

PERSEUS - the digital library of Greek, Roman, and Renaissance culture

(<http://www.perseus.tufts.edu>)

Esej v rámci předmětu PV070 – Digitální knihovny

Stručná charakteristika projektu

Projekt Perseus univerzity Tufts se týká jedné z největších digitálních knihoven shromažďující textové i obrazové materiály především z dob antického Řecka a Říma, anglické dokumenty z období renesance a 19. století, ale také spoustu sbírek germánských a arabských, noviny Richmond Times Dispatch z několika let nebo dokumenty sbírky The Duke Databank of Documentary Papyri, což je elektronický korpus řeckých a latinských textů nalezených na papýrech, ostrakonech a dřevěných tabulkách. Tento projekt je neziskový, podporován různými organizacemi a granty.

Doba řešení, aktuální stav

Plánování projektu započalo už v roce 1985, avšak skutečný vývoj až v roce 1987 s iniciální snahou sbírat a prezentovat ke studijním účelům díla starého Řecka. Postupem času se projekt začal více rozrůstat a s příchodem webu měl možnost účastnit se experimentů online prezentace (konkrétně v roce 1995 vznikly sbírky na www). Začaly se objevovat nové formy publikace a digitalizace se ukázala jako velmi významný krok ve vývoji. V dnešní době se může zájemce přes webové stránky projektu dostat k velké databázi článků v různých jazycích, která čítá přes 100 mil. slov. Tato digitální knihovna obsahuje ale i desítky tisíc obrázků. Kromě toho byl projekt publikován i na několika CD.

Cíle projektu

V počátcích byl zájem výzkumu projektu Perseus v souhře mezi technologií a intelektuálním bohatstvím. Nejdříve se vývojáři zaměřili na starověké Řecko, kde dokázali vytvořit velký objem informací sestávající z původních textů, moderních překladů, slovníků, komentářů, obrázků, ... Tyto materiály byly publikovány na několika CD nosičích. Za podpory se později mohly zpracovávat i zdroje z dalších kultur a publikovat na web. V akademickém roce 2008/2009 mají vyvojáři v úmyslu se soustředit na tři nejrozšířenější jazyky oblasti středozemí: řečtinu, latinu a arabštinu. Jeden z důvodů je existence některých původně řeckých děl jen v arabštině.

Obecné cíle (nebo také poslání) projektu jsou stanoveny jako vytvoření úplného záznamu lidstva – lingvistických zdrojů, hmotných artefaktů, historických míst (také v podobě virtuálního světa) – intelektuálně dostupných jak je to jen možné co největšímu publiku i přes různá jazyková či kulturní zázemí.

Popis projektu a jeho výsledků

Momentálně je Perseus tvořen sbírkou XML dokumentů (dříve SGML), neboť díky takovému strukturovanému značkovacímu jazyku tvůrci získali jistou obecnost a abstrakci, která dovoluje dodávat obsah bez problémů na různorodé platformy.

Co se týče obecných autorských práv vztahujících se k těmto dokumentům, pro osobní použití studentů, vědců a veřejnosti jsou materiály volně dostupné. Pro komerční účely jsou bez svolení standardně zakázány. Konkrétní materiály však mohou mít autorská práva stanovena jinak. Licence jsou obvykle nastaveny pomocí Creative Commons.

Běžné on-line procházení sbírek digitální knihovny Perseus je možné přes webové stránky <http://www.perseus.tufts.edu>, které jsou označeny jako Perseus 4.0 nebo také Hopper. Mimo to lze k dokumentům přistupovat i přes mirrory berlínského Max Planck Society anebo University of Chicago. Samotné prohlížení sbírek lze intuitivně provést zvolením konkrétní sbírky (jako je například Greek and Roman collection) a následným výběrem jména autora, či titulku dokumentu

nebo jednoduše vyhledáváním klíčových slov. Pro textové dokumenty, které jsou součástí webové stránky, je přehledně naznačena struktura rozdělením na jednotlivé stránky, kapitoly, verše nebo jiné celky.

Jelikož je hojně využívána technologie XML, nechybí samozřejmě možnost zobrazení aktuálně načteného textu ve formátu XML.

Pro práci se samotným textem je poskytováno hned několik nástrojů, které však zřejmě nejsou poskytovány pro všechny dokumenty stejné. Jedním z nich jsou možnosti zobrazení textu, v nichž lze například zvolit kódování zobrazující řečtinu, arabštinu. U arabsky psaných textů je možné použít i slovníček některých pojmů s vypočtenou frekvencí výskytů. Kliknutím na slova z textů arabských, řeckých, latinských se zobrazí okno s informacema jako je třeba překlad do angličtiny, morfologický popis nebo struktura, frekvence výskytů v daném korpusu. U anglických textů se jinak mohou do tabulek automaticky vypsát názvy míst, jména osob a datové údaje jakožto odkazy na výskyty těchto výrazů. Případně nám může posloužit vyhledávání námi zadaného takového údaje. Vyhledávat samozřejmě můžeme i běžná slova nebo i podslova ve vybraném jazyce. K dispozici je angličtina, řečtina, latina, stará angličtina, stará norština a němčina. Zajímavostí je určitě zadání anglických slov, která jsou automaticky přeložena buď do řečtiny, latiny, arabštiny nebo staré norštiny a následně v takovém tvaru vyhledávána v dokumentech.

Nástrojů pro vyhledávání je ale k dispozici více, především ze starších stránek této digitální knihovny. Seznam těchto nástrojů se dá nalézt na stránce <http://www.perseus.tufts.edu/cgi-bin/perscoll>. Zdá se však, že Atlas tool, London Atlas a Virtual Reality Interface neposkytují (v době vypracovávání tohoto eseje) plnou funkčnost. Z ostatních nástrojů mohou být vybrány:

- Art & Archaeology Browser – procházení databáze objektů typu architektura, mince, klenoty, sochy, místa, vázy. Součástí jsou většinou obrázky či fotografie, krátký popis a údaje jako je místo výskytu, doba výskytu aj.
- Colleccion Viewer – klasické procházení dokumentů podle zařazení.
- Dictionary Entry Lookup – vyhledávání řeckého, latinského, italského nebo anglického slova nebo podslova. Vyhledává se výskyt slova v některém slovníku. Lze si nechat zobrazit samotnou definici ze slovníku, další podobné výrazy, či statistiky výskytů v textech.
- English Index – vyhledávání slov a frází v textech nejen anglických, ale také francouzských, německých, řeckých, italských nebo latinských ze zadané kategorie sbírky dokumentů.
- English to Greek Word Search – vyhledávání řeckých slov v textech a jejich frekvencí výskytu, případně odkazem do slovníku. Zadává se však anglický ekvivalent, zmíněno již výše.
- Greek Morphological Analysis – jak již název napovídá, zobrazí se analýza výrazu a samozřejmě i frekvence výskytů.
- Greek Vocabulary Tool – Po vybrání některého řeckého dokumentu a zobrazovacích parametrů se vypíše tabulka slovíček ze zvoleného dokumentu. Parametry zobrazení jsou například: setřídění abecedně nebo podle frekvence výskytu, počet zobrazených slov, zobrazení dat v tabulce nebo v řádcích oddělených čárkou, zda poskytnout odkaz na morfologickou analýzu a které všechny údaje u jednotlivých výrazů vlastně zobrazit.

Pozn: poslední tři jsou možné i pro latinu. Greek (resp. Latin) Words in Context a Lookup Tool jsou vlastně to samé co English Index.

Verze Perseus 4.0 je od roku 2007 k dispozici pro uživatele také jako open-source aplikace. Lze si stáhnout Java zdrojový kód, který je nutné před spuštěním zkompileovat. Texty a data obecně je pak nutné stáhnout zvlášť. Ty jsou licencovány pod Creative Commons jako Attribution-Noncommercial-Share Alike.

Vlastní zhodnocení projektu a jeho přínosu

Perseus Digital Library je obsáhlá sbírka vhodná pro všechny, kteří se aspoň trochu zajímají o dějiny a umění. Sice neumím řečtinu a podobné jazyky, ale aspoň jsem si prohlédl anglické texty, konkrétně jak vypadala stará angličtina, kterou používal třeba Shakespear. Velkou výhodou je jistě volná dostupnost. Bohužel jsem se ale často setkával se zatížením serveru a stránky se načítaly velmi pomalu, především ty s textovými sbírkami. Na druhou stranu je k dispozici off-line řešení v podobě stažitelné open-source aplikace. Dále pak jsem nenašel na poslední verzi webových stránek všechny nástroje jako na té starší a taky se zdá že mirrory této digitální knihovny mají nejspíš taky nějaké hardwarové potíže.

Seznam zdrojů

<http://www.perseus.tufts.edu>
<http://www.ariadne.ac.uk/issue25/rydberg-cox/>
http://en.wikipedia.org/wiki/Perseus_Project
<http://www.cultivate-int.org/issue2/perseus/>
<http://www.perseus.tufts.edu/cgi-bin/perscoll>
<http://scriptorium.lib.duke.edu/papyrus/texts/DDBDP.html>

Metadata v Dublin Core

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Perseus" />
<meta name="DC.Creator" content="Karel Vaculík" />
<meta name="DC.Subject" content="Perseus Digital Library" />
<meta name="DC.Subject" content="the digital library of Greek, Roman, and Renaissance
culture" />
<meta name="DC.Description" content="digitální knihovna shromažďující historické textové a
obrazové dokumenty světa" />
<meta name="DC.Date" content="2008-11-30" />
<meta name="DC.Type" content="essay" />
<meta name="DC.Format" content="application/pdf" />
<meta name="DC.Format" content="computerFile" />
<meta name="DC.Source" content="http://www.perseus.tufts.edu" />
<meta name="DC.Source" content="http://www.ariadne.ac.uk/issue25/rydberg-cox/" />
<meta name="DC.Source" content="http://en.wikipedia.org/wiki/Perseus_Project" />
<meta name="DC.Source" content="http://www.cultivate-int.org/issue2/perseus/" />
<meta name="DC.Source" content="http://www.perseus.tufts.edu/cgi-bin/perscoll" />
<meta name="DC.Source" content="http://scriptorium.lib.duke.edu/papyrus/texts/DDBDP.html" />
<meta name="DC.Language" content="cze" />
```