

MASARYKOVA UNIVERZITA

FAKULTA INFORMATIKY



World Digital Library

PV070

Filip Sollár
456187

26.11.2017

O Projekte	3
Vznik	3
Ciele projektu	4
Obsah WDL	5
Hlavné výhody WDL	5
Implementácia	6
Verejné API	6
Priebeh pridávania obsahu do WDL	9
Zhodnotenie projektu	10
Zdroje	11
Metadata v Dublin Core	11

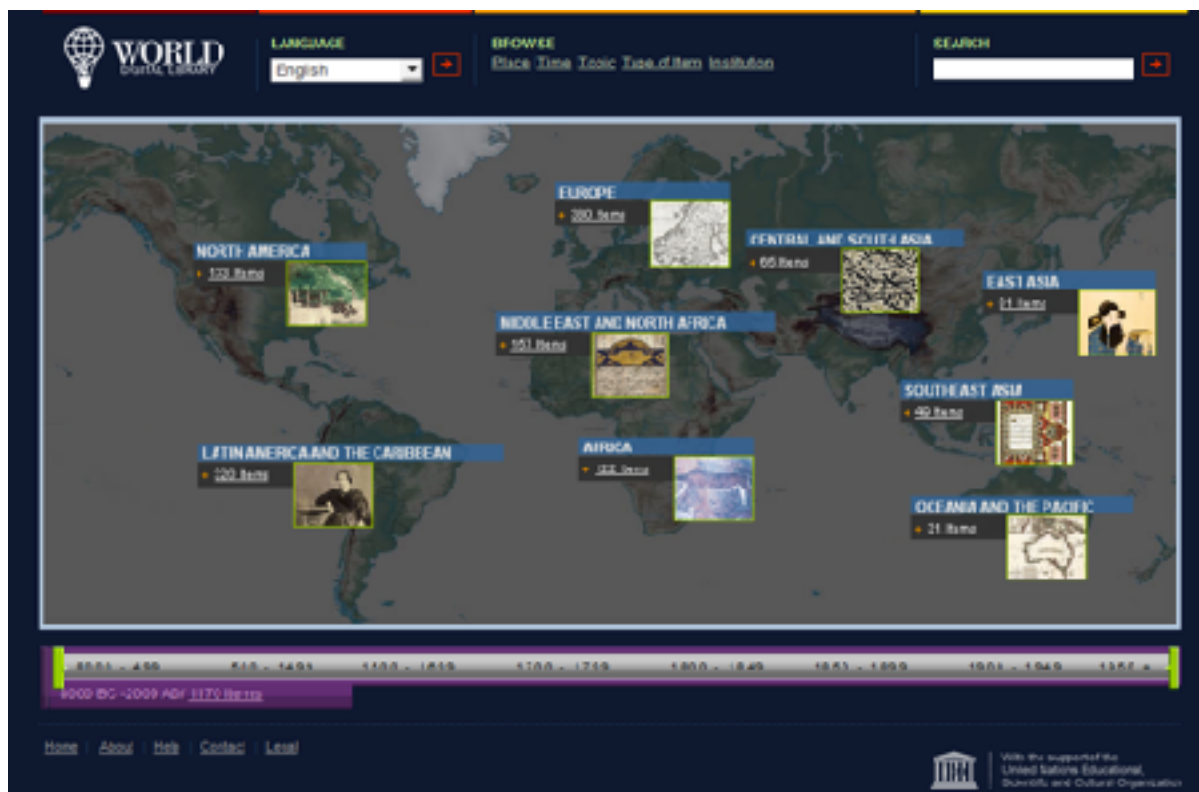
O Projekte

World Digital Library je nekomerčná digitálna knižnica, vytvorená Library of Congress, za pomoci svetovej organizácie UNESCO. Dostupná na adrese www.wdl.org umožňuje užívateľom z celého sveta pristupovať zadarmo k materiálom z rôznych kultúr, hlavne knihám, mapám a iným vzácnym dokumentom vo viacerých svetových jazykoch. Jej obsah pochádza z viac ako 200 krajín, z toho najstaršie diela sa datujú až do roku 8000 p. n. l.

Vznik

V júni 2005 James H. Billington (LoC) vo svojom prejave na Georgetown University predstavil svoj nápad na vytvorenie digitálnej knižnice, ktorá bude hlavným zdrojom informácií o úspechoch jednotlivých krajín. Pán Billington bol knihovníkom v LoC už od roku 1989 a s príchodom internetu začal myslieť dopredu. Bol jedným z hlavných podporovateľ projektov American Memory a National Digital Library Project. Jeho motiváciou k vybudovaniu WDL bolo, že si myslel, že takýto projekt zbližuje národnosti a zmenší kultúrne rozdiely. Preto požiadal o pomoc partnerské univerzity po celom svete a UNESCO, ktoré by umožnilo pomocou dotácii otvoriť tento projekt pre každého.

V decembri 2005 dostala LoC prvú finančnú dotáciu od Google-u, čo vzbudilo obrovský záujem aj iných investorov. Viacero národných knižníc ponúklo spoluprácu, ale takisto to zdvihlo vlnu skepticizmu. V Európe už začal vývoj Európskej digitálnej knižnice, krajiny tretieho sveta zas videli problém v Americkom privlastňovaní si kultúrneho dedičstva a nepáčila sa im ani počiatočná investícia Google-u, za čím hľadali snahu o vlastné obohatenie. V LoC bol poverený tím, ktorý by ďalej menil Billingtonovu víziu na realitu. Pridalo sa aj päť ďalších partnerských inštitúcií, Alexandrijská knižnica, Národná knižnica v Brazílii, Národná knižnica v Rusku, Ruská štátna knižnica a Egyptská národná knižnica. Na konferencii UNESCO v roku 2007 bol odprezentovaný prototyp, ktorý bol úspešný a viacero ďalších knižníc začalo kontribuovať svoj obsah. V apríli 2009 začala byť WDL verejne prístupná a mala vtedy okolo 60 partnerských inštitúcií. Takto vyzerala vo svojej prvej verzii.



Po spustení WDL pokračuje v pridávaní obsahu a získavaní nových partnerských inštitúcií z celého sveta. V roku 2010 bola vytvorená zmluva, týkajúca sa všetkých partnerských inštitúcií, ktorá zahŕňa aj každoročné stretnutie partnerov, volenie projektového manažéra a výkonnej rady. Projektový manažér sa volí kým ho nenahradia a jeho hlavnou úlohou je získavať financie na ďalšie rozvíjanie a údržbu projektu. Úzko spolupracuje s výkonnou radou a informuje ich o vývoji projektu. Výkonná rada je volená na dobu piatich rokov, dvaja z členov sú ex officio (zastávajú iný úrad) a tí sú v rade, až kým ich nenahradia. Posledné voľby prebehli v roku 2015 na sneme v Alexandrii a kompletný zoznam členov rady je na <https://project.wdl.org/about/org.html>. Posledné stretnutie bolo v máji 2016 v Katare a plánovala sa tam expanzia WDL do arabských krajín. V posledných rokoch na väčšine stretnutí sa diskutovalo hlavne ako zachovať kultúrne dedičstvo v Arabských krajinách.

Ciele projektu

World digital library poskytuje svetu zadarmo a vo viacerých jazykoch dôležité materiály z krajín celého sveta. Hlavné princípy WDL sú propagovať a podporovať medzinárodne a medzikultúrne porozumenie, rozširovať obsah a jeho pestrosť na Internete, poskytovať zdroje pre učiteľov a študentov a zjednotiť partnerské inštitúcie. WDL umožňuje ľuďom objaviť a študovať kultúrne klenoty z celého sveta.

Obsah WDL

V dobe písania tejto eseje bolo vo WDL presne 17 219 položiek, ktoré hovorili o 193 krajinách, medzi rokmi 8000 p. n. l. a 2000 n. l. Je to napísané hneď na úvodnej stránke pri vyhľadávaní. Ďalej sa dajú položky triediť podľa krajiny, časovej oblasti, témy, inštitúcie, typu predmetu a jazyku. Momentálne existujú materiály WDL v 143 jazykoch. Spolu majú všetky súbory obsiahnute vo WDL skoro 28 terabajtov. Podrobnejšia štatistika s presným počtom rôznych typov súborov je na <https://www.wdl.org/en/statistics/>. WDL obsahuje veľa unikátnych dokumentov, mňa osobne zaujala prvá práca na Slovensku v Slovenčine, ktorú dodala Slovenská národná knižnica. Veľmi príjemne ma prekvapilo, že aj SNK je zapojená ako partner tohto projektu. Pokúšal som sa aj niečo z tej knihy prečítať, bohužiaľ tá slovenčina bola veľmi odlišná od teraz používanej a viac sa podobala na poľštinu a češtinu tak som sa toho veľa nedozvedel. Nachádza sa tam plno prvotných textov alebo máp z rôznych krajín, kompletný list je na http://project.wdl.org/publications/WDL_Firsts.pdf. Knižovňa obsahuje 43 dokumentov zameraných na Slovensko a 34 na Česko. Väčšinu z dokumentov zo Slovenska tvoria portréty osobností, ktoré boli súčasťou formovania Slovenska ako napríklad Ľudovít Štúr alebo Milan Rastislav Štefánik. Ale nájdeme tam aj rôzne staršie maľby Bratislavy (vtedajšieho Pressburgu). Veľa z materiálov je zameraných na boj o nezávislosť Slovenska. Pri prezeraní som si všimol aj viaceré básne a sonety, o ktorých sme sa učili na literatúre. Najpozoruhodnejšou vecou čo som objavil, bola vkladná knižka Martina Kukučina alebo jeho poznámky, ktoré písal pomocou svojich vlastných vytvorených skratiek, takže celý text sa podobá na mimozemské písmo. Doteraz ich nikto nebol schopný úplne prepísať.

Hlavné výhody WDL

World Digital Library je poprednou digitálnou knižnicou, pretože jej obsah je popísaný veľmi konzistentnými metadátami. Ako jediná obsahuje preklady v 7 rôznych jazykoch, čo síce komplikuje jej údržbu, ale za to je omnoho viac univerzálnejšia ako iné digitálne knižnice.

Jej implementačný tím je veľmi schopný, snažia sa integrovať nové technológie čo najrýchlejšie, za neustáleho vyvoja ďalších nových v oblasti digitálnych knižníc.

Implementácia

Stránka je napísaná v jazyku Python pomocou frameworku Django a templateovacieho engine-u Jinja2, Django je rozsiahly framework, jeho použitie má viacero výhod ale aj nevýhod, je veľmi robustný a poskytuje rozsiahlu predrobenú funkcionálnu to zrýchľuje vývoj ale na úkor rýchlosti celého systému. Takisto je nevýhodou, že Django núti vývojára používať jeho vstavaný ORM. Stránka WDL používa viaceré JavaScript-ové knižnice ako JQuery, na jednoduchšie narábanie s DOM alebo Modernizr, ktorý dáva vývojárové informácie o tom, akú js funkcionálnu prehliadač poskytuje. Využíva Open Source nástroje ako Apache Solr na vyhľadávanie full-textového obsahu. Na hostovanie a CDN používa stránka platené služby Amazon Web Services. Na Text to speech je použitý Read Speaker. Stránka spĺňa štandard HTML 5 a používa framework Bootstrap, ktorý preddefinováva užitočné CSS triedy na jednoduchú implementáciu responzívneho webu stránok.

Verejné API

World Digital Library ponúka pre svojich užívateľov verejne prístupné API, na ktoré netreba ani vygenerovať kľúč. Vďaka tomuto prehľadnému API môžu vývojári používať zdroje z WDL, rád by som ukázal niektoré zo základných API volaní, celú dokumentáciu je možné si prečítať na <http://api.wdl.org/>. Základná URL API je <https://www.wdl.org> a keďže najväčšou výhodou WDL je jej multilingválnosť tak ako URL parameter sa pripojí kód krajiny v ktorej daný obsah požadujeme. Momentálne podporuje 7 jazykov a to

- ar - Arabčinu
- en - Angličtinu
- es - Španielčinu
- fr - Francúzštinu
- pt - Portugalčinu
- ru - Ruštinu
- zh - Čínštinu

Posledné pridané dokumenty môžeme získať buď vo formáte RSS alebo Atom. RSS a Atom sú oba dialekty XML, zamerané na získavanie nového obsahu, s tým rozdielom, že Atom je

novší, robustnejší, núti používateľov vyplniť viacero informácií o danom obsahu atď. Preto je jeho jednoduchší ekvivalent RSS častejší.

Príklad volania:

<https://www.wdl.org/en/feeds/latest.rss>

alebo

<https://www.wdl.org/en/feeds/latest.atom>

vráti informácie o správovom kanály a aktualizované položky,

Príklad zjednodušeného výstupu:

```
<channel>
  <title>World Digital Library</title>
  <link>https://www.wdl.org/en/</link>
  <description>New and updated content from the World Digital Library</description>
  <language>en</language>
  <lastBuildDate>Thu, 16 Nov 2017 17:08:53 +0000</lastBuildDate>
  <item>
    <title>Killed Seagull</title>
    <link>https://www.wdl.org/en/item/21375/?
      utm_source=feed&utm_medium=rss&utm_campaign=feed</link>
    <description>This photograph is from a collection ...</description>
    <pubDate>Thu, 16 Nov 2017 17:08:53 +0000</pubDate>
    <guid>https://www.wdl.org/en/item/21375/?
      utm_source=feed&utm_medium=rss&utm_campaign=feed</guid>
  </item>
  ... more items
</channel>
```

ako vidíme vo výstupe sa používajú obvyklé RSS tagy.

Takisto WDL používa aj Sitemapy, čo je XML súbor ktorý informuje internetové vyhľadávače o všetkých URLs webstránky s metadátami, ktoré popisujú kedy sa daná stránka naposledy aktualizovala, ako je daná URL dôležitá a s akou frekvenciou sa aktualizuje. Nie vždy však crawlers podporujú sitemapy, takže využitie Sitemapu na stránke negarantuje obsiahnutie URL z mojej stránky vo vyhľadávaní.

URL sitemapy WDL:

<https://www.wdl.org/en/sitemap.xml>

Príklad sitemapy:

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>https://www.wdl.org/en/item/1/</loc>
    <lastmod>2017-04-04</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.5</priority>
  </url>
  ... more URLs
</urlset>
```

Na vyhľadávanie vo WDL sa používa OpenSearch, čo je zbierka formátov na zdieľanie výsledkov hľadania pre prehliadače. Klient načíta konfiguráciu OpenSearchu na stránke a podľa toho zistí ako na stránke vyhľadávať.

OpenSearch description document sa nachádza na

<https://www.wdl.org/en/opensearch.xml>

a vyzerá nasledovne:

```
<?xml version="1.0" encoding="UTF-8"?>
<OpenSearchDescription xmlns="http://a9.com/-/spec/opensearch/1.1/" xmlns:moz="http://
www.mozilla.org/2006/browser/search/">
  <ShortName>WDL</ShortName>
  <Description>Search the World Digital Library</Description>
  <Image width="16" height="16" type="image/x-icon">/static/img/favicon.8b766365aaa1.ico</
Image>
  <Image width="152" height="152" type="image/png">/static/img/logos/
wdl-152x152.3c48b49bd60d.png</Image>
  <InputEncoding>UTF-8</InputEncoding>
  <Url type="text/html" template="http://www.wdl.org/en/search/?qla=en&q={searchTerms}"/>
  <Url type="application/x-suggestions+json" template="http://www.wdl.org/en/search/opensearch-
suggestions/?qla=en&q={searchTerms}"/>
  <moz:SearchForm>http://www.wdl.org/en/search/</moz:SearchForm>
  <Url type="application/opensearchdescription+xml" rel="self" template="http://www.wdl.org/en/
opensearch.xml" />
</OpenSearchDescription>
```

Ako vidíme wdl poskytuje dva rôzne spôsoby vyhľadávania, vo formáte html a dopĺňanie vo formáte json.

WDL používa v HTML kóde HTML 5 Microdata používa schému na popísanie metadát, väčšina programovacích jazykov obsahuje knižnice na narábanie s Microdata-mi, preto api neobsahuje aj výstup v json formáte, bolo by to zbytočné. Poskytuje ale výstup v metadátovom formáte METS a to tak že posledné lomítko v URL nahradíme .mets napríklad:

<https://www.wdl.org/en/item/1.mets>

Na interoperabilitu sa používa International Image Interoperability Framework, ktorý je zatiaľ iba v beta verzii pre obrázky sa používa IIIF Image API 2.0 Pre informácie o obrázku môžeme zavolať <https://content.wdl.org/iiif/{wdlId,volume,page}/info.json>, kde wdlId je id položky v katalógu WDL, volume je číslo knihy ak existuje viacero dielov diela a page je stránka, všetky parametre sú povinné takže aj jedno obrázkové položky musia mať parameter 1,1,1 tým získame json, ktorý popisuje obrázok, neskôr k nemu môžeme prísť tak, že ako je už popísané v dokumentácii IIIF Image API zavoláme

<https://content.wdl.org/iiif/{identifier}/{region}/{size}/{rotation}/{quality}.{format}> kde

- id je identifikátor už hore zmienený
- región je časť obrázku ktorú chceme vybrať, pre celý použije full
- size určuje dimenzie na ktoré chceme obrázok natiahnuť
- rotation je rotácia obrázku
- quality určuje farebnosť obrázku, či je obrázok vrátený farebný alebo čiernobiely
- format určuje file format v ktorom je obrázok vrátený

napríklad <https://content.wdl.org/iiif/1,1,1/full/full/0/default.jpg>

IIIF Presentation API 2.0 sa používa na interoperabilitu medzi záznammi vo WDL manifest daného diela získame tým, že za URL itemu pridáme /manifest napríklad

<https://www.wdl.org/en/item/1/manifest> to nás presmeruje na príslušný .json alebo .jsonld podľa headerov v requeste.

Priebeh pridávania obsahu do WDL

Ak by som chcel pridať obsah do WDL, tak musím postupovať nasledovne. Najprv sa musím stať partnerom WDL tým, že kontaktujem povereného človeka a ostatní partneri WDL musia schváliť moju žiadosť. Potom môžem vytvoriť žiadosť o pridanie obsahu. Upresním o aký

obsah ide a kedy ho môžem dodať. Komisia alebo rôzne špecializované skupiny vyberú obsah, ktorý bude pridaný do knižnice. Vyberajú hlavne materiály, ktoré pomáhajú pochopiť históriu ľudstva, teda hlavne veci, ktoré súvisia s históriou a kultúrou krajín. Najviac sa zameriavajú na materiály ako historické knihy, manuskripty, prvé filmy, prvé zvukové nahrávky atď.

Nahraté média musia spĺňať viaceré kvalitatívne kritériá, preferujú TIFF formát s 24 bitovou farebnou hĺbkou, ale prijímajú aj bezstratové JPEG-y ako obrázky. Čo sa týka videí, tak jedine Apple-ovsky .mov formát alebo .mp4. Podliehajú aj rôznym pomenovávacím konvenciám. Prekvapivo často sa však stáva, že kontribútori nahrávajú súbory, ktoré daným obmedzeniam nepodliehajú.

Každý súbor musí obsahovať popisné metadáta či už vo formáte MARC, MODS, Dublin Core alebo UNIMARC. Takisto majú aj vlastný formát WDL Descriptive Metadata Element Set, ktorý sa jednoducho mapuje na iné metadátové štandarty. Napríklad pri WDL tag publisher sa mapuje v Dublin Core na Publisher a v MARC na Publication. Potom partnerské inštitúcie WDL recenzujú obsah a metadáta nahraných súborov a zhodnotia, či treba ešte daný obsah upravovať alebo nie. Celý proces trvá maximálne dva týždne, ale pri väčšom obsahu sa to môže líšiť. Keď je obsah zrecenzovaný, pošle sa na preklad do ostatných šiestich jazykov.

Zhodnotenie projektu

Stránka bola veľmi dobre interaktívna, intuitívna a veľmi vyspelá čo sa týka použitých technológií a UX. Na stránke je veľa zaujímavých materiálov a dáta su veľmi celistvé vďaka zložitému procesu vkladaniu nových materiálov. Zaujalo ma otvorené api, ktoré stránka ponúka a keďže pracujem ako Android Developer rozhodol som sa vytvoriť klientskú aplikáciu, ktorú by som rád ukázal pri prezentácii. Páčili sa mi interaktívne mapy a časové osy, ktoré stránka poskytuje. Avšak vytkol by som design stránky, ktorý sa mi na niektorých miestach javil byť nedorobený a bol použitý nedoštýlovaný Bootstrap.

Zdroje

http://www.unesco.org/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Van_Oudenaren_26_A_1620.pdf

<https://www.wdl.org>

<http://api.wdl.org/>

<http://www.opensearch.org/>

<https://validator.w3.org/feed/docs/rss2.html>

<https://www.sitemaps.org/>

<http://www.digitalpreservation.gov/meetings/documents/ndiipp13/wdl.pdf>

Metadata v Dublin Core

```
<link title="Dublin Core Metadata Schema" rel="schema.DC" href="http://purl.org/DC/elements/1.1/" />
```

```
<meta name="DC.Title" content="World Digital Library" />
```

```
<meta name="DC.Creator" content="Filip Sollár" />
```

```
<meta name="DC.Subject" content="Digital libraries" />
```

```
<meta name="DC.Subject" content="WDL">
```

```
<meta name="DC.Subject" content="World Digital Library" />
```

```
<meta name="DC.Subject" content="Digitálne knižnice" />
```

```
<meta name="DC.Description" content="Essay about World Digital Library" />
```

```
<meta name="DC.Type" content="text" />
```

```
<meta name="DC.Date" content="2017-11-25" />
```

```
<meta name="DC.Format" content="application/pdf" />
```

```
<meta name="DC.Language" content="sk" />
```