|  |  |  |
| --- | --- | --- |
|  | Zpráva ze služební cestyProjekt „Vytvoření Národní digitální knihovny“ |  |
| CZ 1.06/1.1.00/07.06386 |

|  |  |
| --- | --- |
| Jméno a příjmení účastníka cesty | Rudolf Kreibich |
| Pracoviště – dle organizační struktury | Oddělení technické podpory NDK, 2.2.1 |
| Pracoviště – zařazení |  |
| Důvod cesty | Shromáždění/Konference IIPC |
| Místo - město | Ljubljana |
| Místo – země | Slovinsko |
| Datum (od – do) | 22. 04. 2012 – 26. 04. 2012 |
| Podrobný časový harmonogram | 21. 11. 2012 – odjezd vlakem z Prahy v 00:0222. 11. 2012 – Shromáždění IIPC23. 11. 2012 – Mini workshopy24. 11. 2012 – Pracovní skupiny25. 11. 2012 – Konference26. 11. 2012 – Workshopy, odjezd do Prahy v 16:00 |
| Spolucestující z NK |  |
| Finanční zajištění | Projekt 512 – WebArchiv a Big Data |
| Vztah k projektu | Seznámení s výzkumem a nasazenými technologiemi v partnerských web archivech. |
| Cíle cesty | Zjistit možnosti extrakce webových dat, technologií pro extrakci, způsobu jejich ukládání, optimalizace práce s webovými daty a výzkumu nad webovými daty. |
| Plnění cílů cesty | Cíle byly splněny. |
| Další podrobnější informace | Viz níže |
| Podpora publicity projektu |  |

|  |
| --- |
| Související materiály |
| Materiál | Místo uložení |
|  |  |
|  |  |

|  |  |
| --- | --- |
| Datum předložení zprávy |  |
| Podpis předkladatele zprávy |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Datum | Podpis |
| Podpis nadřízeného |  |  |
| Vloženo na intranet |  |  |
| Přijato v mezinárodním oddělení |  |  |

Zkušenosti

Alexandrijská knihovna přestala používat pro přístup k archivním balíčkům nestabilní NFS ve prospěch AFS. Viz. <http://en.wikipedia.org/wiki/Andrew_File_System>. Dále komprimuje CDX index pomocí zero compression, čímž komprimuje záznamy bez vlivu na výkon. Pro více informací kontaktovat Youseffa Eldakara, vývojáře z Alexandrijské knihovny: youssefeldakar@gmail.com

Lepší indexace Wayback Machine pomocí: <https://github.com/nicosensei/webarchiveaccess>

Memento plugin pro Chrome: <https://github.com/ukwa/tachyon>

Vývoj H3 směřuje ke konceptu Headless browser, IIPC se chopí vývoje.

Monitorovací Frontend k Heritrixu: <https://github.com/ukwa/monitrix>

Výzkum

Nástroje

**ArcLink** – Nástroj budování Webgraphu s časovými řezy, umí využívat jak Wayback API, tak CDX či samotné archivní balíčky. SW je vyvíjen v Old Dominion University, byl budován na rozdíl od WAT, jako veřejný nástroj analýzy WA. Měl by být časem dostupný na: <https://code.google.com/p/arcsys/> - případně na vyžádání u Ahmeda AlSuma: aalsum@cs.odu.edu

Nápady, přejaté i vlastní

Ukládat data nikoliv do WARCů, ale přímo do Hbase.

Advokacie http chyby 451: Unavailable For Legal Reasons - jako přesnější informace vůči 403: Forbidden: <http://www.451unavailable.org/>

Nainstalovat Memento Plugin do referenčního centra a seznámit s ním knihovníky.

Vztah mezi televizí a webem – cross media studies. Viz. např. Hydepark v ČT

Bibliografie by měla odkazovat spíše na WA, než na živý web. Viz. <http://arxiv.org/ftp/arxiv/papers/1105/1105.3459.pdf>

Mapa mrtvých stránek -

Projekty

<https://wiki.digitalmethods.net/Dmi/Winter13SearchingTheArchive>

<http://www.netlab.dk>

<http://dighumlab.com/>

<http://staff.science.uva.nl/~kamps/webart/>

[http://mignify.com](http://mignify.com/)

<http://archivethe.net/en/>

<http://www.domaindarkarchive.blogspot.co.uk/>

<http://data.webarchive.org.uk/opendata/>

<http://ws-dl.blogspot.cz/2013/04/2013-04-19-carbon-dating-web.html>

<https://www.e-helvetica.nb.admin.ch/pages/main.jsf>

Kolekce

Metadata o Kolekci

Pro výzkum je třeba definovat důvody ke vzniku kolekce a zaznamenat proces selekce. Výzkumníky bude zajímat proč je co archivováno, proč tak hluboko a proč to tak vypadá.

Nápady

Budování kolekce jako sociální aktivita: <http://www.slideshare.net/phonedude/old-dominion-university-computer-science-iipc-new-member>

Budování kolekce k ochraně citací ve vědeckých pracích.

Kolekce jako ochrana osobního digitálního odkazu (personal heritage). Tj. WA jako archivační nástroj pro uživatele.

Institucionální budování kolekce

Crowd sourced heritage

Propagovat archivovatelný web jako další koncept vedle good code aj. konceptů. Orientovat se na weby alá: jakpsatweb.cz, propagovat mezi marketéry, studenty nových médií, programátory – ukázat jak jejich web mizí. Kontaktovat skupiny kolem větších CMS

Projekty

Archiv ke konci prezidenství: <http://eotarchive.cdlib.org/index.html>

Topsy – twitter archiv: <http://topsy.com/>

Nástroje pro budování kolekce

**Twittervane**: Umožňuje kurátorům definovat webové kolekce a na základě nich identifikovat relevantní trendy a URL v tweetech. <http://www.webarchive.org.uk/twittervane/>, <https://github.com/ukwa/twittervane>

**AtN2**: ArchiveTheNet2 hledání relevantních URL k archivaci, na základě sociálních sítí. Přístup do privátní Bety: <http://archivethe.net/beta>. Základní info: <http://archivethe.net/en/index.php/news/leveraging_social_web_to_propel_your_archiving_campaign_with_atn2/>

Web Archiving Integration Layer – mohl by být Frontend nad Heritrixem a Waybackem pro naše kurátory. <http://matkelly.com/wail/>

Jak propagovat WA

Memento jakou standardní součást webového prohlížeče

Zdá se, že nejzásadnějším rozšířením povědomí o WA , vede cestou implementace Memento Plugin jako standardní součásti prohlížečů. Uživatelské paradigma bude počítat s posuvníkem času, který bude načítat historické verze URL z různých světových WA. Čímž bude uživatel bezprostředně konfrontován s temporalitou webové zkušenosti, obdobně jako je s ní konfrontován na časové ose facebooku.

Google nás nezachrání

Google je opatrný s obsahem který spadá pod autorský zákon, z této perspektivy bude těžko soupeřem WA. Jeho cache se zabývá primárně dostupností poslední textové verze obsahu na dané URL. Ponaučení: Google nás nezachrání.

Výzkum

Je potřeba připravit nástroje nad daty a metadaty WA a seznámit s nimi výzkumníky. Jednou z forem je spolupráce s vytipovanými katedrami napříč Vysokými školami a prezentování možností WA na příkladě nástrojů a hotových analýz. Budovat výzkumné skupiny nad daty WA. Např. <https://wiki.digitalmethods.net/Dmi/Winter13SearchingTheArchive>

Prezentace IIPC Staff Exchange 2012

Prezentoval jsem své zkušenosti z pobyt v Internet Archive v San Franciscu na jednom z workshopů a diskutoval očekávání jednotlivých knihoven.