brání odvoláváním na tzv. [Fair Use](#), což v tomto případě znamená, že výsledný produkt nikterak

nekonkuruje původnímu dílu (dokonce naopak, podporuje jej). Snáží se tedy přesvědčit soud, že "když se vám někdo vloupá do domu, aby uklidil v kuchyni, je to v pořádku". :-)

Žalobci se snaží přitížit Googlu mj. tím, že např. upozorňují na skutečnost, že nedělá kopie pouze pro sebe, ale také pro partnerské knihovny. Možnost vyjmutí knihy z programu je také právně irelevantní, copyright vyžaduje explicitní povolení kopírování (opt-in).

Precedenční právo zná případy, které jsou podobné. Např. vyhledavač obrázků Arriba Soft, který vytvářel z originálů náhledy (ano, vyhrál, dělají to dnes všichni). Nebo Sony v. Betamax, kde se kopie v přehrávači udržovala pro time-shifting (odložení sledování pořadu).

Závěr je poměrně jasný. Současná podoba copyrightu je nevyhovující. Vytváření kopií je de facto nutnost pro běžné používání digitálních dokumentů. Bohužel kopírovat znamená de jure nelegálně rozšiřovat.

Úsměvné obvinění

Londýnská dětská nemocnice Great Ormond drží od roku 1929 (v UK) copyright na Petra Pana a z provize za prodané knihy spolufinancuje léčebnou péči. Ústy svého mluvčího, který ještě stránky neviděl, doufá, že Google si dvakrát rozmyslí publikování knihy, protože by to znamenalo mnohamilionové ztráty. Google Print zraňuje děti! Peter Pan je volně dostupný přes [Project Gutenberg](#) ([The Adventures of Peter Pan](#)).

Subjektivní zhodnocení

Oblíbenost vyhledávání v knihách roste, už nyní je vyšší než další zavedené služby. [8] Toje podpořeno i tím, že odpovídající knihy nabízí před běžnými výsledky hledání na webu.

Nepochybují, že i méně úspěšné firmy, než je Google Inc., by dokázaly komerčně využít tak rozsáhlou digitální knihovnu. Na kontextové reklamě zatím větší díl vydělávají vydavatelé [9]. V budoucnu možná nezanedbatelná část z hledajících za přiměřený poplatek ráda využije možnost stáhnout si požadované stránky (např. do specializovaného zařízení podobného dnešním mp3 přehrávačům).

Myslím, že si Google jasně uvědomuje, že porušuje literu zákona, ale spoléhá na to, že není v rozporu s jeho duchem. Pokud zvítězí, může to být velký impuls k přepracování copyrightu. Naopak by mohlo dojít k přitvrzení, kdyby prohrál.

U nás se k méně ambicióznímu (ale komerčně jistě zajímavému) projektu chystá Seznam.cz. Můžeme se těšit na "české Sample Page View" knih, u kterých se podaří domluvit s vydavatelem. [4]

"Veřejnoprávní alternativa"

Tempo, s jakým se Google pustil do práce, převálcovalo podobné projekty veřejných institucí. Nepříjemné je využití proprietární technologie (byť je to pravděpodobně ta dnes nejlepší možná).

[Open Content Alliance](#) je sdružení, ve kterém je členem také Yahoo!, MSN, Internet Archive, několik kanadských univerzit, Adobe nebo O'Reilly Media. Cílem je všem přístupná digitalizace publikací s "vyřešeným copyrightem".

Zdroje

[1] [About Google Book Search](#)

[2] [Google adds major libraries to its database](#), ZDNet

[3] [Google, Publishers, Copies and "Being Evil"](#), IPTAblog, Andrew Raff, září 2005

[4] [Seznam objednává skenery](#), Dušan 'Yuhů' Janovský, listopad 2005

[5] [Publishers Sue Google Over Plans To Digitize Books](#)

[6] [Making books easier to find](#), Google Blog

[6] [Authors Guild Sues Google, Citing “Massive Copyright Infringement”](#)

[7] [Michigan Digitization Project](#)

[8] [Google Print Is Google's Ninth Most Popular Service](#), Search Engine Watch, listopad 2005

[9] [The point of Google Print](#), Google Blog

Pro ilustraci: http://books.google.com/books?as_auth=havel, pro názory ze světa: Technorati [Google Book Search](#).

Metadata DC viz zdrojový text.