

Metadata Management and Future Plans to Generate Linked Open Data in the Hungarian Web Archiving Pilot Project

In this article we would like to offer a short overview about the metadata management model of our web archiving pilot project together with international recommendations as a major background of modelling. It is including an outlook to the scope of metadata management (archive-level and website-level), an overview of major metadata types and description of some major metadata fields (more than one hundred fields are available). Metadata based full-text search and retrieval capabilities are also being described in the article.

The second chapter of the article points out that the absence of efficient and meaningful exploration methods of the archived content is a really major hurdle in the way to turn web archives to a usable and useful information resource. A major challenge in information science can be the adaptation of semantic web tools and methods to web archive environments. The web archives must be a part of the linked data universe with advanced query and integration capabilities, and must be able to directly exploitable by other systems and tools. We would like to describe some basic considerations in order to successfully manage this semantic web integration process as a plan to the future.

A short introduction to Web archiving

Web archiving is a process to collect segments of the world wide web. It is rather important to ensure the information is preserved in an archive for future researchers and to the general public. Without preserving the web appearance of our daily life and the cultural heritage on the web, a significant portion of our past could disappear forever for the future generations. Web archiving has started from the mid 1990's by the appearance of the Internet Archive. It is the largest web-archive however the public content has not indexed and only retrievable by very limited capabilities. The lack of full-text search and retrieval tools in a big data environment in order to get relevant hits by multiple types of queries is a significant deficiency of web archiving services. This challenge has to inspire future research activities. (Figure 1) The International Internet Preservation Consortium (IIPC)¹ has coordinated the professional efforts of national public heritage projects and the activities of major stakeholders in web archiving field. The basic software infrastructure of web archiving was born by the help of these coordinated activities.

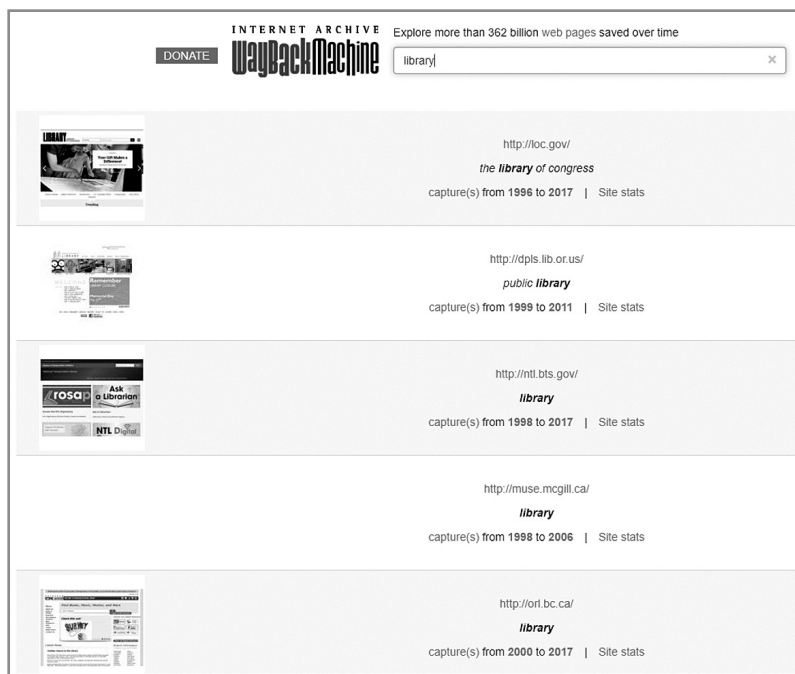


Figure 1 An example to the ineffectiveness of Internet Archive information retrieval services

In one hand web archivists generally archive various types of web content including HTML web pages, style sheets, JavaScript, and graphic elements. Saving video streams mainly can be assigned to media archives as a special task. A major challenge is that the content formats of the web can become outdated in a really fast manner (e.g. the rise and fall of the Flash video format). From an archiving perspective it is a special challenge to preserve special software, build-up emulation techniques in order to access these kind of web documents in the future.

On the other hand, the most common web archiving technique uses web crawlers to automate the process of collecting web pages. By the help of an IIPC - based collaboration the generally used web crawler (robot) Heritrix has been developed. Internet Archive have made available the Open Wayback software in order to display the archived content. Web crawlers access web pages in the same way that users with a browser see the Web, and therefore provide a comparatively simple method of remote harvesting web content. The main challenge is that these crawlers are being optimized for standard web content (web 1.0). Social media content and web-based databases are inaccessible by this preserving technique. Nowadays a new method of

¹ <http://www.netpreserve.org>

browser emulation is being developed by several projects. In this case a headless browser (without graphic interface) can be adjusted to crawl web content by imitating the human browsing behaviour. In our hope this new software generation will enable us to archive content more broadly and effectively in the near future.

Another major challenge of web archives is the legal background, especially in Europe. While the collection of Internet Archive is available publicly, the overall majority of the web archive collections in the European web archives are inaccessible for the public, or access can be made in a really restrictive manner (e. g. by a network of public terminals in a memory institution) due to legal reasons. Each archived website can be made public only by a written agreement with the content owner (legal owner). It is a major drawback for representing this important segment of cultural heritage to the general public.

Financial and professional background of each web archiving project is a critical factor of success. In general, most of the projects are suffering somehow by the lack of financial and human resources. It has a certain effect to the relatively low dynamics of development of web archiving research field.

Web archiving in Hungary

In April 2017 the National Széchényi Library has started a web archiving pilot project as a part of its comprehensive IT infrastructure development programme. The deadline is the end of 2019 and the technological background is provided by the government informatics agency KIFÜ. The aim of the research and development activities in the pilot project is the establishment of a prospective Hungarian Internet Archive. A technological infrastructure must be ready for collecting, harvesting, preserving a large amount of data from the public internet. A service framework must be managed by this infrastructure to make the preserved resources available in accordance to the legal rules and regulations.

Experts must be trained to grant web archiving skills in the Hungarian GLAM sector (galleries, archives, libraries, museums). An accredited course is being offered in collaboration with the Library Institute. An active membership of the Training Working Group of the International Internet Preservation Consortium (IIPC) helps us to learn from international best practices in this field.

Relevant documentation must be made to establish a web archiving workflow. (E. g. collection description, the limits of the Hungarian web sphere, metadata structure of the harvested sites, contract samples for the content providers to grant the harvesting of their content, legal regulations for the web archiving workflow and the access rules of the preserved material).

In the pilot phase, some hundred cultural and scientific websites are being selected (e. g. homepages of libraries, archives, museums, universities, research institutes, e-journals and blogs). The owners of these websites are being notified to ask permission for the harvesting and archiving of their webpages and to grant public access to these contents through a demo service.

Introduction to metadata in web archiving

The living web is one single document, a huge ever-changing unlimited hypermedia. However, the archived web is a versioned (timed-stamped) file depository. These files can be harvested by several collection methods. Selective method can be focused to a subject or for a special event. Domain-wide harvest is focusing on a certain range of web-domains. Comprehensive harvests are snapshots of a national domain reflects to a certain time period. Web-content can be harvested automatically but can be submitted by a content provider as a deposit copy.

When we would like to generate metadata in a web archive environment, we have a couple of questions that have to be answered. The subject of the description and the types of metadata that are being needed have to be determined and metadata format must be chosen. Another central question is that how metadata can be produced and how can it be used for? User needs, scale of archiving, infrastructure background (e. g. staff, financial resources) can be relevant on this matter.

The major metadata types can be the following: bibliographic (title by lots of variations, creator/contributor/publisher with uncertain roles, rights, possibly by unclear legal status), dates (but what kind of dates?), subject/type – that can be very mixed content), administrative (e. g. curator, nominator, urgency, permission request, harvesting schedule, quality assurance, access etc.), and technical (original content management system, harvester software, harvest parameters, size of the downloaded content, storage, long-term preservation etc.). Another central issue is the level of description: several levels can be chosen, like collection, sub-collection, website, website unit, document, file.

The Hungarian metadata model and its international environment

When we started to build-up our own metadata model, some international set of recommendations could help our job. ISO/TR 14873:2013 – Statistics and quality issues for web archiving (offers collection level indicators). Descriptive Metadata for Web Archiving/OCLC Web Archiving Metadata Working Group (mostly site-level bibliographic data fields, based on the Dublin Core schema with rich mapping options including semantic mapping). Metadata Application Profile for Description of Websites with Archived Versions / New York Art Resources Consortium (offers site-level indicators on MARC/RDA format). Mainly the OCLC recommendations had been used through the creation of our metadata model.

Metadata are being collected on website-level and in sub-collection level in our web archive in XML records. These includes XSD (XML Schema Definition) and XSLT (Extensible Stylesheet Language Transformations) files. More than one hundred XML metadata fields are available including all three major types of metadata in the data model. (figure 2) Metadata can be managed on a quite flexible way. Due to the overwhelming number of the collected websites only a small portion of these sites can be enriched by detailed metadata description. These are available as a part of a small demo collection in public²

² <http://mekosztaly.oszk.hu/mia/demo/>

A	B	C	D	E	F	G	H	I	J	K
Mezőkód	Mezőnév	Címke	XML tag	MARC 21 mező	Minőtlők	Státusz	Adattípus	Ismételhető	Kötött szótár	Megjegyzés, példa
EGYEDI WEBHELYEK										
LÉÍRÓ METAADATOK										
L01	ID	azonosító	mia_id	001	nincs	kötelező (automatizált)	"MIA-" előtag, majd azonosító	nincs	nincs	MIA-000123
L02	Títle	cím	dc_title	245; 246; 730; 740	egységesített cím (+besorolás)	kötelező (legálább)	rovid szöveg	korlátlanul	nincs	A besorolási cím az egységesített cím
L03	URL	URL cím	dc_identifier	856 u almező	eredeti URL, alternatív eredeti URL-je, ...	opcionális	URL	korlátlanul, de az eredeti URL-je	nincs	
L04	Source of Description	a leírás forrása	source_of_desc	588	forrás neve, forrás URL-je, ...	opcionális	rovid szöveg, URL, ...	korlátlanul csoportosan	nincs	
L05	Collector	begyűjtő	collector	583 h almező 6, 710; 852 a	intézmény, szervezeti egység	ajánlott	rovid szöveg	korlátlanul csoportosan	nincs	
L06	Creator	létrehozó	dc_creator	100; 110; 700; 710	intézmény, szervezeti egység	ajánlott	rovid szöveg	korlátlanul	nincs	
L07	Contributor	közreműködő	dc_contributor	700	szerző, közreadó, szerkesztő	opcionális	rovid szöveg	korlátlanul csoportosan	nincs	
L08	Publisher	kiadó	dc_publisher	280 a; 280 b	kiadó neve, kiadó e-mail címe	ajánlott	rovid szöveg, e-mail cím	korlátlanul csoportosan	nincs	
L09	Rights	jogok	dc_rights	508; 540; 542	tartalmi jogtulajdonos neve, jogok	ajánlott	rovid szöveg, e-mail cím	korlátlanul csoportosan	nincs	CC licenck: h
L10	Genre/Form	műfaj/típus	dc_type	655	műfaj, típus, változékony neve	ajánlott	rovid szöveg	korlátlanul	nincs	mindháromhoz
L11	Subject	téma	dc_subject	600; 610; 611; 630; 647	főtéma, téma, altéma, tárgy	ajánlott	rovid szöveg	korlátlanul	nincs	Pl. "honlap", "blog", "periodika", "portál"
L12	Description	tartalmi leírás	dc_description	500; 505; 520; 540; 541	nincs	ajánlott	rovid szöveg	korlátlanul	nincs	Pl. "Kulturális források", "Oktatási források"
L13	Relation	kapcsolat	dc_relation	580; 730; 773; 787	részgyűjtemény azonosító	ajánlott	rovid szöveg	korlátlanul csoportosan	nincs	
L14	Language	nyelv	dc_language	008 35-37 byteok; 041	nyelvkód, megjegyzés	ajánlott	3-betűs kód, rovid szöveg	korlátlanul	nincs	https://www.loc.gov
L15	Date	dátum	dc_date	008 07-14 byteok; 245 f	arítás kezdése, utolsó módosítás	ajánlott	YYYY-MM-DD formátum	korlátlanul	nincs	Pl. "naponta többször", "naponta", "hét naponta"
ADMINISZTRÁCIÓS METAADATOK										
A01	Curator	felelős archíváló munkatárs	curator		munkatárs neve, e-mail címe	ajánlott	rovid szöveg	korlátlanul	nincs	
A02	Nominator	javaslattevő	nominator		javaslattevő neve, javaslattevő e-mail címe	ajánlott	rovid szöveg, e-mail cím	korlátlanul csoportosan	nincs	Részben automatizálható a kitöltése a "javasolt" szótárral
A03	Status	javaslattétel állapota	status		nincs	opcionális	nem	korlátlanul csoportosan	nincs	Pl. "javasolt", "még nem mentett", "még nem engedélyezett"
A04	Urgency	sürgősség	urgency		sürgősség státusza, megjegyzés	opcionális	Y/N logikai érték, rovid szöveg	nem	nincs	
A05	Permission	engedélyezés	permission		engedélyezés státusza, engedélyezés dátuma	opcionális	rovid szöveg, YYYY-MM-DD formátum	korlátlanul csoportosan	nincs	Pl. "nem kértünk engedélyt!", "kértünk engedélyt", "ellenőrzve", "ellenőrzve", "nincs ellenőrzve", "van oldalkép", "nincs oldalkép"
A06	Quality assurance	ellenőrzés	quality_assurance		ellenőrzés státusza, ellenőrzés dátuma	opcionális	rovid szöveg, YYYY-MM-DD formátum	korlátlanul csoportosan	nincs	Pl. "nincs ellenőrzve", "ellenőrzve", "nincs ellenőrzve", "van oldalkép", "nincs oldalkép"
A07	Screenshot	oldalkép	screenshot		oldalkép státusza, oldalkép dátuma	ajánlott	rovid szöveg, YYYY-MM-DD formátum	korlátlanul	nincs	A nyilvános demo gyűjteménybe bekerülő MNB-be való felvitelre javasolt. A nyilvános szolgáltatható, "zárt háttérű"
A08	Demo set	demo	demo		demo státusz	opcionális	Y/N logikai érték	nem	nincs	
A09	National bibliography	MNB-be javasolt	mnb		MNB státusz	opcionális	Y/N logikai érték	nem	nincs	
A10	Access	hozzáférés	access		hozzáférés státusz, hozzáférés dátuma	ajánlott	rovid szöveg, YYYY-MM-DD formátum	nem	nincs	
A11	Administrative note	adminisztrációs megjegyzés	administrative_note		nincs	opcionális	rovid szöveg, YYYY-MM-DD formátum	korlátlanul	nincs	
TECHNIKAI METAADATOK										
T01	CMS	tartalomkezelő rendszer	cms		nincs	ajánlott	rovid szöveg	nem	nincs	Pl. "Drupal", "WordPress", "Dspace", "Moodle"
T02	Harvester	aratószóver	harvester		nincs	ajánlott	rovid szöveg	korlátlanul	nincs	Pl. "Heritrix 3.x", "Heritrix 1.0", "HTTrack"
T03	Harvest parameters	aratószóver paraméterei	harvest_parameters		mélység, ugrások száma, max. URL-ek száma, max. mélység	opcionális	rovid szöveg	korlátlanul	nincs	
T04	Shutdown conditions	leállítási feltételek	shutdown_conditions		max. URL-ek száma, max. mélység	opcionális	rovid szöveg	korlátlanul	nincs	
T05	Config file	konfigurációs fájl	config_file		nincs	opcionális	URL	korlátlanul	nincs	Ha egyedi config fájlra van szükség, akkor a robots.txt leírásba kell beírni a robots.txt fájl nevét.
T06	Robots rule	robots.txt kezelés	robots_rule		nincs	opcionális	Y/N logikai érték	nem	nincs	Annak jelzésére, hogy a robots.txt meg van jelenítve a robots.txt fájlban.
T07	Deduplication	duplikáció	deduplication		duplikáció státusza, megjegyzés	opcionális	Y/N logikai érték, rovid szöveg	nem	nincs	A megjegyzésbe pl. beírható, hogy a duplikáció nem történt meg.
T08	Schedule	mentések ütemezése	schedule		első mentési időpont, ismétlődés	ajánlott	YYYY-MM-DD HH:MM:SS formátum	korlátlanul	nincs	
T09	Harvest report	aratószóver napló	harvest_report		nincs	opcionális	URL	korlátlanul	nincs	Automatizálható a kitöltése a Heritrix 3.x szótárral.
T10	Extent	mentések jellemzői	extent	300; 347	mentések száma, webhely neve	opcionális	számok, dátum	korlátlanul csoportosan	nincs	Részben automatizálható a kitöltése a "mentések" szótárral.
T11	Archive format	archívformátum	archive_format		nincs	ajánlott	rovid szöveg	korlátlanul	nincs	Részben automatizálható a kitöltése a "mentések" szótárral.
T12	Storage	raktári hely	storage		nincs	ajánlott	rovid szöveg	korlátlanul	nincs	Részben automatizálható a kitöltése a "mentések" szótárral.
T13	Long-term preservation	hosszú távú megőrzés	long-term_preservation		nincs	opcionális	rovid szöveg	korlátlanul	nincs	Részben automatizálható a kitöltése a "mentések" szótárral.
T14	Technical note	technikai megjegyzés	technical_note		nincs	opcionális	rovid szöveg	korlátlanul	nincs	Részben automatizálható a kitöltése a "mentések" szótárral.

Figure 2 Database model on website level

Azonosító	A webhely neve	URL címe	OSZK mentés	Képernyőfotó	Linktérkép	IA mentés	Eredeti	Metaadat
MIA-000034	Berzsenyi Dániel Városi Könyvtár, Marcali	www.marcalikonyvtar.hu	●	●	●	●	●	●
MIA-000029	Csuka Zoltán Városi Könyvtár, Érd	www.csukalib.hu	●	●	●	●	●	●
MIA-000106	Evangélikus Országos Könyvtár	konyvtar.lutheran.hu	●	●	●	●	●	●
MIA-000006	Esztergomi Főszékesegyházi Könyvtár	www.bibliothea.hu	●	●	●	●	●	●
MIA-000007	EX LIBRIS Könyvtár	exlibriskonyvtar.hu	●	●	●	●	●	●
MIA-000032	Fejér György Városi Könyvtár, Keszthely	www.fgyvk.hu	●	●	●	●	●	●
MIA-000030	Helischer József Városi Könyvtár, Esztergom	www.vkesztergom.hu	●	●	●	●	●	●
MIA-000035	Jókai Mór Városi Könyvtár, Pápa	www.jmvk.papa.hu	●	●	●	●	●	●
MIA-000026	Keresztély Gyula Városi Könyvtár, Bátaszék	www.kgyvk.hu	●	●	●	●	●	●
MIA-000129	Dr. Kovács Pál Megyei Könyvtár és Közösségi Tér, Győr	www.gyorikonyvtar.hu	●	●	●	●	●	●
MIA-000125	MaNDA DB	mandadb.hu	●	●	●	●	●	●
MIA-000033	Martonsi Pál Városi Könyvtár, Kiskunhalas	www.vkhalas.hu	●	●	●	●	●	●
MIA-000014	Nagy Gáspár Városi Könyvtár, Budakeszi	ngvk.hu	●	●	●	●	●	●
MIA-000031	Széchényi István Egyetem Egyetemi Könyvtár, Győr	lib.sze.hu	●	●	●	●	●	●

Azonosító	A webhely neve	URL címe	OSZK mentés	Képernyőfotó	Linktérkép	IA mentés	Eredeti	Metaadat
MIA-000065	Baranya Megyei Levéltár	www.bml.hu	●	●	●	●	●	●
MIA-000012	Budapest Főváros Levéltára	bparchiv.hu	●	●	●	●	●	●

Figure 3 Hungarian Internet Archive Pilot Project Demo Portal

(figure 3). Some basic fields can be filled out automatically by our plans. In this way at least a very reduced level of description will be offered by each website that will be archived through the ordinary workflow in the future. Each set of metadata description on site or sub-collection level can be enriched further anytime in the future. Currently during the pilot project XML Notepad software is being used to record metadata. Predefined lists are being used through the metadata recording process (e. g. genre, type, topic, subtopic, change frequency, harvest frequency, quality level). Namespace links can be also added (person names, geographic names). Related sites can be recorded (both from the living web and from the archive). (figure 4). The metadata descriptions of the demo collection are being displayed in a converted html format (figure 5).

Figure 4 MIA XSD file

Figure 5 The description of Óbuda Múzeum's blog in original xml and converted html format

Full-text metadata search and retrieval tools

A main challenge of the web archives is the lack of full-text search capability (e. g. because of the sheer size of the archive, or because GDPR considerations). Another major challenge is the lack of relevance, to get relevant hits from a huge set of archived webpages full of partially duplicated content and „garbage“. The international web archiving community has to accept that such a knowledge base, that Google has, is out of question for their needs (Google uses currently over 200 ranking factors³: geo targeting, user browsing and search history, freshness of the content, number of organic clicks, artificial intelligence etc.). Web archives however can develop metadata supported full-text search services and semantic web based query tools and services. The current chapter is focusing on the former and the next chapter is with the latter.

On the SolrWayback Search interface a full-text search function is available for the whole content of the harvested websites in the small demo collection. Query results can be filtered by domain name, file type, crawl year, status code and public suffix. (figure 6) From the set of results, the harvested and live version of a certain website is also available. An advanced interface of metadata is available by clicking on the full post button. (figure 7). SolrWayback Search has developed in collaboration with the Danish National Library.

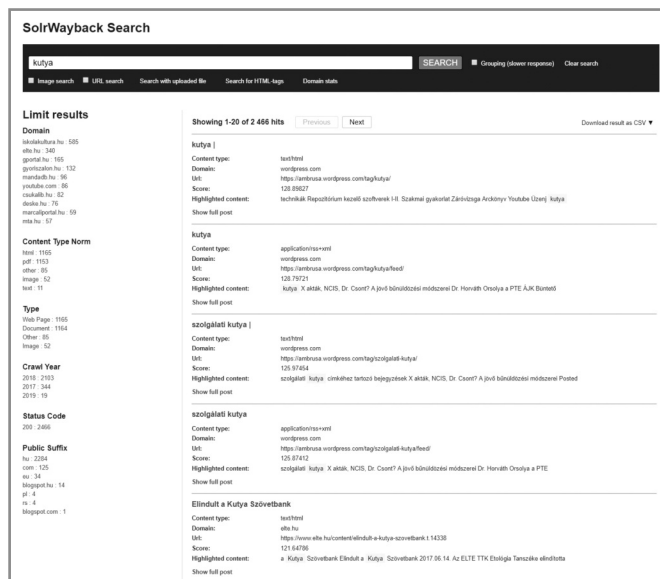


Figure 6 SolrWayback interface

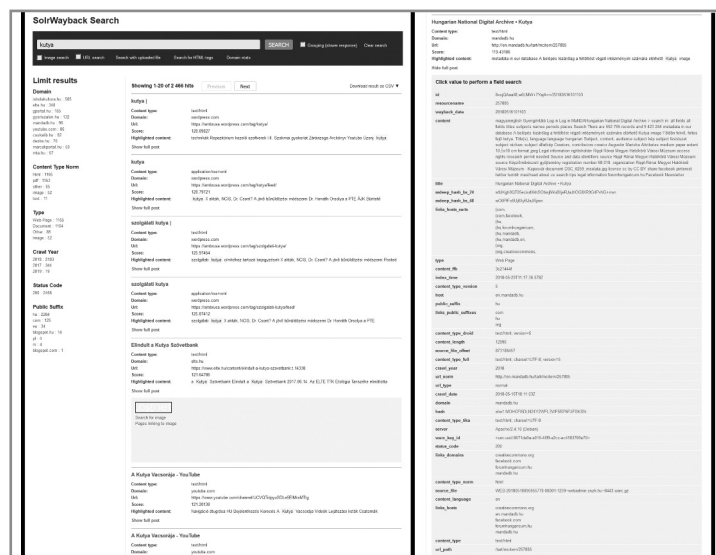


Figure 7 SolrWayback advanced metadata set

³ <https://backlinko.com/google-ranking-factors>

The in-house developed (SolrMIA a Solr based full-text search engine) offers even more advanced filtering functions by topic, subtopic genre and type. Filtering options are being generated from the XML metadata records. The unified name of the original website is also being displayed (metadata origin also the XML record). (figure 8).

The screenshot displays the SolrMIA search interface. At the top, there is a search bar with the text 'kutya' and a 'Keresés' button. Below the search bar, the title 'KERESÉS A DEMÓ WEBARCHÍVUMBAN' is visible. On the left side, there are several filtering options categorized by domain, file type, year, topic, and type. A box highlights these filtering options with the text 'Filtering options are being generated from the XML metadata records'. The main search results area shows a list of items, each with a thumbnail image of a dog and a title in blue text. A box highlights the unified name of the original website (in blue in white) with the text 'The unified name of the original web site (blue in white) comes from the XML metadata record'. The results include items like '(kutya-thumb.jpg)', '(kutya-alvas2-thumb.jpeg)', '(alvo-kutya-thumb.jpg)', '(kutya-egeszsegvedelme-thumb.jpg)', and '(kutya-flickr-thumb.jpg)'. Below the results, there is a section for 'Hungarian National Digital Archive • Reklámfotó - Szőke hajly hölgy kutyával' and a section for 'Ki az ijesztőbb a kutya számára: egy férfi vagy egy nő?'.

Figure 8 SolrMIA interface and main functions

Future plans in metadata management

The metadata collection from the current database will have to be integrated to the central database of the prospective national digital library system. A form-based data entry interface has to be developed on this framework. Cooperation forms have to be established with partner institutions from the cultural heritage sector, especially shared cataloguing services needs to be established (based on the above mentioned national platform). An essential national library task is to incorporate metadata of important archived websites to the national bibliography. We would like to develop further the SolrMIA full-text search engine with additional faceted full-text hit lists filtered by metadata.

Web archives and the semantic web

The absence of efficient and meaningful exploration methods of the archived content is a really major hurdle in the way to turn web archives to a usable and useful information resource. A major challenge in information science can be the adaptation of semantic web tools and methods to web archive environments. The web archives must be a part of the linked data universe with advanced query and integration capabilities, must be able to directly exploitable by other systems and tools.

Extraction of main content and metadata is the initial step. Automatic entity identification and extraction from the full text can be a major task for the future in order to build-up linked data compatible datasets (e. g. names, events, concepts). Then generation of RDF triples is also an essential task. Another major task to the future that a linked-data compatible collection's metadata can be enriched from external resources (e. g. DBpedia). The semantic dataset must be published as linked open data. Advanced queries and ranking models can be set based on semantic data.⁴

⁴ See the details: Pavlos Fafalios et al.: Building and querying semantic layers for web archives (extended version). International Journal on Digital Libraries. 05 July 2018. <https://doi.org/10.1007/s00799-018-0251-0>

In the future for the Hungarian Web Archive we would like to develop automatic entity identification and extraction services, enriching metadata from external resources. We would like to combine our prospective dataset with the national namespace⁵ of the National Széchényi Library (still in beta phase). Other namespaces and thesauri can also be used for metadata enrichment.

Conclusion

Metadata management is a core area of the web archiving workflow. The financial and staff background, the availability of appropriate level of IT-support of a web archiving project is determining the major limits of metadata management activities. Throughout Europe web archiving projects has limited capabilities in this context. The general limits of human and financial resources means that a major future task is to automatize as much working process as it is possible. However, metadata management will require a significant amount of human job in the foreseeable future. It would be really important to support the effectiveness of these human metadata management activities with proper working models and data entry interfaces.

References

International Internet Preservation Consortium (IIPC): <http://www.netpreserve.org>

Hungarian Internet Archive Pilot Project Demo Portal: <http://mekosztaly.oszk.hu/mia/demo/>

Google Ranking Factors: <https://backlinko.com/google-ranking-factors>

Fafalios, P. et al.: Building and querying semantic layers for web archives (extended version). International Journal on Digital Libraries. 05 July 2018. <https://doi.org/10.1007/s00799-018-0251-0>

Drótos, L. & Németh, M. (2017). Hungarian web archiving pilot project in the National Széchényi Library. In 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000209–000212). IEEE. <https://doi.org/10.1109/CogInfoCom.2017.8268244>

Drótos, L., & Németh, M. (2018). Web museum, web library, web archive The responsibility of public collections to preserve digital culture. In L. Petrovska, B. Īvāne-Kronberga, & Z. Meldere (Eds.), The Power of Reading: Proceedings of the XXVI Bobcatsss Symposium, Riga, Latvia, January 2018 (pp. 124–126). Riga: The University of Latvia Press. http://bobcatsss2018.lu.lv/files/2018/08/BOBCATSSS_2018_TheProceedings.pdf

Drótos L. & Németh, M. (2018.10. 23.). How to catalogue a web archive? Information Interactions (INFINT) 2018 Conference. Department of Library and Information Science, Faculty of Arts, Comenius University & University Library, Bratislava. http://mekosztaly.oszk.hu/mia/doc/How_to_catalogue_a_web_archive.pptx

Drótos L. & Németh, M. (2018.11. 08.). The Education of Web-Archiving. CDA 2018 Conference, Bratislava. http://mekosztaly.oszk.hu/mia/doc/The_education_of_web-archiving_CDA_2018.pptx

Németh M. & Drótos, L. (2019.01. 24.). Web archives as a research subject? BOBCATSSS 2019. Department of Information Sciences, Faculty of Humanities and Social Sciences at Osijek University, Croatia. http://mekosztaly.oszk.hu/mia/doc/Web_archives_as_a_research_subject-Bobcatsss2019.pptx

Márton Németh

nemeth.marton@oszk.hu

László Drótos

drotos.laszlo@oszk.hu

(National Széchényi Library Department of E-Library Services)

⁵ Current beta namespace interface: <http://www.abcd.hu>