

Project Gutenberg

<http://www.gutenberg.org>

Stanislav Kunc, student Bc. Ap., 7. semestr

Zpracováno dne 5. ledna 2005

Stručná charakteristika projektu

Project Gutenberg (PG) byl založen v roce 1971. Jedná se o dobrovolné hnutí, které je oficiálně zastřešeno neziskovou organizací The Project Gutenberg Literary Archive Foundation 501(c)(3) a je financováno z darů. Cílem je digitalizovat, archivovat a distribuovat knihy, časopisy, noviny, hudební záznamy, fotografie a jiná kulturní díla. Předmětem zájmu jsou hlavně díla, která spadají do public domain.

Historie projektu

V roce 1971 napadlo pana Michaela Harta (což je zakladatel, ideový vůdce a současný Executive Coordinator PG), že počítače představují ideální nástroj pro zpřístupnění literárních děl. Využil tedy darovaného počítačového času u mainfraimu a přepsal americkou Deklaraci nezávislosti do textového souboru, čímž vznikl první příspěvek do PG (viz [PGHIST]).

Vzhledem malé publicitě projektu a k tomu, že před třiceti lety byla počítačová infrastruktura daleko méně rozvinutá a rozšířená, další literární díla přibývala velice pomalu. Konkrétně v 70. letech byly přidány některé klasické americké texty. Následně se v 80. letech pracovalo na Bibli a sesbíraných pracích Shakespeara, které však kvůli změně copyrightu nebyli v rámci PG vydány.

Dnešní podobu získal PG v roce 1991, kdy se rozšířila sbírka i oblast zájmu, nicméně opravdový posun nastal až po rozšíření počítačů a internetu. Internet přilákal nové dobrovolníky a umožnil efektivněji využívat prostředky.

Cíle projektu

PG se snaží plnit některé funkce digitální knihovny. Především jde o zachování kulturního dědictví, tj. shromáždit a zpracovat do vhodné formy co nejvíce prací, aby byly přístupné dalším generacím. Dále se PG snaží o co nejmenší finanční náklady při následném přístupu k uloženým záznamům a distribuci objektů z archivu (viz [PGLAF]).

Přesnou formulaci prvotních cílů lze najít např. ve [FAQ0]:

- We want to provide as many eBooks in as many formats as possible for the entire world to read in as many languages as possible.
- The mission of Project Gutenberg is simple: "To encourage the creation and distribution of eBooks."

Michael Hart na počátku PG prohlásil, že cílem je do roku 2000 shromáždit 10 000 nejpoužívanějších knih. Tento cíl se zdál být nerealizovatelný, neboť zpočátku knihy přibývaly velice pomalu (např. shromáždit prvních 750 knih se podařilo až v roce 1996, viz [PGNEWS96]). Přesto díky exponenciálnímu nárůstu příspěvků bylo 10 000 knih shromážděno v roce 2003, což vzhledem k omezeným zdrojům není příliš velké zpoždění.

Po dosažení 10 000 příspěvků navrhl Hart novou ambiciózní metu: zvětšení archivu na 1 milion (volně a bezplatně distribuovatelných) knih, které budou dodány 1 miliardě lidí a to do konce roku 2015 (viz [PROGRESS]).

Pro opravdu snadný přístup k uchovaným duševním dílům nestačí pouze dílo samotné. Důležité je také, v jakém jazyce se dílo nabízí. Je jistě vhodné, aby byly např. knihy dostupné v co největším počtu jazyků. Překlad děl vidí pan Michael Hart jako další činnost PG. Nepochybně se jedná o úkol mnohem náročnější než samotné zpracování díla do vhodného formátu. Zda se to podaří, v jaké míře a kvalitě ukáže až čas.

Způsob financování

Nezisková organizace The Project Gutenberg Literary Archive Foundation (viz [PGLAF]) je jediným zdrojem příjmů PG. Projekt není přímo podporován granty, vládní iniciativou či jiným oficiálním způsobem. Tato skutečnost samozřejmě poněkud omezuje možnosti a rozsah PG.

Na druhou stranu to znamená nezávislost na dotujících subjektech. PG není ohrožen končícím grantem, politickou či fiskální krizí a zodpovídá se de facto těm, kteří PG dotují a často zároveň realizují. Převážná část příjmů pochází od drobných dárců a velké množství dobrovolníků představuje základní pilíř PG.

Přidání díla do PG

Do PG si nemůže vložit kdokoli cokoli. Ačkoli PG z licenčního hlediska postačí souhlas autora nebo ověření, že dílo už patří do public domain, neznamená to, že PG automaticky přijme zaslané dílo. PG dává přednost knihám klasicky vydaným a respektovaným (umělecká či informační hodnota je uznávána nejen autorem). To je nutné vzhledem k omezeným prostředkům a pro možnost kontroly autorských práv. Podrobněji viz [SUBMIT].

PG nijak necenzuruje konkrétní díla, která splňují nutné podmínky pro zahrnutí do archive. Neexistuje oficiální seznam knih, které chce PG zpracovat. Vše záleží na dobrovolnících a jejich preferencích.

Kontrola korektnosti obsahu

PG striktně trvá na editovatelných formátech souborů, aby se mohly opravovat případné chyby objevené po vydání díla. Typicky se jedná o překlepy či špatně rozpoznané znaky při skenování. Samotný proces přidání díla nemůže zajistit úplnou korektnost zpracovaného díla, ale existují nástroje a projekty, které ulehčují a zlepšují kontrolu.

Souborové formáty

Od počátku je v PG pravidlem, že základem nového příspěvku je čistě textová verze dokumentu (samozřejmě tam, kde to má smysl, hudební nahrávka zřejmě v ASCII nezní hezky). Dále se preferují otevřené a standardní formáty oproti proprietárním formátům. Více viz [FAQF1].

PG vychází z jednoduché úvahy, že formáty se časem mění a zanikají, zatímco prostý text je zde od počátku historie počítačů a dlouhodobě zůstane zachován. Důsledkem této politiky je převaha knih zastoupených prostým textovým formátem (viz [FAQG17])

Lze očekávat, že ve střednědobém horizontu se budou knihy ukládat v nějakém XML formátu a z něj se budou generovat čistě textové a jiné verze.

Čeho bylo dosaženo

Nejvíce jsou v PG zastoupeny knihy. Ke dnešnímu datu se jich podařilo archivovat zhruba 14 000. PG je jazykově nezávislý, tudíž si čtenář může vybírat z 36 různých jazyků. Z historických důvodů je však dominantně zastoupena angličtina.

PG obsahuje řádově stovky zvukových záznamů knih, které čte počítač. Další generované záznamy už se nebudou přidávat, spíše se nahradí vysláním na požádání. Je nutné uvést, že tyto záznamy jsou určeny hlavně nevidomým, protože ti často nemají jinou možnost, jak se seznámit s dílem.

Desítky knih čtené lidmi. Většinou jde o záznamy, které jsou výsledkem projektu The Sound of Literary Works [SOUNDWORK]. Další zvukové záznamy, např. z oblasti vážné hudby, projevy apod.

Praktickým produktem jsou CD a DVD projekty z roku 2003. Na [CD/DVD] je k dispozici "Best of Gutenberg" CD obsahující 600 digitálních knih a DVD obsahující zhruba 9 400 digitálních knih.

Především však vzniklo jakési informační podhoubí, které umožnilo vznik dalších návazných projektů jako jsou specifičtější zaměřené organizace Projekt Gutenberg-DE [GUTDE], Project Gutenberg Europe [GUTEU] a Project Gutenberg of Australia [GUTAU], projekty přidávající hodnotu k archivovaným textům např. už zmíněné The Sound of Literary Works nebo TellTale Weekly [TELLTALE] atd.

Důležitým dodavatelem obsahu se stal Project Gutenberg's Distributed Proofreaders [PGDP], který v současné době zajišťuje zhruba polovinu všech nových knih a zřejmě se v budoucnu stane dominantním dodavatelem knih pro PG.

Vlastní zhodnocení projektu

PG je jeden z mnoha pokusů o zachování naší kultury. Specifický je tím, že se zaměřuje (převážně) na knihy a řídí se americkou legislativou. Unikátní je v tom, že jde o čistě dobrovolnou aktivitu a byl první v oblasti literatury volně dostupné na internetu.

Nejedná se o digitální knihovnu v pravém slova smyslu, ale spíše o velikou sbírku digitálně zpracovaných materiálů, což se projevuje v tom, že uživatel nemá k dispozici komfort na který je zvyklý v klasické knihovně. Např. identifikátory definované v rámci PG platí pouze v PG, neexistuje ekvivalent ISBN. Metadata jsou minimální a nemají standardní strukturu, navíc některá metada PG záměrně vypouští, např. vydavatele a místo vydání. PG samozřejmě nepokrývá všechny oblasti kultury, ale zároveň není propojen s jinými projekty, takže nemohu efektivně získávat souvislosti apod.

V ideálním světě bychom měli vyřešeny všechny problémy, které nám znemožňují efektivně uchovávat, sdílet a rozvíjet naše díla a poznatky. My však nežijeme v ideálním světě, a tudíž máme spoustu nedořešených otázek jako jsou autorská práva, jak získat a dlouhodobě uchovat díla, kdo to zaplatí apod. Tyto nedořešené záležitosti tvoří teoretické limity pro PG.

I přes problémy, které jsem zmínil v předchozích dvou odstavcích považuji Project Gutenberg za úspěšný a prospěšný a doufám, že v následujících letech, desetiletích a staletích se bude dál rozvíjet a zlepšovat.

Použité informační zdroje

- [FAQ0] Project Gutenberg Mission Statement, dostupné na URL:
<http://www.gutenberg.org/about/faq0>
- [PGLAF] The Project Gutenberg Literary Archive Foundation,
dostupné na URL:
<http://www.gutenberg.org/fundraising/pglaf>
- [PGHIST] History and Philosophy of Project Gutenberg, dostupné na
URL: <http://www.gutenberg.org/about/history>
- [PROGRESS] Project Gutenberg Progresses, autorka Paula Hane,
dostupné na URL:
<http://www.infoday.com/it/may04/hane1.shtml>
- [SUBMIT] Submitting your own eBook or other Work to Project
Gutenberg HOWTO, dostupné na URL:
<http://www.gutenberg.org/howto/scopy-howto.php>
- [PGNEWS96] PG Newsletter December 1996, dostupné na URL:
<http://promo.net/pg/nl/9612-750.html>
- [SOUNDWORK] The Sound of Literary Works, dostupné na URL:
<http://literalsystems.org/>
- [GUTDE] Projekt Gutenberg-DE, dostupné na URL:
<http://gutenberg.spiegel.de/>
- [GUTEU] Project Gutenberg Europe, dostupné na URL:
<http://www.gutenberg.nl/>
- [GUTAU] Project Gutenberg of Australia, dostupné na URL:
<http://gutenberg.net.au/>
- [TELLTALE] TellTale Weekly, dostupné na URL:
<http://www.telltaleweekly.org/>
- [PGDP] Project Gutenberg's Distributed Proofreaders, dostupné na
URL: <http://www.pgdp.net/>
- [FAQF1] The Project Gutenberg FAQ - F-1, dostupné na URL:
<http://www.gutenberg.org/faq/F-1.php>
- [FAQG17] The Project Gutenberg FAQ - G-17, dostupné na URL:
<http://www.gutenberg.org/faq/G-17.php>
- [CD/DVD] The CD and DVD Project, dostupné na URL:
<http://www.gutenberg.org/cdproject/>

Metadata v Dublin Core

Název	Schéma	Hodnota
DC.Title	-	Project Gutenberg
DC.Creator	-	Stanislav Kunc
DC.Publisher.address	-	standa.kunc@mail.muni.cz
DC.Date	ISO8601	2005-01-04
DC.Type	-	Text
DC.Language	ISO639-1	cs
DC.Description	-	Esej do předmětu Digitální knihovny na téma Project Gutenberg
DC.Subject.keyword	-	Project Gutenberg, esej, digitální knihovna