

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



Internet Memory

Uchování kulturního dědictví na webu <http://internetmemory.org/en/>

Esej do předmětu PV070 Digitální knihovny

David Holík, UČO: 324970

10. prosince 2013

Úvod

Od úsvitu lidské civilizace až po současnou modernu 21. století se dá v dějinách vystopovat celá řada charakteristických projevů lidstva. Ta jako červená nit protkává stejně tak dobře životy dvou po sobě jdoucích generací, tak i lidských kultur, mezi nimiž je na časové ose rozdíl tisíců let. Jedním z takových rysů je tendence lidí uchovávat záznamy o sobě samotných, o způsobu jejich života, okolním světě zahaleným rouškou tajemna, který byl předmětem dobrodružného poznávání, o nastolených zvyklostech a v neposlední řadě o poznatcích a vynálezech, jež byly s velkou slávou vyvádány. Potřeby a důvody pro takové ukládání záznamů, stejně tak úroveň ovládnutí technologických procesů, se v průběhu času měnily. Avšak samotné přemítání o tom, jak nejlépe se vypořádat s nelítostným působením času a jaké schopnosti je potřeba si osvojit, aby vyprodukované informace bylo možno odkázat generacím příštím, zůstává neměnné. Leckoho by mohlo svádět označit takovou problematiku v dnešní době za triviální. Technologický vývoj za posledních pár desetiletí přinesl výrazný pokrok mimo jiné na poli výpočetní síly počítačů, diskových kapacit či zkvalitnění a rozšíření síťové architektury. A přesně tyto jmenované prvky jsou předurčeny k tomu, aby pomáhaly ukládat informace, které naše společnost vytváří a zakonzervovat je i pro budoucí pokolení. Ovšem jak již bývá nepsaným pravidlem, pokrok nové doby většinou sice přinese odpovědi na některé otázky, které dříve nebylo možné zodpovědět, avšak příchod nových technologií často předurčí i zrození otázek zcela nových. A není jistě překvapením, že i v rámci naší soudobé společnosti dochází k vyvstávání celé řady podnětů k přemýšlení a následnému hledání optimálního řešení nastolené problematiky.

Specifický svět Internetu

Internet je bezpochyby médiem nového tisíciletí. Celosvětová rozšířenost, rychlost toku dat a jejich ohromné množství, rozmanitá škála služeb a oblastí využití, to vše při aktivním zapojení stovek milionů lidí dělá z Internetu hlavní zdroj informací v rámci dnešní společnosti. Internet jako takový rovněž představuje ideální nástroj, prostřednictvím kterého lze široké veřejnosti umožnit přístup k archivovaným položkám. Kromě klasických knihoven, které již několik staletí fungují jako instituce bdící nad uchováním kulturního dědictví budoucím generacím, hovoříme i o tzv. digitálních knihovnách. Ty shromažďují různorodé digitální kolekce a skrze internetové rozhraní dávají lidem možnost k těmto digitálním sbírkám přistoupit. Obsah kolekcí může být tvořen artefakty, jejichž původ nemá s počítačovými technologiemi nic společného, ale procesem digitalizace získaly i svoji digitální kopii (knihy, mapy, časopisy, noviny, staré audio nahrávky a další). Na druhou stranu existuje ohromné množství děl, které již v digitální podobě vznikly (označováno termínem *born digital*) a u nichž je přirozené, že se v určitých digitálních kolekcích budou organizovat. Kromě tradičně chápaných děl se však nesmí zapomenout ani na Internet samotný, myšleno ve smyslu webových stránek, které jsou významným zdrojem informací a které by bylo rovněž žádoucí archivovat.

S příchodem nového milénia spatřilo světlo světa mnoho institucí či organizací, které si webovou archivací vytyčily jako předmět svého projektového zájmu. Za jednu z takových organizací lze považovat i neziskovou instituci *Internet Memory Foundation* [1] (dále v textu zkracováno na IMF). Ta byla založena roku 2004, tenkrát ještě pod názvem *European Archive*. Korporátně spravována z Paříže a Amsterdamu si tato instituce vymezila cíl podniknout kroky vedoucí k archivaci webového obsahu.

Internet Memory Foundation

Důvod pro ustanovení názvu IMF z původního European Archive spočíval zejména ve vyjádření podstaty, že tato instituce nemá zájem jen na tom provozovat archivační platformu, ale že se chce problematice permanentního ukládání webového obsahu věnovat poněkud komplexněji, nejlépe s využitím co nejširšího portfolia dostupných technologických postupů (data mining, sémantický web, práce s metadaty, vyhledávací algoritmy a mnohé další), které mohou výrazně přispět k lepším výsledkům při samotném provozu archivu. Bylo by tedy velkou chybou chápat instituci Internet Memory jen jako provozovatele archivu pro uchovávání webového obsahu. Naopak IMF si klade za cíl z pozice neziskové organizace poskytnout informační a jiný servis pro subjekty různého typu mající společně to, že se zabývají problematikou archivace webu. IMF je rovněž aktivním hráčem na poli pořádání různých konferencí či workshopů, které mají za úkol vytvořit prostor pro diskusi a šíření nejnovějších poznatků ze studované oblasti.

Pokud se na celou problematiku podíváme jinou optikou, konkrétně z pohledu geograficko-politických poměrů v Evropě, je přirozené domnívat se, že takto nastolená otázka archivace webového obsahu, zejména toho mající kulturní a společenský přesah, je v souladu s vizí Evropské unie. Konkrétně Evropská komise, která má v popisu práce ustanovení programů, do nichž se přihlašují různé projekty, které podporují myšlenky těchto programů, a tím pádem i vize o vývoji a směřování Unie jako takové. Informační a komunikační technologie (ICT) představují jeden z klíčů pro vybudování moderního společenství a z tohoto důležitého statutu pramení i hojnost programů pro rozvoj v oblasti ICT, ať již z hlediska reálného provozu, tak i výzkumného charakteru. A právě k evropskému výzkumu ve sféře ICT neodlučitelně patří tzv. *Framework Programmes for Research and Technological Development*, často zkráceně označováno jako *Framework Programmes (FP)*. [2] Tento typ programu se spustil v roce 1984 a je sekvenčně očíslován pro jednotlivé časové periody (od FP1 až po FP8), každému z těchto programů se přidělí určitý rozpočet, z něhož projekty zapsané do těchto programů čerpají finance. Pro účely naší eseje postačí zmínka o FP7, tedy The Seventh Framework Programmes. Ten je vymezen časovým obdobím mezi rokem 2007 a 2013. Celková výše rozpočtu pro FP7 činí přes 50 miliard € a k programu byla přidružena celá řada projektů, mezi nimiž pochopitelně nechybí ani projekty zabývající se archivací webového obsahu. Již o pár řádků níže bude v krátkosti představena šestice takových projektů, které probíhají (popřípadě již proběhly) pod záštitou IMF. Tady ovšem pozor na správné pochopení situace, níže uvedené projekty nejsou přímým produktem IMF. Projekty mají svoji vlastní řídicí i výzkumnou strukturu, IMF zde hraje roli spíše externí záštity, popřípadě partnera, která poskytuje informační servis, dodává své know-how či zprostředkovává komunikační kanály s dalšími relevantními subjekty. A jak uvidíme později, některé prvky, které jsou předmětem výzkumu těchto projektových skupin, implementovala IMF do své archivační služby, respektive IMR (Internet Memory Research), kterou lze chápat jako výzkumnou sekci tohoto institutu.

Vybrané výzkumné projekty v rámci FP7

Nyní se dostáváme k náhledu na šest výzkumných projektů, které běží v rámci FP7 a v nichž IMF v jisté míře participuje, neboť předmět výzkumu spadá do problematiky archivace webového obsahu.

• Longitudinal Analytics of Web Archive data (LAWA)

Projekt LAWA [3], jak již samotný název napovídá, přinesl poznatky zejména v oblasti porozumění datům jako takovým. Analyzování dat, která se sbírala do v rámci projektu zbudovaného „testbedu“, přineslo celou řadu uspokojivých výsledků, které již nacházejí praktické využití v reálně fungujících systémech. V obecné rovině lze konstatovat, že přínos LAWA projektu se může odrazit do oblastí, jakými jsou skladování, indexace, efektivní distribuce dat, dolování informací z datových skladů nebo analýza datových struktur mající povahu grafových úloh. Za zmínku rozhodně stojí uvést, že projekt LAWA byl na podzim roku 2013 oficiálně uzavřen hodnotícím orgánem Evropské komise, který dosažené výsledky a přínos v rámci řešené problematiky ohodnotil termínem „excellent“. [4]

• Scalable Preservation Environments (SCAPE)

Zaměření projektu SCAPE [5] se dá ve zkratce popsat jako vývoj škálovatelných služeb pro plánování a samotné provádění uchovávání digitálních informací na úrovni jednotlivých institucí a organizací. Strategie takové archivace musí být založena na platformě open source, doplněná o poloautomatický workflow, který umožní organizování velké, škálovatelné a heterogenní kolekce digitálních objektů. Takto vytvořené služby budou mimo jiné schopny identifikovat požadavky pro uchování digitálních entit, umožní využít výsledky na poli analytických trendů či nástroje pro podporu škálovatelnosti a v neposlední řadě připraví prostředí pro monitoring kvality samotného archivačního procesu.

• Large-scale, Cross-lingual Trend Mining and Summarization of Real-time Media Streams – TrendMiner

Projekt TrendMiner [6] reflektuje masivní rozvoj v oblasti online médií, sociálních sítí, vzájemného propojování relevantních odkazů v rámci nesusoudných internetových aplikací, společně s nárůstem šíření mediálního obsahu v reálném čase, a to vše častokrát s vícejazyčným kontextem. Všechny uvedené aspekty vedou ke vzniku výzvy, jak nejlépe přistupovat a interpretovat tato silně vícevrstvá data, přirozeně časově, efektivně a dostupně přijatelným způsobem. Předpokládá se, že výsledky projektu budou moci být použity jednak v soukromé „business“ sféře (zejména komunikace a podpora zákaznické komunity), tak i ve státní správě (například lepší přístup k elektronickým zdrojům shromažďující informace o činnosti vlády nebo samosprávných celků). Své využití rovněž najde ve statistických prognózách vyžadující přehled o názorech napříč sociální strukturou.

• Collect-All Archives to Community Memories (ARCOMEM)

Nosnou ideou projektu ARCOMEM [7] je, že při archivaci internetového obsahu se nesmí zapomenout na sociální sítě, které v moderní době představují velice důležitý zdroj informací, které by bylo žádoucí uchovávat. Komunikaci na sociálních sítích již nelze dávno jednoduše označit nálepkou neformální. S rostoucím vlivem takovýchto sítí roste i počet využití v oblasti formální komunikace (např. v politice při užším spojení s voličstvem). Specifické prostředí sociálních sítí však vyžaduje specifický přístup k archivaci jejich obsahu. To vše díky hlubokému pochopení fungování sociálních sítí, chování uživatelů i přesnému vymezení sémantiky, která se v tomto prostředí objevuje.

• Living Web Archives – LiWA

Již proběhnuvší projekt LiWA [8] přinesl sadu výsledků v oblasti archivace webu, ať se jedná o metodiku či použité nástroje. Projekt se zabýval studiem významu sémantického popisu webu pro zlepšení kvality a věrnosti při archivaci. Pozornost byla rovněž věnována metodikám pro filtraci spamu, detekce pastí, které mohou zkomplikovat archivační proces, pohyb na úrovni tzv. „deep webu“ a mnoho dalšího.

• LivingKnowledge – LK

Informace na Internetu se vyznačují velikou rozmanitostí, které jsou přirozeně dány geografickými, kulturními, ekonomickými a dalšími kontexty. Široká škála takové rozmanitosti je samozřejmě vítána, ovšem například vyhledávání v takto heterogenním prostředí nemusí přinést očekávané výsledky. Hlavní ideou projektu LivingKnowledge [9] je při zachování rozmanitosti archivované kolekce podpořit rysy jako výsledovatelnost, srozumitelnost a využitelnost informací. Pozitivní dopad projektu by se měl projevit zejména na zlepšení vyhledávacích mechanismů, které přinesou komplexnější a přesnější výsledky hledání.

ArchivetheNet – obecný rámec

ArchivetheNet (AtN) [10] představuje platformu pro webovou archivaci, jež je spravována institucí Internet Memory, konkrétně její sekci Internet Memory Research (IMR), která reprezentuje interní oddělení IMF zabývající se výzkumem technologických aspektů, jež úzce souvisí s archivací internetového obsahu. V předchozí kapitole jsme si představili šestici evropských projektů, v nichž Internet Memory vystupovala v roli externí spolupracující entity, ArchivetheNet je na rozdíl od těchto projektů zcela interní záležitostí, běžící tedy plně pod správou IMR. Jak již ale bylo nastíněno a jak ještě k této záležitosti bude věnována pozornost, je přirozené, že myšlenky a rovněž pochopitelně výsledky výše zmíněných projektů implementovala IMR i do své archivační platformy.

V současné době je ArchivetheNet nabízen zejména kulturním institucím, které jeví zájem tendenci ve využívání webové archivace k ukládání nabízeného kulturního dědictví pro současnou i budoucí generaci. Pokročilá archivační metodika je důvodem zlepšení kvality výsledné digitální kolekce, takže není problém archivovaný webový obsah prohlížet a procházet způsobem, jako by byl plně ve standardním online režimu. Použitím sdílené platformy ArchivetheNet získávají zúčastněné organizační jednotky výhody v obou směrech. Na jednu stranu zůstává těmto institucím plná kontrola nad spravovanou archivační sbírkou, spolu s individuální možností ustanovení archivační politiky, nastavení parametrů pro provádění archivace. Na stranu druhou však dostanou právo k využití dedikované a optimalizované infrastruktury pro sběr dat (tzv. data crawling).

Výhody platformy ArchivetheNet [10]

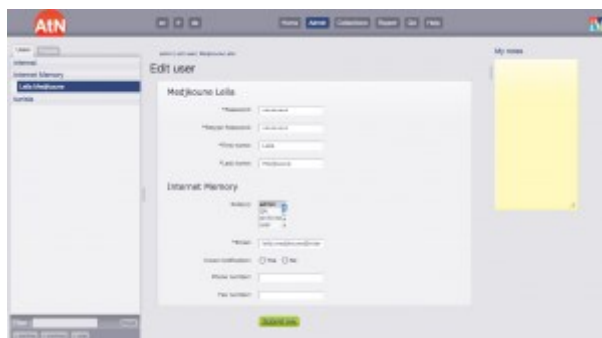
- Nižší finanční náklady dané kontrolou peněžních poplatků, které jsou založeny na aktuální podobě digitální sbírky (velikost, heterogenní povaha formátů archivovaných dat).
- Čerpání výhod z pokračujícího vývoje v oblasti archivace webu, a to v rámci mnoha projektů, do nichž je instituce Internet Memory zapojena. Jedná se kupříkladu o výše zmíněné evropské projekty Liwa, LivingKnowledge a další.

- IMF je významným, aktivním členem IIPC (International Internet Preservation Consortium), což vede ke konstatování, že AtN bude mít možnost implementovat nejlepší metodiky v oblasti permanentního uchovávání webových dat, ať již se jedná o využití metadat, provádění redundantních záloh atd.
- Fulltextové vyhledávání v archivované kolekci.
- Data jsou uložena na speciálně zřízeném datovém uložišti, které splňuje nejpřísnější kritéria pro bezpečné a dlouhotrvající ukládání digitálních informací. Zapojené instituce a organizace se tak nemusejí starat o způsob fyzického ukládání dat, ale mohou svoji pozornost zaměřit na ustanovení parametrů pro samotný výběr archivovaných údajů a následné usnesení o tom, jaké politice bude podléhat zacházení s již archivovanou kolekcí dat.
- Možnost užití pokročilých modulů pro sběr dat mající multimediální charakter a rovněž dat, které jsou obsahem sociálních sítí.
- Ve sběračích dat je implementována antispamová detekce, která s poměrně vysokou spolehlivostí dokáže redukovat množství nechtěného obsahu. Díky tomu lze ušetřit 20 - 30 % místa na paměťovém uložišti, a tím pádem přirozeně snížit dobu archivace i cenové náklady.
- Je zde funkcionality, jež dokáže přistoupivšího uživatele na web organizace automaticky přesměrovat na archivační verzi namísto původní webové stránky, která již v danou chvíli může být smazána, přetížena nebo z jiného důvodu nedostupná => potlačení nutnosti zobrazení „404 error message“.
- Poskytování pravidelných dávek kopií archivované kolekce ve formátu ISO 28500:2009 WARC.

ArchivetheNet – konkrétní podoba řešení

Platforma ArchivetheNet je navržena jako softwarový produkt typu SaaS (Software as a Service), což je zajímavý koncept, který poslední roky nabývá na oblíbenosti. Čtenářům, kteří o pojmu SaaS neslyšeli, se pokusím v následujících řádcích nastínit, jakých principů tento přístup využívá. Pod konceptem SaaS [11] se jednoduše řečeno skrývá model nasazení softwaru, kdy dochází k hostování aplikace provozovatelem služby. Služba je zákazníkům distribuována skrze rozhraní Internetu. Výhodou takového řešení je fakt, že dochází k eliminaci potřeb instalace a provozu služby na vlastním zařízení. Dalším plusem je bezpochyby možnost rychlého nasazení a snižování nákladů na software, alespoň co se počátečních investic týče. Software se totiž neplatí na začátku jednorázovou platbou, nýbrž ve formě paušálních poplatků, jejichž výše reflektuje rozmanitost portfolia využívané funkcionality. A to je přesně koncept, který je v této archivační oblasti notně prospěšný, neboť nepředstavuje veliké riziko spojené s finanční ztrátou vysokých počátečních vkladů v případě, že by se od archivace upustilo. Kromě toho takto nastavený finanční rámec reflektuje heterogenní povahu samotného Internetu, kdy každá organizace má zájem ukládat rozdílné objemy dat, data různých formátů apod.

V momentě zapojení do platformy ArchivetheNet, dostane příslušná organizace možnost přístupu k rozhraní AtN administrace. Ta je koncipována ve formě modulů, které jsou logicky členěny a poskytují veškerý servis ke správě související s archivací dat. Níže uvádím tři stěžejní moduly se zachováním originálního pojmenování v anglickém jazyce. Jednotlivý přehled poskytované funkcionality daných modulů již pochopitelně předkládám v češtině.



Obrázek 1: Zmenšená podoba rozhraní administračního AtN modulu

- **AtN Administration Module: Who will use your Web archiving platform**

V tomto administračním modulu (viz obr. 1) je možno spravovat uživatelské účty související s danou archivační kolekcí. Lze vytvářet, modifikovat a deaktivovat profily uživatelů, kteří v různých rolích a tím pádem na různé autorizační úrovni mohou přistupovat k příslušné kolekci.

- **AtN Collection Module: Where you will put in place your selection policy**

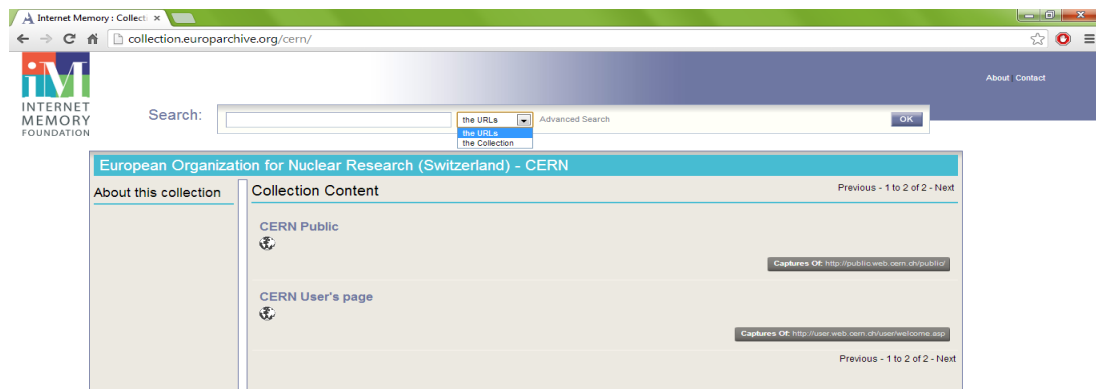
Jak již dává tušit název modulu, jedná se o rozhraní, skrze které je povoleno přistupovat k archivované datové sbírce. V případě, že autentizační proces potvrdí požadovanou úroveň oprávnění, může přistoupivší uživatel spravovat archivovaná data. Stejně tak je tento modul místem, kde se nastavují parametry, kterými se řídí procesy v rámci sběru a ukládání dat.

- **AtN Report Module: All you want to know about your captures**

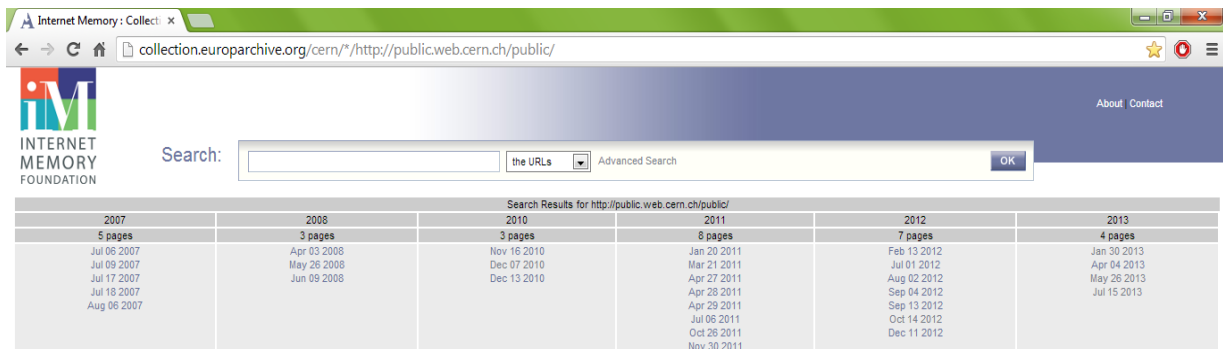
Poslední modul, který zde uvedu, představuje sice ne úplně nezbytně nutné prostředí pro manipulaci s kolekcí archivovaných dat, přesto by měl být nedílnou součástí pro takovou datovou správu. Tento modul totiž poskytuje statistické výstupy, které jsou předmětem analýzy uložených dat. Lze zvolit ze široké škály statistik, měřených parametrů, které v důsledku mohou příslušné organizaci umožnit lepší pochopení struktury archivovaných dat (globální pohled na data, statistické analýzy v rámci jedné kolekce, analýza jednotlivých entit sdružovaných do kolekcí).

Rozhraní archivu

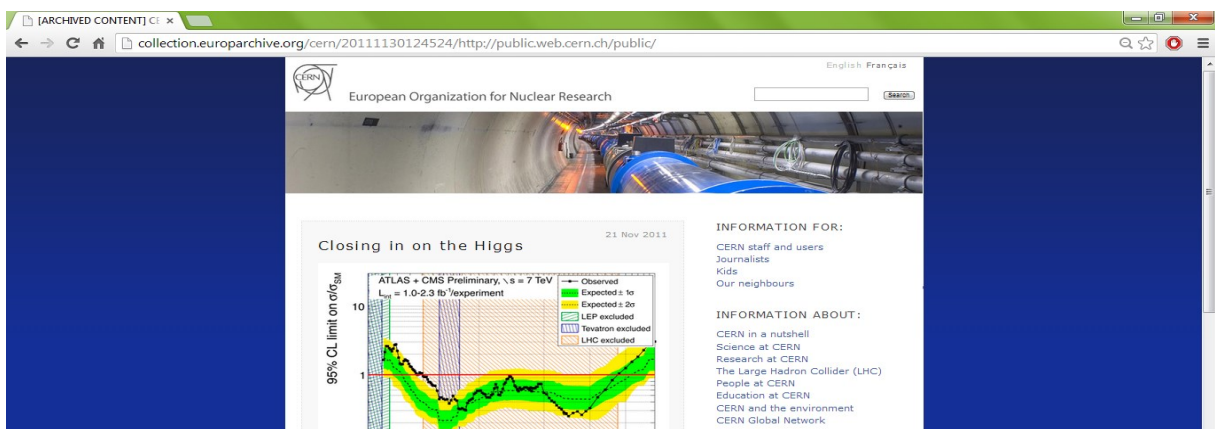
Nyní nastává nejvyšší čas představit si samotné rozhraní archivu. V něm libovolný internetový uživatel může přistoupit k archivované kolekci webových stránek organizací, které takový přístup povolily. Vše si ukážeme na archivaci webového obsahu institutu CERN, konkrétně si předvedeme, jak je snadné na tři kliknutí zobrazit podobu internetových stránek z listopadu roku 2011 (obr. 2, 3 a 4)



Obrázek 2: Rozhraní pro výběr archivované kolekce organizace CERN (Volíme odkaz CERN Public)



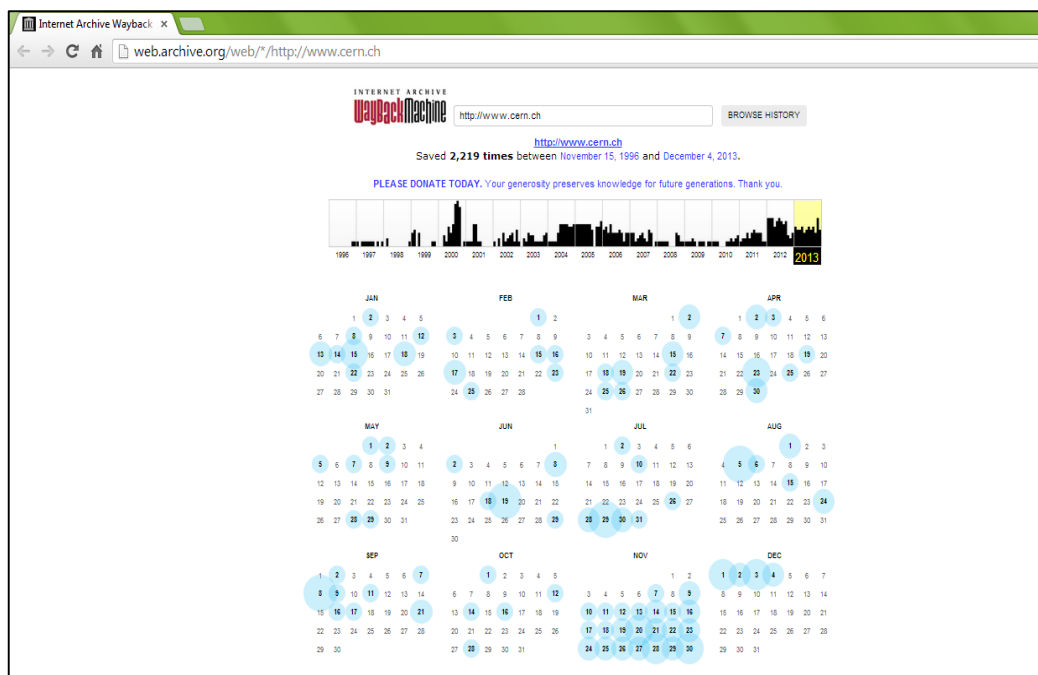
Obrázek 3: Rozhraní pro výběr archivované jednotky v rámci kolekce CERN Public (Volíme odkaz Nov 30 2011)



Obrázek 4: Zobrazení webové stránky aktuální k datu 30. listopadu 2011

Aktuální stav

Osobně ve mně současný stav archivační platformy od organizace Internet Memory evokuje přiřadit k tomuto projektu termín mezidobí. Podle mě totiž nejlépe vystihuje aktuální situaci, která je ve velké míře ovlivněna samotným procesem změny názvu instituce, spolu s hledáním a posunem své budoucí působnosti. Sekce Internet Memory Research sice spravuje webovou archivační platformu AtN, rovněž jsme si na příkladu CERNu ukázali, že ke kolekci ukládaných dat lze pohodlně přistupovat a archivace jako taková funguje. Problém by někteří lidé mohli spatřit v tom, že díky úzkému cílení na politické, vědecké či kulturní instituce, které navíc prozatím ve velkém měřítku nepřišly tomuto archivu na chuť, je množství uložených kolekcí poněkud malé. Kromě zmíněného institutu CERN, existuje možnost se obdobným způsobem dostat k archivovaným webovým informacím například u subjektů: *Biblioteca Nazionale Centrale de Firenze*, *UK Parliament*, *National Library of Ireland* a další. [10] Na jednu stranu lze v tomto vidět limitující omezení, na straně druhé je potřeba chápat širší kontext. Internet Memory si neklade za cíl ohrožit jedničku na archivačním trhu, kterým je bezpochyby *Internet Archive* [12], který míří na širší klientelu a k součnému dni eviduje bezmála 370 miliard archivovaných stránek. Pro lepší srovnání si ještě jednou vypomůžeme CERNem, který obsah svých webových stránek archivuje jednak skrze produkt IMF, tak i pomocí již zmíněné platformy Internet Archive. Sami si můžete porovnat výše vystavené obrázky 3 a 5, které krásně dokumentují, že archivační služba Internet Archive je jasným způsobem preferována.



Obrázek 5: Přehled dnů v roce 2013, ve kterých došlo ke sběru dat na stránkách <http://www.cern.ch> pro následnou archivaci na platformě Internet Archive

Internet Archive – lídr webové archivace

Jak jsem již uvedl, je to právě Internet Archive (IA) [12], který hraje prim v oblasti ukládání webového obsahu. O této instituci pojednával ve své esaji Marek Hlaváč [13], do níž mohou případní čtenáři nahlédnout a ve které se dozvědí spoustu informací o principech fungování IA a spravovaných digitálních kolekcí. Od doby svého založení (rok 1996) urazil IA dlouhou cestu, na níž vybudoval robustnou archivační platformu, která seskupuje miliony digitální objektů různých typů. Jedná se o texty, audio, video nahrávky nebo i software (zejména programy a hry staršího data vydání) a v neposlední řadě zde přirozeně figuruje i archivace webových stránek. Sesbírané internetové stránky se ukládají v datovém centru IA, kde se třídí a uspořádávají do kolekcí, které jsou pak zpřístupněny uživatelům. Neexistují-li žádná legislativní a vlastnická omezení, může k těmto kolekcím přistoupit jakýkoli uživatel, konkrétně skrze webovou službu Wayback Machine implementovanou jako nedílnou součást IA. Ta umožňuje navštívit archivní verze internetových stránek, a to velice intuitivním způsobem v uživatelsky příjemném prostředí. Jediné, co je potřeba, je zadat příslušné URL. V případě, že internetová stránka s daným URL identifikátorem má v archivu uložené položky, lze si tyto archivní verze dané webové stránky prohlížet. Podoba rozhraní Wayback Machine je zachycena na výše uvedeném obrázku číslo 5.

Závěrečné slovo

Své dojmy ze studované problematiky bych označil slovem rozporuplné. Na jednu stranu mi svět webové archivační technologie přišel fascinující a z důvodů, které jsem nastínil v samém úvodu své práce, rovněž velice potřebný. Na straně druhé mě poměrně výrazně zklamal samotný archiv spravovaný institucí Internet Memory, od kterého jsem si zkrátka sliboval více (zejména, co se množství obsahu a počtu zapojených subjektů týče). Pokud by tato archivační platforma měla do budoucna zůstat v neměnné podobě, stěží najdu důvod pro její znovu navštívení.

Částečné vykompenzování tohoto negativa přinesla možnost nahlédnout při bádání do zákoutí mnoha zajímavých projektů, do kterých se IMF zapojila. Rovněž jsem dostal možnost připomenout si významný podíl Evropské unie na rozvoji v oblasti informačních technologií. A přesně v tomto duchu, ve kterém mě řešené téma oslovilo, jsem se snažil sepsat tuto esej. Věřím, že případným čtenářům poskytne vkusně sepsaný úvod do světa archivace webu a také, že se mi dostatečně podařilo podnítit zvědavost a zájem dozvědět se další hlubší informace z této oblasti. Práce se sice zaměřuje na aktivity instituce Internet Memory, ale jak je již patrné ze zmínky o „konkurenčním“ Internet Archive, existuje mnoho podobných platforem, které umožňují archivovat webový obsah. A právě v tomto kontextu je potřeba činnost IM chápat – nikoli jako provozovatele jediného podobně zaměřeného archivu, na jehož bedrech by stálo veškeré současné dění kolem archivace webu, ale spíše jako jednu z mnoha institucí, která si webovou archivaci vytyčila jako svůj cíl. Je logické, že takto široce vymezená tematika se dá pojmout mnoha způsoby a až čas ukáže, jestli směr, kterým se vydala IMF, přinese více pozitiv a úspěchů nebo naopak zklamání z nenaplněných očekávání.

Citované internetové zdroje

- [1] *Official website of „Internet Memory Foundation“*, [online], 2013, [cit 2013-11-26]. Dostupné z WWW: <<http://internetmemory.org/en/>>
- [2] European Commission. *Community Research and Development Information Service*, [online], 2013, [cit 2013-11-27]. Dostupné z WWW: <http://cordis.europa.eu/projects/home_en.html>
- [3] *Official website of „LAWA Project“*, [online], 2013, [cit 2013-11-29]. Dostupné z WWW: <<http://www.lawa-project.eu/>>
- [4] Internet Memory Foundation. *„News: Successful end of LAWA project“*, [online], 2013, [cit 2013-11-29]. Dostupné z WWW: <http://internetmemory.org/en/index.php/news/successful_end_of_lawa_project>
- [5] *Official website of „SCAPE Project“*, [online], 2013, [cit 2013-12-02]. Dostupné z WWW: <<http://www.scape-project.eu/>>
- [6] *Official website of „TRENDMINER Project“*, [online], 2013, [cit 2013-12-02]. Dostupné z WWW: <<http://www.trendminer-project.eu/>>
- [7] *Official website of „ARCOMEM Project“*, [online], 2013, [cit 2013-12-02]. Dostupné z WWW: <<http://www.arcomem.eu/>>
- [8] *Official website of „LiWA Project“*, [online], 2013, [cit 2013-12-02]. Dostupné z WWW: <<http://www.liwa-project.eu/index.php>>
- [9] *Official website of „LivingKnowledge Project“*, [online], 2013, [cit 2013-12-02]. Dostupné z WWW: <<http://livingknowledge.europarchive.org/>>
- [10] Internet Memory Research. *ArchivetheNet*“, [online], poslední aktualizace 3. prosince 2013, 14:00, [cit 2013-12-04]. Dostupné z WWW: <<http://archivethe.net/en/index.php/>>
- [11] Wikipedia. *Software as a service*, [online], 2013, [cit 2013-12-02]. Dostupné z WWW: <http://en.wikipedia.org/wiki/Software_as_a_service>
- [12] *Official website of „Internet Archive“*, [online], [cit 2013-12-04]. Dostupné z WWW: <<https://archive.org/>>
- [13] HLAVÁČ, Marek. *Internet Archive*, [online], 2011, [cit 2013-12-04]. Dostupné z WWW: <https://dspace.muni.cz/handle/ics_muni_cz/975>

Metadata (Dublin Core)

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content=" Internet Memory – Uchování kulturního dědictví na webu
      http://internetmemory.org/en/" />
<meta name="DC.Creator" content="David Holík" />
<meta name="DC.Language" content="cs-CS" />
<meta name="DC.Date" content="2013-12-05" />
<meta name="Type" content="text" />
<meta name="Type" content="esej" />
<meta name="DC.Format" content="application/pdf" />
<meta name="DC.Format.medium" content="computerFile" />
```