

Technológie pre spracovanie a sprístupňovanie dát v knižniciach: od katalogizačných lístkov k dátam vo virtuálnom svete

Nadežda Andrejčíková

andrejcikova@cosmotron.sk

Čas je neúprosný a dnes, keď sa všetko okolo nás neustále mení, sa nám môže zdať, že aj ten čas plynie akosi rýchlejšie. Nemôžeme poprieť skutočnosť, že prostredie, v ktorom žijeme, stále viac ovplyvňuje práve technológie a tiež to, že azda nič nestarne tak rýchlo ako technológie samotné. No a v tom je práve problém, pretože sa musíme neustále prispôbovať a učiť sa používať nové technológie, ale nielen to – ich zavádzanie si častokrát vyžaduje aj zmenu prístupu a pracovných zvyklostí, čo je ešte väčší problém, ako sa naučiť používať nové prostriedky. Celý svet sa mení a riaditeľ Mestskej knižnice v Prahe, pán Tomáš Rehák, raz povedal: „Ten, kto chce prežiť, musí sa zmeniť, pretože sú len dve skupiny – tí, čo sa zmenia a tí, čo zaniknú.“

Cieľom tohto článku je ponúknuť nový pohľad v kontexte technológií, a zároveň objasniť, čo predstavujú z pohľadu spracovania informácií v knižniciach.

Na začiatku bol katalogizačný lístok, ktorý obsahoval bibliografický záznam a mal svoje špecifické rozmery, aby sa mohli unifikovať kartotéky, ktoré slúžili pre vyhľadávanie v týchto katalogizačných lístkoch. Pre vyhľadávanie podľa viacerých hľadísk budovali knihovníci súčasne viacero typov katalógov. Tým najhlavnejším, ktorý ešte aj dnes môžeme v niektorých knižniciach nájsť, bol generálny katalóg, ktorého úlohou bolo uchovávať prehľad o celom fonde knižnice. Čitateľom slúžili katalógy rozdelené na autorský, názvový, predmetový, atď. Rozdiel medzi týmito katalógmi bol v tom, že katalogizačné lístky v nich boli radené podľa daného ukazovateľa. Ak nerátame písací stroj – aj keď elektrický – a ani kopirovacie stroje, ktoré slúžili na rozmnožovanie týchto lístkov, aby sa mohli zakladať do jednotlivých katalógov, tak o technológiách hovoríme až v období nástupu automatizácie. Najprv to bolo spracovanie bibliografických dát na veľkých sálových počítačoch, ktoré neskôr nahradili osobné počítače a tiež počítačové siete. Aj keď sa katalogizačné lístky naďalej tlačili a zakladali do kartoték, tak vyhľadávanie bolo možné už aj prostredníctvom počítačov. Tie pomerne jednoduchým spôsobom ponúkli oveľa širšie možnosti vyhľadávania. Keď katalogizačné lístky ponúkali vyhľadávanie napr. len podľa prvých troch autorov, či troch kľúčových slov, v počítačoch bolo možné vyhľadávať podľa ľubovoľného počtu. Výhody elektronickej formy dát si knihovníci rýchlo uvedomili a hľadali spôsoby pre efektívny spôsob komunikácie, výmeny a vzájomného zdieľania dát, pretože tušili, že už sú len krôčik od toho, aby mohlo byť vyvrátené tvrdenie „*At the present time, if a specially valuable book is published it finds its way to at least a thousand different libraries, in all of which it must be catalogued. ... Can librarians complain if practical business men call this sheer extravagance?*“ (Melvil Dewey, 1877)

V tom čase si asi len málokto uvedomoval, že prvú štruktúrovanú formu bibliografických dát – aj keď len v papierovej podobe – predstavovali už katalogizačné lístky, resp. to boli pravidlá ISBD, ktoré túto štruktúru prostredníctvom interpunkčných znamienok definovali. Avšak za prvý všeobecne známy formát pre výmenu a zdieľanie dát v počítačovom svete je označovaný formát MARC, čo predstavuje akronym pre Machine Readable Cataloging, zo začiatku tiež označovaný aj ako LCMARC, ktorý bol vydaný a ďalej aj spravovaný knižnicou Library of Congress. (Avram, 2003). Pre potreby vzájomnej spolupráce a pre možnosť efektívneho využívania týchto dát bolo však potrebné vyriešiť tiež otázky komunikácie. No a boli to knihovníci, ktorí definovali protokol pre komunikáciu informačných systémov všeobecne, bez nutnosti vedieť podrobnosti o tom, na akej platforme či nad akým databázovým systémom a s akými indexami ten druhý informačný systém pracuje. Tento protokol je známy pod označením Z39.50. Možnosti, rozsah a význam tohto protokolu bol štandardizovaný a v roku 1995 prijatý ako ANSI/NISO norma a v roku 1998 aj ako ISO 23950:1998, Information and documentation – Information retrieval (Z39.50) – Application service definition and protocol specification. (/). Charakteristické pre tento protokol je, že ide o protokol typu klient – server, ktorý danému informačnému systému umožňuje vyhľadávať a získavať údaje z iného informačného systému vlastnými prostriedkami. Neskôr bola funkcionálnosť tohto protokolu rozšírená aj o možnosti tvorby a modifikácie záznamov. Keďže Z39.50 je všeobecným protokolom, efektívnosť a správnosť jeho využitia závisí od toho, aby klient aj server používali v komunikácii rovnakú množinu atribútov. Implementátori protokolu Z39.50 v prostredí knižníc, ktorí tvorili v jednotlivých krajinách samostatné skupiny, pracujúce pod označením ZIG (Z39.50 Implementor's Group), sa na nevyhnutnosti definovať takýto profil – spoločnú množinu atribútov pre všetky funkčné oblasti vyhľadávania a objavovania informačných zdrojov na medzinárodnej úrovni – dohodli na svojom stretnutí v meste Bath v auguste 1999. Ešte v tomto roku bola publikovaná prvá verzia s označením Bath Profile – BIB1 attribute set, ktorá sa postupne upravuje a rozširuje podľa nových požiadaviek a skúseností. Čitatelia ocenili silu a význam tohto protokolu vtedy, keď vyhľadávanie v jednotlivých kartotékach nahradili počítače.

Knižnično-informačné systémy, či skôr ich časť – modul OPAC – Online Public Access Catalog – prostredníctvom tohto protokolu začali ponúkať rovnakú možnosť vyhľadávania nad katalógom a informačnými zdrojmi danej knižnice, ako aj nad katalógmi iných knižníc. Dáta v elektronickej podobe tak na jednej strane ponúkli jednoduchou formou širšie možnosti vyhľadávania, na druhej strane však pomerne rýchlo odhalili pre knihovníkov aj nové výzvy. Ukázalo sa totiž, že systémy pri vyhľadávaní v katalógu, napr. podľa autora, nevrátia vždy všetky jeho záznamy, ktoré sú v katalógu spracované. Dôvodom je subjektívny prístup spracovateľa, pretože pri počítačoch platí, že ak je čo i len jeden znak v danom reťazci iný, tak pre počítač to je úplne iný reťazec. Preto sa ponúka otázka, ako zabezpečiť, aby každý katalogizátor zadal meno toho istého autora rovnako, alebo pre vyjadrenie témy, ktorá je kľúčová u daného dokumentu, použil rovnaký výraz? Druhá výzva súvisí so samotným formátom, ktorý si knihovníci definovali s nástupom automatizácie pre výmenu a zdieľanie dát. Pýtate sa prečo, keď si ho jasne definovali? Na rozdiel od prvej otázky, odpoveď na túto ponúkam hneď. Áno, knihovníci boli asi prví, ktorí sa na medzinárodnej úrovni dohodli na spoločnom formáte pre výmenu a zdieľanie dát, ale tento formát bol prvým spoločným formátom, kde ešte neboli vychytené všetky problémy. Nezhľadňoval totiž všetky možné špecifiká, ktoré boli typické napr. len

pre určitú krajinu, či komunitu knihovníkov, a tak vznikli postupne rôzne verzie MARC-u, ktoré sa navyiac menili v čase, a tak komplikovali preberanie a zdieľanie bibliografických záznamov rovnako, ako aj budovanie súborných katalógov na medzinárodnej úrovni.

Knihovníci tak mali pred sebou novú výzvu, a to unifikovať formát dát. Tým nemám na mysli len formát ako taký, teda z hľadiska jeho rozsahu polí a podpolí či indikátorov, ktoré určovali to, čím je možné jednotlivé dokumenty charakterizovať a prostredníctvom čoho ich je možné identifikovať, ale aj jeho samotný obsah. Teda, unifikovať to, ako a presnejšie akými hodnotami – dátami tieto jednotlivé polia a podpolia pre konkrétny dokument vyplniť. Jednoducho povedané, ako zabezpečiť to, čo zodpovie prvú otázku, teda aby katalogizátor vedel, aký termín zvolí pre identifikáciu napr. konkrétneho autora, a to taký, ktorý bude zhodný s termínom, ktorý použili, či použijú pre daného autora katalogizátori aj v iných knižniciach. Preto sa medzinárodná federácia knižničných asociácií a knižnic rozhodla hľadať nové riešenie. Formát, ktorý bude vyhovovať požiadavkám a špecifikám všetkých krajín, a zároveň riešiť aj zjednotenie termínov pre vybrané prvky. Riešenie, ktoré bude spĺňať požiadavky interoperability, čo zároveň zabezpečí možnosť budovať súborné katalógy aj discovery systémy na medzinárodnej úrovni. Významnú zmenu v spracovaní priniesol v roku 1977 návrh nového medzinárodného formátu – UNIMARC. Dovolím si tvrdiť, že zmeny, ktoré do procesu spracovania dokumentov a tvorby bibliografických záznamov tento formát prináša, majú revolučný charakter. Je to totiž prvýkrát, čo formát definuje nie jednu, ale v prvopočiatku hneď dve, a to ešte vzájomne prepojené metadátové schémy. Tou prvou je schéma pre bibliografické záznamy, ako tomu bolo aj vo formáte MARC a druhou schéma pre Autority. Je to vôbec prvýkrát, kedy sa aj tento pojem – „Authority“ – v spracovaní dokumentov objavil. Služí pre budovanie súborov autorít, ale záznam autority tu nepredstavuje záznam o osobe, ktorá je akýmsi vzorom pre spracovateľa, ale autority sú chápané z iného pohľadu. Hana Vodičková ich v svojej práci definovala ako „súbor otvorených a unifikovaných selekčných údajov menných a vecných, určených pre spracovanie a vyhľadávanie dokumentov a informácií s nevyhnutným odkazovým a poznámkovým aparátom“, pričom ja k tejto definícii ešte dodávam: „ktoré sú identifikované prostredníctvom trvalého, unikátneho a tiež jedno-jednoznačného identifikátora.“ (Vodičková, 1999)

Viem, znie to príliš komplikovane, ale je to jednoduché. Predstavme si napríklad to, že každý autor má pridelený svoj jedno-jednoznačný, unikátny a ešte k tomu trvalý identifikátor, pod ktorým sú uvedené a vzájomne prepojené všetky rôzne varianty mena tohto autora, vrátane prepojenia na mená a záznamy iných typov autorít, ktoré s ním úzko súvisia. Hovorí sa, že jeden príklad je viac ako niekoľko strán textu. Takže sa to pokúsím vysvetliť na príklade. Máme autora – Martin Kukučín, či skôr s ohľadom na pravidlá – Kukučín, Martin, ktorý inde môže byť uvedený pod menom Bencúr, Matej, alebo Bencúr-Kukučín, Martin, či Bencúr, Matthias, atď. Keďže to, akú formu mena zvolí pri popise diela katalogizátor, môže byť ovplyvnené jeho subjektívnym prístupom, prichádza formát UNIMARC Authority s tým, že sa vytvorí pre tohto autora záznam autority, identifikovaný unikátnym a trvalým jedno-jednoznačným identifikátorom, ktorý sa pre tohto autora použije všade tam, kde chcem vyjadriť nejaký súvis s ním. Podľa pravidiel spracovania bude teda vytvorený samostatný záznam autority s vlastným identifikátorom, kde bude za hlavnú zvolená jedna z uvedených foriem mena, pričom ostatné formy mena budú priamo napojené na tento identifikátor, takže pri vyhľadávaní podľa ľubovoľnej z uvedených foriem mena systém vždy odkáže na jeden a ten istý identifikátor. Záznam autority bude taktiež obsahovať ďalšie údaje o autorovi, minimálne jeho krátku biografickú charakteristiku, ale aj dátum narodenia, úmrtia, miesto narodenia či úmrtia, pričom pre toto miesto bude vytvorený samostatný záznam autority s vlastným identifikátorom a tento bude použitý k tomu, aby bolo zabezpečené prepojenie na tento záznam. Schematicky je to vyobrazené na obr. č. 1.

Pre uvedený príklad Martina Kukučina môžeme v Národných autoritách v Českej republike nájsť záznam s identifikátorom „jn19990218058“. Národné autority v Českej republike som nevybrala náhodne, ale sú krásnym príkladom toho, ako funguje

Kukučín, Martin, 1860-1928

Kukučín, Martin

Martin Kukučín

Kukučín, Martin, pseud. di Matej Bencúr, 1860-1928

VIAF ID: 10637721 (Personal)
 Permalink: <http://viaf.org/viaf/10637721>
 ISNI: 0000 0001 2095 5849

Preferred Forms

- 100 1 _ \$a Kukučín, Martin, pseud. di Matej Bencúr, pseud. di 1860-1928
- 100 1 _ \$a Kukučín, Martin
- 100 1 _ \$a Kukučín, Martin (undifferentiated) (sparse)
- 100 1 _ \$a Kukučín, Martin pseud. di 1860-1928
- 100 1 _ \$a Kukučín, Martin pseud. di 1860-1928
- 100 1 _ \$a Kukučín, Martin pseud. di 1860-1928
- 100 1 _ \$a Kukučín, Martin
- 100 1 _ \$a Kukučín, Martin
- 100 1 _ \$a Kukučín, Martin pseud. di (1860-1928)
- 100 1 _ \$a Kukučín, Martin pseud. di (1860-1928)
- 100 1 _ \$a Kukučín, Martin pseud. di 1860-1928
- 100 0 _ \$a Martin Kukučín

National Library of the Czech Republic
 NKC-jn19990218058

Obr. 1

spomenutý protokol Z39.50 s využitím funkcií pre tvorbu a modifikáciu záznamov v inom informačnom systéme prostriedkami vlastného informačného systému. S formátom UNIMARC v našich podmienkach začali prvé kroky smerujúce k štandardizácii v procese spracovania a sprístupňovania bibliografických záznamov, a teda vytváraním interpretácií k jednotlivým medzinárodným normám, pravidiel a iným štandardom a odporúčaniam. UNIMARC, rovnako ako spracovávanie súborov autorít, boli pre knihovníkov úplnou novinkou, s ktorou sa postupne zoznámali. Najťažšie bolo uvedomiť si fakt, že autority predstavujú akoby samostatný register a pre zdieľanie katalogizačného záznamu je potrebné mať aj všetky súvisiace záznamy autorít. Taktiež sa zakrátko ukázalo, že aj keď jednotlivé knižnice pracujú v UNIMARCU, používajú k tomu rôzne knižnično-informačné systémy, bolo nutné nájsť spôsob a presný pracovný postup, ako budovať centrálnu databázu autorít a mať tak pre každý termín centrálny, unikátny, jedno-jednoznačný, trvalý identifikátor. Tak v roku 2000 uviedli v Národnej knižnici do prevádzky systém pre kooperatívnu tvorbu Národných autorít v heterogénnom prostredí on-line, ktoré sa úspešne využíva aj dnes. Toto riešenie je založené na vyššie spomínanom protokole Z39.50 s využitím funkcie pre tvorbu a modifikáciu záznamov v externom informačnom systéme prostriedkami vlastného systému. To je krásny príklad využitia štandardov v praxi. Národná knižnica ČR sa po zavedení kooperatívnej tvorby Národných autorít stala aj garantom a spolupracovníkom za ČR na budovaní medzinárodného systému autorít, ktorý jednoducho mapuje identifikátory, používané pre danú autoritu v jednotlivých krajinách medzi sebou. Toto riešenie je dostupné na stránke <http://viaf.org/> a je v ňom krásne vidieť prepojenie na záznam autority z nášho príkladu vyššie.

Neustále rastúca potreba jednotnej a jednoznačnej identifikácie na medzinárodnej úrovni dospela k tomu, že v marci 2012 bola publikovaná norma ISO 27729:2012 International standard name identifier (ISNI), ktorej základom sa stal VIAF a jej garantom, rovnako, ako aj garantom VIAF.org je OCLC. Dostupná je na stránke <http://isni.org/>, kde je možné získať aj aktuálne informácie o počte spolupracujúcich organizácií či počte jednotlivých typov záznamov, tak ako aj odkazy na pravidlá, štandardy, iné súvisiace materiály, ap.

V katalógu kníh ČR môžeme zase nájsť bibliografický záznam, reprezentujúci jedno z diel tohto autora, kde môžeme vidieť tiež prepojenie na záznam autora v súbore autorít práve cez tento identifikátor „jn19990218058“.

```
LDR      -----nam-a22-----1--4500
FMT      BK
BAS      |a 02
001      bknzdr15877
003      CZ-PrNK
005      20021103000000.0
008      021103s1939----xo-----slo--
040      |a ABA001 |b cze |c ABA001|d ABA001
1001     |a Kukučín, Martin, | d 1860-1928 |7 jn19990218058 |4 aut
24510    |a Čas tráť – čas platí. na hradskej ceste; |b Obecné trampoty; Hody; Ako sa kopú poklady; Visitatio canonica; Preháňanky; Pod Vládou ženy/ |c [Napísal] Martin Kukučín [pseud.]
260      |a V Turčianskom Sv. Martine: |b Matica slovenská, |c 1939 |f (Knihtlačiarsky účastinársky spolok)
300      |a 232-[I] s.; |c 8°
4901     |a Sobrané spisy / Martin Kukučín; |v Sv. 12
500      |a Vlastní jm. autora: Matej Bencúr
74002    |a Obecné trampoty 74002 |a Hody
74002    |a Ako sa kopú poklady
74002    |a Visitatio canonica
74002    |a Preháňanky
74002    |a Pod Vládou ženy
```

Obr. 2

830 0 |a Sobrané spisy (Martin Kukučín)
 9101 |a ABA001
 SYS 001207371

Práve toto prepojenie prostredníctvom trvalého, ale predovšetkým – jednoznačného, unikátneho identifikátora ja považujem za „revolučnú“ zmenu, ktorú priniesol UNIMARC.

Dôležité na tom je to, že pri tvorbe bibliografického záznamu totiž nevzniká už len jeden záznam. Tak ako tomu bolo predtým, ale s príchodom formátu UNIMARC sa súčasne vytvárajú či prepájajú viaceré záznamy medzi sebou – a to ako bibliografické – ak ide napr. o záznam článku v periodiku, tak zároveň prepojenia medzi bibliografickým záznamom a záznamami autorít či prepojenia medzi záznamami autorít navzájom, a to aj rôznych typov záznamov autorít. Samozrejme, tieto záznamy nevznikajú len tak, ale pri ich tvorbe je dôležité prísne dodržiavať jasne definované pravidlá, pretože vďaka internetu už knižnica nesppracováva fond len pre seba, ale internet pomyselne ruší všetky reálne a fyzické bariéry, ktoré boli spojené s prístupom do katalógov iných knižníc v prípade klasických listkových katalógov (geografické vzdialenosti a steny budov).

Možno sa vám to nebude zdať, ale dovoľm si tvrdiť, že aj napriek tomu, že v tom čase možno Tim Berners Lee ešte ani len netušil o webe, tak knihovníci UNIMARC-om už položili základné princípy sémantického webu. Sú nimi práve unikátne, trvalé a jedno-jednoznačné identifikátory autorít, ktoré sú použité pre vyjadrenie vzájomného prepojenia záznamov. Autority a pravidlá pre ich tvorbu tiež riešia veľmi zložitú problematiku homonym a synonym, ktoré do značnej miery komplikujú vyhľadávanie v prostredí klasického, a dokonca aj sémantického webu.

Tým sa dostávame k dvom technológiám, ktoré pre knihovníkov predstavovali ďalšie nové výzvy. Nielen knihovníci prijali web s veľkým očakávaním, pretože až ten v skutočnosti zbúral bariéry, ktoré dovtedy stáli v ceste prístupu k informáciám. Avšak to, čo sa spočiatku ukazovalo ako riešenie, zakrátko generovalo oveľa väčší problém. Web ako taký rozšíril predovšetkým hromadné využívanie internetu. Ten otvoril používateľom nové možnosti komunikácie, a tým aj získavania informácií.

Všetci však veľmi dobre vieme, že každá minca má dve strany. Jednoduchý spôsob publikovania informácií v prostredí internetu viedol k nárastu informačného obsahu a vyhľadávanie sa tým začalo komplikovať. Na jednoduchý dotaz webové prehliadače, či vyhľadávacie systémy, zapuzdrené do rôznych informačných systémov, nachádzali obrovské množstvo informácií a vybrať z nich tie pravé nebolo jednoduché. Stroje – teda počítače – vedia takéto informácie síce čítať, ale už nemajú nástroje a prostriedky, ktoré by im umožnili ich aj ďalej spracovať. To v takomto prípade vie len človek, ktorý keď si informáciu prečíta, vie si ju zatriediť podľa toho, čo momentálne hľadá. Lenže na to, aby človek dokázal prečítať a pomyslene spracovať také množstvo informácií, ktoré mu web na jeho dotaz ponúka, potrebuje stále viac času. Preto bolo potrebné nájsť riešenie, ktoré by strojom umožnilo všetky tieto informácie nielen čítať, ale hlavne spracovávať. To viedlo k definovaniu nového vývojového stupňa webu, ktorý v roku 2001 predstavil priamo Tim Berners Lee. Tento stupeň vývoja webu označil za web s významom, či za sémantický web – web2.0. Nechcem tu rozoberať problematiku web2.0, pretože je jasné, že sa o ňom od tej doby popísalo už mnoho, ja som sa s kolegyňkou Jankou Šubovou tejto téme bližšie venovala na INFORE 2007. Význam vložených informácií, ktorý by bol strojom spracovateľný, otvára totiž úplne nové možnosti vyhľadávania, ktorého výsledky môžu dosahovať vyššiu presnosť a hlavne relevantnosť vyhľadaných informácií v porovnaní s klasickým webovým prostredím. Avšak web2.0 = sémantický web, bol len zámer. V skutočnosti sa vývoj webu2.0 podstatne odklonil od pôvodného zámeru a dnes môžeme povedať, že web2.0 sa stal skôr webom sociálnym ako sémantickým. Webom, kde používatelia môžu ešte jednoduchšie zdieľať informácie medzi sebou, ale aj jednoduchšie komunikovať, vytvárať si vzájomne komunity vo virtuálnom priestore, vyjadrovať svoje názory, publikovať svoje zážitky a poznatky, atď. S týmto obdobím sa spájajú pojmy ako sociálne siete, tagovanie, blogy, krátke správy, atď. Publikovanie v prostredí webu sa tak stalo oveľa, ale oveľa jednoduchším. Nárast množstva informácií v prostredí webu tak začal naberať ešte väčšiu rýchlosť. Ocitli sme sa tak v prostredí informačnej explózie, kde hľadanie riešenia smerujúceho k zvýšeniu relevancie a presnosti vyhľadaných informácií je priam nutnosťou. Riešením má byť nový stupeň v evolučnom vývoji webu – teda sémantický web, ktorého princípy a základy boli prezentované už skôr. Viaceré spoločnosti a významné združenia preto ešte viac zacielili svoje aktivity smerom k snahe vyvinúť prostriedky a nástroje, ktoré by umožnili naplniť smelú myšlienku sémantického webu. Ten, ako ho predstavil Tim Berners Lee, pozostával z viacerých úrovní – jednotlivých technológií, tak, ako sme sa s tým mohli stretnúť už v množstve článkov zameraných priamo na sémantický web. Ja si tu z nich vyberiem len niektoré, ktoré považujem z hľadiska spracovania dokumentov za dôležité.

S jazykom XML (Extensible Markup Language) ste sa už isto stretli. Aj keď je to jazyk, častokrát sa môžeme stretnúť s tým, že ho označujú za formát, ale podľa môjho názoru formátom možno skôr označiť to, čo prostredníctvom tohto jazyka chceme definovať, aby sme si tieto informácie mohli zdieľať a vzájomne sme si rozumeli. V súvislosti so sémantickým webom budeme tento jazyk vnímať ako jazyk, ktorý bude použitý na to, aby sme ním vyjadrili význam, ktorý chceme, aby to vedeli spracovávať stroje. Ďalšou skratkou, ktorá sa často v súvislosti so sémantickým webom používa, je RDF – Resource Description Framework, čo predstavuje databázovú technológiu, prostredníctvom ktorej vieme zapísať význam jednotlivých vlastností – napríklad vyjadriť to, že Martin Kukučín je autorom publikácie Dom v stráni ap., a to tak, aby to vedel spracovať počítač. Lenže samotné RDF nestačí, pretože neposkytuje žiadne prostriedky, ktoré by umožnili zachytiť terminologický aparát používaný na popis dokumentov. Mám na mysli slová, ako je napr. kniha, zvukový dokument, článok, film, ale aj ich atribúty. Na to je potrebné rozšírenie RDFS, čo predstavuje *RDF Vocabulary Description Language*, alebo sa tiež často označuje ako RDF Schéma, z čoho pochádza aj akronym RDFS. RDFS slúži teda aj pre vyjadrenie slovníkov a tezaurov, ktoré sa v našom prípade používajú pri popise dokumentov. Platí to však všeobecne; RDF aj RDFS či ostatné technológie samozrejme nie sú určené len knihovníkom.

Knihovníci však boli tí, ktorí definovali jasné pravidlá, ako použiť RDFS pre vyjadrenie slovníkov a tezaurov tak, aby im všetci – teda aj stroje – správne rozumeli. Definovali formálny jazyk určený na prezentáciu systémov organizácie znalostí, inak povedané SKOS – Simple Knowledge Organisation System. V prípade vášho záujmu môžete viac informácií o ňom, ale aj o jeho praktickom využití v Národnej technickej knižnici, získať z článku Kristíny Kožuchovej (Škuta a Kožuchová, 2010). Ak sa ale vrátim k RDF, tak ten si prakticky môžeme predstaviť ako zápis veľmi jednoduchých tvrdení, ktoré svojou skladbou pripomínajú jednoduché vety. Azda každý z nás si pamätá na to, ako sme sa v prvej triede učili prvé vety typu „Mama má Emu“ či „Ema má mamu“. Teda podobne ako v týchto vetách, aj v tvrdeniach reprezentovaných v RDF, ide o spojenie podmetu a predmetu prostredníctvom prísudku. Preto, aby sme si vedeli jasnejšie predstaviť význam, prečo by mali tieto tvrdenia poskytovať širšie možnosti vyhľadávania, je dôležité si uvedomiť, že podmet aj prísudok predstavujú vždy „zdroje“, teda niečo, čo sa vyskytuje viackrát, alebo podľa čoho (akej vlastnosti) chceme vyhľadávať či nazerať na dáta. Na druhej strane predmet môže byť zdroj, ale taktiež môže byť vyjadrený len reťazcom – literálom. Označuje niečo, čo nie je významné z hľadiska

vyhľadávania. No a celý princíp spočíva v tom, že zdroje identifikujeme pomocou trvalých globálnych unikátnych identifikátorov – URI (Uniform Resource Identifier), čo veľmi nápadne pripomína princíp, ktorý zaviedol UNIMARC. Združenie W3C, rovnako ako pre ostatné veci, súvisiace s WWW, definuje jasné pravidlá, ako tvoriť URI. V zásade ide o to, že sa skladá z identifikácie domény a samotného identifikátora. Doména musí byť trvalo dostupná, rovnako, ako aj celé URI a doporučuje sa ešte, aby samotný identifikátor neobsahoval v sebe žiadnu významovú informáciu, ale bol skôr nezávislým identifikátorom, napr. určený poradovým číslom.

K technológiám sémantického webu neodmysliteľne patria tiež koncepty a ontológie. V nich môžeme formálne zachytiť vzťahy medzi jednotlivými pojmami a tieto potom využiť pri vyhľadávaní či agregácii súvisiacich informácií. Za takéto koncepty môžeme považovať aj už spomenuté súbory autorít. Vyšší stupeň predstavujú ontológie, ktoré združujú viaceré koncepty, a zároveň zachytávajú vzťahy a súvislosti medzi nimi. Definície k ontológiám existovali z počiatku len vo filozofii. V súvislosti s informačnými technológiami sa o nich hovorí len v posledných 20-tich rokoch. Bolo už o nich tiež popísané dosť, ja som o nich viac uviedla v príspevku s názvom Dotknú sa nás ontológie či Od zpracovávaní fondů k ontológiám, alebo okrajovo sme sa ich dotkli aj s Jankou Šubovou v článku Nevyhnutné podmienky pri tvorbe digitálnych knižníc či v ďalších.

Knihovníci si silu konceptov a ontológií jasne uvedomovali a už v začiatkoch ich to priviedlo k zamysleniu sa nad tým, čo používateľ v skutočnosti potrebuje. Výsledkom bol už v roku 1998 publikovaný entitno-relačný model funkčných požiadaviek na bibliografické záznamy, FRBR – Functional requirements for bibliographic records. Tieto neskôr doplnili požiadavky na záznamy autorít (FRAD) a aj záznamy vecných autorít (FRSAD). V praxi sa tiež stretávame s ontologickým rámcem CIDOC CRM – Conceptual Reference Model (<http://www.cidoc-crm.org/>). Tento definovali pracovníci z múzeí vychádzajúc pri tom z objektovo orientovaného prístupu. CIDOC CRM predstavuje formálnu štruktúru implicitných a explicitných konceptov, ako aj ich vzájomných vzťahov, používaných pri spracovaní informácií v doméne kultúrneho dedičstva. Tento model nehovorí nič o tom, ako sa majú veci robiť, ale len dokumentuje všetko to, čo pracovníci v múzeách a iných pamätových inštitúciách skutočne robia. S cieľom umožniť vzájomné zdieľanie údajov bibliografických záznamov aj s múzeami prepracovali knihovníci FRBR aj s použitím objektovo orientovaného prístupu. Výsledkom je FRBRoo, ktoré je vlastne rozšírením ontologického rámca CIDOC CRM o triedy či entity a vlastnosti, ktoré sú nevyhnutné pre zachytenie toho, čo spracovávajú a používajú knižnice. Aktuálne je platná verzia FRBRoo 2.0, ktorá zároveň predstavuje jeden z hlavných prostriedkov interoperability v doméne kultúrneho dedičstva. Musíme mať však na pamäti ten najdôležitejší fakt, ktorým je kvalita dát. Len kvalitné dáta je možné efektívne zdieľať s cieľom prepájať ich na základe významu, a aj v nich odhaľovať nové poznatky. Spôsob ich prezentácie je potom už len forma, ktorá v súvislosti s inými nástrojmi môže udávať mieru efektivity, ako aj možností spracovania a sprístupňovania týchto dát.

Z uvedeného je zrejmé, že knihovníci nemajú problém s technológiami ako takými. Preto sa natiska otázka, prečo teda knižnice dnes neponúkajú viac, a aj kvalitnejších služieb? A to nemám na mysli len vyhľadávanie, i keď je jasné, že kvalitné vyhľadávanie sa možnosťami súčasných technológií je základom každej služby. Dôvodov, prečo technológie nie sú v plnej miere knihovníkmi využívané, je viaceró. Jednu skupinu tvoria informačné systémy. Tie sú totiž stále rozsiahlejšie a nie vždy sú tak pripravené, alebo aj schopné tie najnovšie technológie aplikovať v rozsahu a spôsobom, ako by užívatelia očakávali. Druhá oblasť týchto problémov súvisí so samotnými knihovníkmi. Na jednej strane knihovníci síce veľmi rýchlo pochopili možnosti nových technológií, resp. v predstihu definovali to, čo by mali takéto technológie splňať, ale na druhej strane, ako je vidieť, potrebujú pomerne veľa času na to, aby si tieto poznatky osvojili do takej miery, ktorá im umožní ich zavedenie do každodennej praxe. Mrzí ma to, ale stále mám pocit, akoby sa knihovníci stále báli prekročiť svoj tieň, opustiť svoje staré paradigmy a v mnohých prípadoch môžeme badať ich snahu pri zavádzaní nových technológií o to, aby tieto čo najviac fungovali podľa ich doterajších zvyklostí. To je samozrejme na škodu veci a investície do technológií tak zatiaľ neprinášajú očakávaný efekt. Ale nie je to nič neobvyklé. Ak to porovnáme so zavádzaním technológií všeobecne, tak môžeme jasne vidieť, že internet tiež potreboval viac ako 20 rokov, kým neprišiel na scénu web. Jeho zásluhou sa potom internet hromadne rozšíril. Preto sa pýtam, čo musí prísť, aby knihovníci vykročili zo svojho tieňa a dokázali svetu, čo naozaj vedia? Kto iný, ak nie knihovníci, vedia lepšie spracovať informácie tak, aby boli všeobecne využiteľné aj budúcimi generáciami? Áno, môžete namietať, že dnes je úplne iná situácia a mnohonásobne viac informácií vzniká priamo v elektronickej podobe, čo je viac vecou informatikov. S tým však súhlasím len čiastočne. Súhlasím s tým, že tých zmien je stále viac a bez informatikov, či skôr znalostí technológií, sa dnes nepohne. Dovolím si však upozorniť na skutočnosť, že tak, ako som uviedla, boli to knihovníci, ktorí dali technológiám význam. Preto tvrdím, že teraz je to správne obdobie, kedy by sa mali knihovníci a informatici spojiť a spoločne vytvoriť nástroje, ktoré budú schopné efektívne spracovávať tú záplavu informácií a navyiac v nich objavovať, a aj z nich získavať nové poznatky. Nebojme sa preto nových vecí, lebo prístup typu „len ten, kto nič nerobí, nič nepokazí“ nám v tejto chvíli určite neprospeje!

Použitá literatúra:

AVRAM, D. Henriette. Machine-Readable Cataloging (MARC) Program. In: *Encyclopedia of library and information science: 3*, New York: CRC Press, 2003, s. 1712-1730. ISBN 978-08247-2079-7

Miller, Paul. Developing the Bath Profil. In: Ariadne [online]. 1999 [cit. 2015-09-12]. ISSN: 1361-3200. Dostupné na [www: http://www.ariadne.ac.uk/issue21/at-the-event/bath-profile.html](http://www.ariadne.ac.uk/issue21/at-the-event/bath-profile.html)

Miller, Paul. Z39.50 for All. In: Ariadne [online]. 1999 [cit. 2015-09-12]. ISSN: 1361-3200. Dostupné na [www: http://www.ariadne.ac.uk/issue21/z3950](http://www.ariadne.ac.uk/issue21/z3950)

ŠKUTA, Ctibor; Kožuchová, Kristýna. Vliv trendů systémů organizace znalostí na vývoj Polytematického strukturovaného hesláře v Národní technické knihovně. *Knihovna plus* [online]. 2010, č. 1 [cit. 2015-09-12] ISSN 1801-5948. Dostupné na [www: http://knihovna.nkp.cz/knihovnaplus101/sktua.htm](http://knihovna.nkp.cz/knihovnaplus101/sktua.htm)

Vodičková, Hana. Soubory autorit – příspěvek nejen terminologický. In: *Národní knihovna: knihovnická revue* [online], 1999, č.2 [cit. 2015-09-12]. ISSN 1214-0678. Dostupné na [www: http://full.nkp.cz/nkkr/Nkkr9902/9902063.html](http://full.nkp.cz/nkkr/Nkkr9902/9902063.html)

http://www.ukoln.ac.uk/interop-focus/activities/z3950/int_profile/bath/agenda.html

Ing. Nadežda Andrejčíková, PhD.

andrejcikova@cosmotron.sk

Cosmotron Slovakia s. r. o.