Webové stránky a webový archív

Archivácia kultúrneho bohatstva každého národa pre budúce generácie je významnou činnosťou kultúrnych inštitúcií. Za archiváciu sú zodpovedné predovšetkým múzeá a depozitné knižnice. Za kultúrne bohatstvo považujeme rôzne druhy hmotných predmetov, ale aj množstvo rukopisných a tlačených dokumentov.
V poslednom období sa však rozširuje množstvo nových typov dokumentov, ktoré sa vyskytujú a ich počet začína narastať, a to predovšetkým rôznych typov elektronických dokumentov.
Táto forma kultúrneho bohatstva a dedičstva má však v niektorých prípadoch veľmi krátkodobé trvanie. Dynamický rozvoj internetu spôsobuje, že obsah množstva stránok sa často presúva, mení sa, mnoho razy úplne zanikne. Životnosť niektorých dokumentov na jednotlivých doménach je niekedy veľmi krátka. Počet webových stránok stále narastá, ale ich životnosť je veľmi rôznorodá. Informácie o ich počte, ako aj ich registrátoroch možno nájsť na www.sk-nic.sk. Aktuálny počet zaregistrovaných slovenských domén bol ku 17.8.2009 - 191 377, počet registrátorov bol 3 402.
Archivácia webu je teda odpoveďou na tieto rýchlo sa meniace trendy. Cieľom je zachytiť obsah webových stránok v priebehu času a ich uchovanie pre budúce generácie.

Základné činnosti pri budovaní webového archívu by sa mohli rozdeliť do troch základných fáz:

- výber webových stránok
- uloženie webových stránok do archívu
- zaistenie ich sprístupňovania

Výber webových stránok a ich uloženie do archívu by bolo možné dokumentovať prostredníctvom vybraných stratégií pri tzv. zbere webu. Po týmto pojmom možno rozumieť vytvorenie kópií webových stránok a ich uloženie do digitálneho úložného priestoru archívu. Zbery môžu byť podľa dohody rôzne, predovšetkým celoplošné – komplexné alebo výberové podľa dohodnutých kritérií či už časových, tematických alebo ďalších dohodnutých kritérií, prípadne aj určité špecifické výbery podľa tém, oblastí, osobností, regiónov atď.
Tu je potrebné stanoviť určitú typológiu jednotlivých webových stránok, ktoré sa môžu vyskytovať, rozpoznať a určiť hlavné webové stránky, ktoré je potrebné zachytiť a stanoviť:

- ktoré sa budú ukladať
- ktoré sa budú ukladať a zároveň analyzovať, t.j. tvoriť metadáta

K základným zaujímavým webovým stránkam patria webové stránky:

- ústredných orgánov a inštitúcií
- regionálnych orgánov a inštitúcií
- firiem všetkých druhov a typov, podľa odborov
- školy – všetky druhy
- zdravotníctvo
- kultúra
- regióny

- rôzne organizácie – tretí sektor

Na vyhľadávanie existujú rôzne medzinárodné softvéry, ktoré je možné používať, napr. aj v rámci spolupráce s medzinárodným združením IIPC (International Internet Preservation Consortium).

Na webe možno nájsť rôzne typy elektronických dokumentov:

- digitálne dokumenty voľne prístupné prostredníctvom Internetu
- publikácie odborného, umeleckého, spravodajsko-publicistického charakteru
- periodiká, monografie, konferenčné príspevky, výskumné a iné správy, akademické práce, zborníky atď.
- textové , ale aj obrazové a zvukové dokumenty existujúce len v digitálnej forme

Ďalšou dôležitou oblasťou činnosti je sprístupňovanie webových stránok širokej verejnosti. Zmyslom je vlastne poskytnúť verejnosti komplexný pohľad na vývoj, obsah, druhy a podoby jednotlivých webových stránok v určitom časovom úseku. Sprístupnenie môže byť komplexné alebo obmedzené.
Existuje totiž množstvo dokumentov, ktoré nemôžu byť prístupné v súčasnosti všetkým používateľom, ale kvôli autorským právam je sprístupňovanie obmedzené.
V tejto oblasti sa stále naráža na množstvo obmedzení legislatívneho charakteru, ktoré negatívne pôsobia na realizáciu dlhodobého cieľa – vybudovanie otvoreného archívu webu. Knižnice majú určité práva získavať a archivovať dokumenty (predovšetkým Slovenská národná knižnica), ale problém je v tom, že môžu poskytovať informácie, ale niektoré iba v priestoroch knižnice, čiže nemôžu byť verejne prístupné všetky informácie, dokumenty a v rámci nich aj webové stránky. V mnohých prípadoch je potrebný súhlas príslušných držiteľov autorských práv.
Tento problém je potrebné riešiť generálnym spôsobom. Zatiaľ by to bolo možné na základe jednotlivých dohôd s príslušnými vydavateľmi resp. registrátormi webových stránok príp. je to možné riešiť aj prostredníctvom medzinárodných licencií Creative Commons, ktoré by bolo potrebné zosúladiť s našim právnym poriadkom, preložiť a sprístupniť v slovenskom jazyku.
Problém je aj v zákone o povinnom výtlačku, ešte stále v ňom nie sú zakomponované elektronické dokumenty, t.j. chýbajú aj webové stránky.
Tu existuje snaha SNK o novelizáciu tohto zákona, ale zatiaľ sa ju nepodarilo presadiť.
V mnohých krajinách sa už tento zákon podarilo presadiť – napr. Nórsko (1990), Chorvátsko (1997), Island (2003), Nová Zéland (2003), Kanada (2004), Dánsko (2005), Slovinsko (2006), Francúzsko (2007), Fínsko (2008). O novelizáciu tohto zákona sa snažia aj v Českej republike.
K základným úlohám Slovenskej národnej knižnica totiž patrí aj ochrana a uchovávanie dokumentov vydaných v Slovenskej republike. V rámci tejto úlohy spracováva, ochraňuje, sprístupňuje a trvale uchováva konzervačný fond a registruje ho v Slovenskej národnej bibliografii. Popri klasických dokumentoch je potrebné v súčasnosti pokrývať tieto úlohy aj v oblasti elektronických dokumentov.
Kritériá výberu by bolo možné rozdeliť do dvoch skupín podľa spôsobu získavania webových zdrojov do archívu resp. podľa legislatívnych podmienok poskytovania prístupu k archivovaným údajom:

- plošná archivácia webových zdrojov
- výberový prístup – zber zdrojov, na ktoré dal vydavateľ súhlas v rámci zmluvy.

Cieľom plošnej archivácie je zachytenie čo najväčšieho počtu domácich webových zdrojov na základe určitých definovaných parametrov v aplikovanom softvéri, ktoré vyžadujú minimum ďalšej intelektuálnej práce.
Základnou podmienkou pre plošnú archiváciu je národná doména, ktorých v auguste 2009 bolo registrovaných v SR niečo pod 200 000. Môžu byť archivované zdroje aj mimo domény, ale musia spĺňať určité kritériá.
Zdroje môžu byť vyberané aj na základe ďalších aspektov, ale nemusia byť povinné, môžu byť stanovené na základe ďalších dohôd, napr. formáty, prístupy, protokoly, veľkosti súborov, počty súborov atď.
Pravidlá výberu dokumentov registrovaných v národnej bibliografii zahŕňajú princípy:

- územné slovaciká
- autorské slovaciká
- jazykové slovaciká
- obsahové slovaciká

V prvom rade musí byť vybratá národná doména alebo ďalšie domény, ktoré spĺňajú výberové kritériá.
Musia byť zabezpečené už spomenuté národné aspekty – územné...
Zdroje musia mať významnú kultúrnu , vedeckú hodnotu, aktuálny obsah atď.
Zdroje musia byť voľne prístupné alebo na základe dohody s vydavateľom.
Preferované sú zdroje, ktoré sú originálnym elektronickým dokumentom, nie kópie diel vydaných iným spôsobom, t.j. webový dokument je originál.
Preferované sú online časopisy, monografie, konferenčné materiály, výskumné a iné správy, akademické práce, vládne dokumenty, predovšetkým také, ktoré majú mimoriadnu kultúrnu, vedeckú hodnotu.
Je potrebné stanoviť aj zdroje, ktoré sa jednoznačne nebudú archivovať, napr. počítačové hry, intranety, , osobné webové blogy, databázy, rozhlasové a televízne vysielania atď.
Do úvahy je potrebné brať:

- prenosový protokol
- autorské práva
- integritu zdrojov
- frekvenciu zberov – tú je potrebné stanoviť podľa možností a potrieb, optimálna by bola denne, ale to nie je realizovateľné, čiže je potrebné túto otázku riešiť podľa technologických možností.

Pre sprístupňovanie týchto obrovských súborov je potrebné zabezpečiť určité technologické podmienky, predovšetkým zodpovedajúci server.
Vo všetkých oblastiach musia byť dôsledne aplikované medzinárodné štandardy – popis online zdrojov podľa MARC 21, Dublin Core, XML, pre identifikáciu medzinárodné identifikátory (ISSN, atď.) a pre archiváciu príslušný vybratý formát.

6.9.2009
AK