6. KAPITOLA

Organizace informací

Prvním krokem k moudrosti je znát věci samotné. To spočívá v pravdivé představě o objektech − objekty se dají odlišit a poznat tím, že je metodicky klasifikujeme a dáme jim vhodné názvy. Proto klasifikace a pojmenování bude tvořit základy naší vědy.

Carl Linnaeus

Katalogizátoři by ztratili mnoho ze svého postavení, kdyby někdo ukázal, že velká část katalogizace je triviální činnost a zvládli by ji i úředníci.

Maurice Line

Úvod

Organizace informací a informačních zdrojů je jedna z hlavních stránek informačních věd. Spočívá v podstatě v klasifikování a pojmenovávání a je stejně tak důležitá v našich vědách, jak podle Lineause v jeho, i když z poněkud jiných důvodů.[1] Jedná se o rozsáhlý a složitý obor sám o sobě, ale naštěstí má k dispozici obzvláště široký výběr učebnic a článků. V této kapitole načrtneme hlavní témata a problematiku této oblasti a ukážeme, kde najít podrobnosti. Jedním z hlavních rysů oboru je způsob, jakým se poměrně staré nástroje a techniky dokáží přizpůsobit modernímu informačnímu prostředí, a proto mohou být velmi užitečné i starší texty, přestože moderní materiály jsou důležité. Základy oboru, zejména část teorie klasifikace a indexace, jsou už velmi staré a, jak jsme viděli ve 2. kapitole, historie některých hlavních nástrojů sahá až do 19. století. Byly vyvinuty v rámci snah zavést bibliografickou registraci tištěných materiálů − zaznamenat, identifikovat a zpřístupnit veškeré intelektuální výstupy lidstva ve formě vyjádřených a zaznamenaných znalostí. Tyto nástroje a jejich modernější ekvivalenty se nyní používají k usnadnění přístupu k rychle se rozšiřujícím zásobám digitálních materiálů.

Nejprve se budeme zabývat některými fundamentálními otázkami organizace informací, poté se podíváme na hlavní nástroje, terminologii, metadata, popis zdrojů, systematický a abecední věcný popis a referování. Podrobnější texty o některých těchto aspektech najdete u Chowdhuryho a Chowdhuryho (2007), Taylora (2004), Chana (2007) a Svenoniouse (2000). Podrobněji prozkoumává všechny stránky organizace informací Robinsonová (2010, 4. kapitola).

Tento obor se nazývá buďto „organizace informací“ nebo „organizace znalostí“. Většinou se tyto termíny používají synonymně, nicméně nám připomínají, že účelem organizace může být buďto pochopení struktury samotných znalostí, nebo praktické uspořádání dokumentů, původně fyzických v regálech, později digitálních ve virtuálním prostoru. Už od nejranějších časů filozofie se hledaly odpovědi na otázky: Existuje jediná „přirozená“ klasifikace všeho na světě? Na jakém základě třídíme věci do kategorií? Jaký je vztah mezi našimi představami a fyzickými věcmi? Jak moc se mohou dvě věci od sebe lišit a přesto mít stejný název? A další. Všechny organizace informací a znalostí používají celou řadu teoretických poznatků − ať už se jedná o klasifikační schémata pro dokumenty nebo vědecké taxonomie rostlin, zvířat, minerálů, hvězd a mnohých dalších.

[1] pozn. editora: Carl Linnaeus byl švédský botanik a zoolog, který vytvořil moderní názvosloví organismů (binominální nomenklaturu).

Přehled teoretických základů pro organizaci dokumentárních informací podávají Svenonious (2000), Tennis (2008) a Hjørland (2003, 2008a), teoretická odůvodnění různých aspektů Hjørland (2008b, 2009, 2011), Hjørland a Pedersen (2005), Hjørland a Shaw (2010) a Weinberg (2009).

Řízený slovník a fasetová analýza

Jedná se o dva fundamentální pojmy a s oběma se setkáváme v několika aspektech organizace informací.

Řízený slovník je v nejjednodušší formě seznam termínů, které se mají používat pro indexaci a rešerši. Příkladem jsou seznamy klíčových slov, předmětová hesla, klasifikace, taxonomie, tezaury, seznamy autorit a jiné. Tyto termíny „řídí" variabilitu a redundanci přirozeného jazyka. Opačným pojmem je neřízený slovník – plný text, volně zvolená klíčová slova, tagy atd. Zcela zbytečně se udržují při životě neutuchající spory, zdali je „lepší" řízený či neřízený slovník. Zcela nasnadě je odpověď, že žádoucí je kombinace obou: řízený slovník pro konzistenci a používání vztahů mezi termíny, neřízený pro přesnost a pro nové termíny.

Součástí fasetové analýzy je rozdělení pojmů v rámci oborové domény na konzistentní oddíly. Například obor „historické budovy" může mít fasety ÚČEL (dům, kostel, škola…), STYL (gotický, klasický, Arts and Crafts…), STAVEBNÍ MATERIÁL (kámen, cihly, dřevo…), DOBA (Viktoriánská, středověk…) atd. Tento styl analýzy se hodně používá při vytváření klasifikací a tezaurů, designu rozhraní, konstrukci komplexní vyhledávací logiky a dalších (La Barre, 2010; Broughton, 2006a).

Nyní se můžeme podívat na nástroje pro organizaci informací. Začneme tím nejznámějším – seznamy terminologií.

Terminologie

Terminologií se obecně rozumí slova a výrazy používané při sdělování informací v konkrétním oboru nebo oblasti, v širším pojetí to znamená obecně používaná slova a výrazy. Terminologické zdroje zahrnují řadu překrývajících se kategorií: obecné slovníky, vědecké, začátečnické nebo zkrácené a ilustrované, multilingvální slovníky, slovníky a glosáře pro specifické obory, speciální slovníky, slovníky rýmů, citátů, křížovkářské slovníky a tezaury, spíše ve smyslu Rogetova hledače slov než rešeršní tezaury diskutované dále. Mohli bychom zde také zahrnout slovníky vlastních jmen osob („biografické slovníky") nebo míst (geografické nebo zeměpisné slovníky). Kdysi ztělesněné kvalitně vytištěnými svazky dostávají nyní všechny digitální podobu a jsou čím dál tím více vytvářeny díky příspěvkům z takzvaného „crowd-sourcingu", zvláště ty pro populární využití.

Mezi nástroji předmětové terminologie, jako jsou výkladové slovníky a seznamy předmětových hesel, tezaury a taxonomie probírané dále, existují jistá překrytí. Všechny se dají použít jako zdroj termínů pro indexaci a stejně tak mohou posloužit při porozumění logice a jazyku nějakého oboru. Určitý překryv je také mezi seznamy „lidí a míst" a soubory autorit používanými při katalogizaci a popisu zdrojů. Mají v prvé řadě splňovat požadavky, jak poznamenává Buckland (2008), na „obecnou referenční" funkci – poskytovat nebo potvrzovat fakta (v tomto případě vysvětlovat význam slova nebo výrazu) a poskytovat kontext pro takové vysvětlení. Počátky a vývoj zdrojů tohoto druhu popisují Hitchens (2005), Hüllen (2004), Mersky (2004) a Mugglestone (2005), současnou problematiku Tackabery (2005), Mugglestone (2011), Cassell a Hiremath (2011, kapitoly 7, 10 a 11) a Ayre, Smith a Cleeve (2006).

Kromě oborově specializovaných slovníků a glosářů existují, zejména u STEM oborů, různé speciální jazyky a notace. Některé notace, např. v matematice, hudbě a tanci, jsou samy o sobě jazyky, jiné jsou podrobnými a specifickými terminologiemi; např. nomenklatury pro chemické struktury a reakce (Leigh, 2011), nomenklatury a taxonomie pro živé věci (Bowker, 2005; Heidorn, 2011) a specializované terminologie pro lékařství a zdravotní péči (Robinsonová, 2010, 4. kapitola).
Různé výše popsané terminologie mohou poskytnout termíny pro indexaci, ale jejich hlavním účelem je spíše komunikace než vyhledávání informací. Nyní si všimneme slovníků a norem vytvořených speciálně pro indexaci a vyhledávání, začneme s metadaty, která slouží jako „standardní kontejnery“ pro popis dokumentů.

Metadata

Metadata jsou vlastně „data o datech“, neboli krátké strukturované popisy informačních zdrojů. Termín samotný se rozšířil v 90. letech minulého století, ale základní myšlenka existuje v katalogizačních pravidlech už od poloviny 19. století. Stručný úvod do principů metadat najdete u Haynese (2004), více podrobností o specifických formátech a schématech uvádějí Zeng a Qin (2008), Miller (2011) a Hider (2012).
Metadatové záznamy zastupují původní jednotky a používají se místo nich. Hlavním účelem metadat je identifikace, vyhledání, používání a správa informačních zdrojů. To zahrnuje vyhledání zdrojů, nalezení konkrétní požadované jednotky prozkoumáním nebo prohlížením, její zobrazení, rozhodnutí pravděpodobné užitečnosti, posouzení právních otázek. Dále si povšimneme přístupových práv, kdo záznam spravuje, kdo je za záznam zodpovědný, kdy by měl být revidován, archivován atd. Je také žádoucí, aby se metadata mohla sdílet a vzájemně vyměňovat, proto existuje požadavek na standardizaci.
Metadata se obvykle skládají ze dvou složek − popisná metadata a předmětová metadata. Tradičně mívala v knihovnách s lístkovým katalogem různé fyzické formy: například katalog „autor/titul“ a zvlášť věcný katalog, nyní se ale obvykle oba slučují do jediného metadatového formátu. Popisná metadata popisují samotnou jednotku − titul, autora, datum publikace nebo vytvoření, fyzickou formu atd. Předmětová metadata popisují obsah, o čem daná jednotka je. Předměty se mohou popisovat řízenými termíny (klasifikačními znaky, předmětovými hesly apod.), nebo neřízenou terminologií (např. uvedením termínů použitých v titulech a abstraktech). Použijeme-li pojmy Popperových tří světů, můžeme si představit, že popisná metadata definují informační objekty světa a předmětová metadata definují jejich obsah ve světě 3. Popisná metadata odpovídají na otázky: Jak se tato jednotka nazývá? Kdo ji napsal? Jak je stará? Jak je velká? Jaká je to věc? Kde je? Předmětová metadata odpovídají na otázku: O čem to je? Podrobnosti popisných metadat se mohou lišit podle fyzické povahy jednotky, např. odpovědí na otázku Jak je velká? bude u knihy počet stránek a údaj v centimetrech a u digitálního zdroje počet megabytů. Předmětová metadata jsou invariantní vůči fyzické formě, protože popisují vnitřní obsah, proto se pro všechny formy používají tytéž nástroje, např. Deweyho desetinné třídění používané v knihovnách pro řazení knih v regálech se využívá i u internetových slovníků.
Jak už bylo řečeno, metadatové záznamy by měly být relativně krátké a musejí být strukturované a standardizované. Strukturované znamená, že metadatový záznam obsahuje pole, prvky nebo atributy reprezentující odlišné typy informací, např. titul, klíčové slovo nebo

jméno autora. Takto lze odlišit záznamy odvolávající se na díla, která napsal Benjamin Disraeli od děl napsaných o něm.
Standardizované znamená, že táž informace se prezentuje vždy stejně − Disraeliho jméno se vždy bude psát „Disraeli, Benjamin“ nebo „Benjamin Disraeli“ nebo „Disraeli B.“. Žádná z těchto možností není špatná ani dobrá, ale měla by být zachována jednotnost.
Mimochodem, Disraeli se později stal lordem Beaconsfieldem a pod tímto jménem napsal i několik knížek. Bylo by dobré, kdyby náš metadatový fond obsahoval odkazy na obě jména. Aby to ale bylo možné, je potřeba mít dva odlišné druhy znalostí: obecnou znalost, že je možné, aby jedna osoba měla dvě jména, a specifickou znalost, že Disraeli/Beaconsfield je jedna a tatáž osoba. Takovéto znalosti si obvykle odborníci pamatují a dovedou je aplikovat, ale s počítači je to horší. I když by bylo velmi žádoucí, aby bylo možné metadata vytvářet automaticky, zvláště dnes, kdy je potřeba zvládat obrovská množství digitálních informací, vytváření metadat lidskými odborníky je stále považováno za nejlepší variantu, jak dosáhnout nejlepší kvality, i když, jak nám připomíná úvodní citát od Maurice Lineho, někteří informační profesionálové zpochybňují mystéria vzniklá okolo některých otázek.
Abychom měli představu o jejich rozmanitosti, uvádíme dále příklady současných významných metadatových standardů pro obsah a prvky záznamů; více podrobností uvádějí Zeng a Qin (2008) a Miller (2011).

- *Machine Readable Cataloguing* (MARC) je výměnný formát pro metadatové záznamy vytvořené podle katalogizačních pravidel AACR2 a RDA. V minulosti existovaly četné národní varianty, například USMARC a UKMARC, mezinárodní verze, konkrétně UNIMARC. V současné době vzniká MARC21 jako nový de facto mezinárodní standard. Obsah záznamu se ukládá v polích a podpolích.

- *Dublin Core* (DC) a deriváty je dosud nejlepší metadatový formát. DC byl navržen jako nejjednodušší možný užitečný metadatový formát původně obsahující pouze 15 prvků, např. „titul“, „tvůrce“, „obor“ a „formát“, s minimálními pokyny pro zadávání obsahu. V současnosti byl rozšířen o četné varianty. Jednou z nich je GMS standard britské vlády, který rozšiřuje DC na 23 polí, včetně některých zohledňujících vládní materiály, například „mandát“ a „zachování“.

- Formát *Learning Object Metadata* (LOM) byl určen pro metadata k výukovým účelům na všech stupních granularity, od diagramu přes videoklip až po učebnici. Má složitý systém polí včetně specificky edukačních prvků jako „typický věkový rozsah“ a „typ interaktivity“.

- Standard *Text Encoding Initiative* (TEI) slouží k reprezentaci textů v digitální formě, což je zvlášť cenné pro „digital humanities“. V době psaní této učebnice existují na adrese http://www.teibyexample.org/ tutoriály a příklady pod názvem „TEI by example“.

- *Metadata Encoding and Transmission Standard* (METS) vyvinutý v Library of Congress pro archivaci digitálních jednotek je modulárním a flexibilním standardem. *Metadata Object Description Schema* (MODS) je rozšíření METS sloužící k reprezentaci katalogizačních záznamů knihovny.

- *International Standard for Archival Description [General]* (ISAD[G]) je metadatový standard poskytující obecné vodítko pro popis archivních záznamů.

- *Visual Resources Association* (VRA) je standard pro popis děl vizuálních kultur a obrázků, které je dokumentují.

Tyto metadatové standardy a formáty, které umožňují vytváření záznamů popisujících informační zdroje, jsou podporovány celou řadou jazyků a standardů, díky nimž jsou kódovány a implementovány ve webových prostředích, což je někdy považováno za posun k sémantickému webu a bude to předmětem další kapitoly. Důležitým příkladem je jazyk XML, který se používá ke kódování několika výše uvedených standardů a jejich variant například LOM, METS, MODS nebo DC. Tyto standardy mají typicky formu XML schématu se jmenným prostorem oznamujícím umístění jednotlivých komponent a definicí.
Resource Description Framework (RDF) je obecný metadatový model založený na popisu zdrojů jako série výrazů, tzv. trojic – subjekt-predikát-objekt, velmi podobný databázovým modelům typu entita-relace, probíraným v další kapitole. Považuje se za standardní jazyk pro reprezentaci znalostí na webu. Mapy témat jsou podobným druhem formální reprezentace znalostí.
Formáty RDF a XML se používají k vytváření specifičtějších metadatových modelů, např. *Web Ontology Language* (OWL) navržený k reprezentaci ontologií či *Simple Knowledge Organization System* (SKOS), model pro reprezentaci řízených slovníků, např. klasifikačních schémat, taxonomií, seznamů předmětových hesel a tezaurů.
Přehled najdete u Antoniou a van Harmelena (2008) a jako příklad viz jazyk XML použitý pro kódování záznamů MARC (Dimic, Milsavljevic a Surla, 2010) a ontologii reprezentovanou v RDF pro vytvoření datového souboru prodejů uměleckých děl (Allinson, 2012).
Nyní se podíváme na to, jak se vytváří první forma metadatového obsahu – popis informační jednotky.

Popis zdroje a katalogizace
Zahrnuje vytváření popisných metadat reprezentujících fyzickou formu daného dokumentu. Termín katalogizace se dosud obecně používá ve vztahu ke knihovním materiálům, popis zdrojů je obecnější pojem a stále častěji se používá. Podrobnější pohled na katalogizaci najdete u Bowmana (2003) a Welshe a Batleyho (2012).
Charles Ammi Cutter, americký knihovník, který byl jedním z iniciátorů moderních myšlenek popisu zdrojů v 19. století, tvrdil, že katalog by měl plnit následující funkce (vyjádřeno modernějším jazykem):

- Umožnit uživateli nalézt zdroj, ke kterému zná jednoho nebo více autorů, titul nebo předmět.
- Ukázat, jaké zdroje napsané danými autory nebo o daném tématu jsou k dispozici.
- Pomoci uživateli vybrat si nejlepší zdroj vyhovující jeho potřebám, podle vydání (datum, vydavatel atd.) a podle povahy (styl, úroveň atd.).

Jsou to dosud velmi relevantní účely. Mezi běžně citované účely katalogů tištěných knihoven patřily:

- lokace − určení, kde se konkrétní zdroje nacházejí,
- kolokace − seskupení souvisejících děl (např. od téhož autora nebo na totéž téma, nebo sérii knih či zpráv),
- informování − přímé poskytnutí některých potřebných informací (např. úplnou bibliografickou referenci, plné jméno autora, přesné jméno kolektivního autora).

Že jsou tyto účely dosud stále relevantní i v digitálním prostředí, potvrzuje jejich představení si místo fyzického seskupení na regálu dynamické seskupení na obrazovce.

Tyto soubory obecných účelů se rozšířily na seznamy specifičtějších „principů“, které pomáhají vytvářet explicitní katalogizační pravidla a další protokoly pro popis zdrojů. Také naplňují snahu o univerzální bibliografickou registraci poskytováním konzistentních popisů pro všechny publikované dokumenty. Viz Bowman (2006), kde najdete historii raných pokusů. Většina moderních knihovních katalogizačních pravidel je založena na důležitých *„Pařížských principech“,* schválených mezinárodní konferencí o principech katalogizace v roce 1961. Nejnovějším souborem takových zásad je „*Statement of International Cataloguing Principles*“ (ICP) mezinárodní federace IFLA, který obsahuje více typů zdrojů než dřívější zásady (Tillett a Cristan, 2009); komentář na toto téma najdete u Guerriniho (2009).

Katalogizační pravidla používaná knihovnami po celá desetiletí podávají velmi podrobné instrukce pro přesný popis například jmen autorů a vydání díla nebo pro specifikaci jmen ilustrátorů a překladatelů, což je dáno potřebou přesně identifikovat specifické tištěné dokumenty, zejména knihy v rozsáhlých fondech. Nejznámější z těchto pravidel jsou *anglo-americká katalogizační pravidla* (AACR). Tištěná podoba druhého upraveného vydání těchto pravidel z roku 1998 má 26 kapitol na 676 stránkách, což napovídá, jak jsou taková pravidla nutně složitá.

AACR se zakládají na obecnějším standardu ISBD (*International Standard Bibliographic Description*) ze 70. let minulého století, který na obecnější úrovni předpisuje, co se smí objevit v popisu bibliografické jednotky. K tomu je nutno specifikovat následující prvky: titul a údaje o původcích [autor], vydání, materiál nebo typ publikace [fyzická podoba], publikování, distribuce atd. [vydavatel], fyzický popis [rozměry], série, poznámka, standardní číslo a podmínky zpřístupnění. AACR specifikuje, dostatečně přesně kvůli konzistenci, jak se tyto prvky mají vyjadřovat.

Omezení katalogizačních pravidel typu AACR vedlo k vyvinutí modelu funkčních požadavků na bibliografické záznamy (FRBR), který začíná mít vliv na katalogizační systémy a praktiky (IFLA, 1998; Zhang a Salaba, 2009). Je založen na modelu entita-relace, o kterém se bude mluvit v další kapitole v části o databázích, kde se popisují vztahy mezi různými typy dokumentů a jejich atributů a lidmi a organizacemi, které je vytvářejí a rozšiřují; příklad vidíte na Obrázku 6.1.

Jak bylo řečeno ve 4. kapitole, podle FRBR se rozlišují čtyři úrovně − dílo, vyjádření, provedení a jednotka − ty ale nejsou zcela uspokojivě vysvětleny. Nicméně existují jisté důkazy, že toto je „přirozený“ způsob, jak nahlížet bibliografické entity (Pisanski a Zumer, 2010) a jak založit formální model.

Knihovní katalogy typicky umožňují hledání na úrovni provedení a zobrazování výsledků na úrovni jednotek. Model FRBR by měl umožnit flexibilnější vyhledávání a zobrazování

a dosáhnout tím větší přesnosti (např. „najdi záznamy pro toto vydání této knihy, jejíž výtisk je k dispozici v knihovně k vypůjčení a mám k němu přístup“) a také větší obecnosti (např. najdi všechno o Shakespearově Bouři − hru samotnou, její různá vydání, romány podle ní napsané, překlady, komentáře, verze audioknih a filmové verze včetně souvisejících filmů například Forbidden Planet nebo Prospero's Books).
Existují však spory, jak se mají tyto čtyři úrovně interpretovat. Tyto spory pocházejí ze základních otázek, co je to dokument (probíráno v předchozích kapitolách) a jak rozumět pojmu „dílo“ (Smiraglia, 2001).
FRBR a přidružené funkční požadavky na autoritní data (FRAD), které udávají standardní tvary, v nichž se mají psát jména lidí, názvy organizací, děl atd. (Patton, 2009), a funkční požadavky na předmětová autoritní data (FRSAD) pro popis oboru (Salaba, Zeng a Zumer, 2011) tvoří „koncepční základ“ pro katalogizační standard *Resource Description and Access* (RDA), jenž má nahradit AACR2. Měl by se používat mimo knihovny v širším prostředí sbírek, například v muzeích a archivech. Přehledy uvádějí Oliver (2010) a Anhalt a Stewart (2012).
Na rozdíl od AACR byl standard RDA navržen na základě formálního datového modelu a vyhýbá se složitým zvláštním pravidlům pro popis různých druhů materiálu. Byl však navržen tak, aby byl kompatibilní s AACR a nemusely se upravovat stávající katalogové záznamy. Už od počátku byl kritizován, že je AACR příliš ovlivněn a tak uvízl v minulosti; viz například Coyle a Hillman (2007). Jeho realizace se zpozdila a dosud není jasné, do jaké míry ho budou v současné formě podporovat velké knihovny a informační služby (Anhalt a Stewart, 2012).
Nyní budeme hovořit o nástrojích používaných k popisu předmětu − „intencionalitě“ dokumentu. Podle ustálené konvence je rozdělíme na systematické a abecední nástroje: klasifikace a taxonomie respektive předmětová hesla a tezaury. Je však třeba všimnout si, že tyto dvě kategorie se poněkud překrývají, klasifikační schémata mají často abecední indexy, aby se dala místa specifických předmětů ve schématu rychle najít, zatímco na abecední slovníky, uvádějící širší a užší termíny, se dá pohlížet jako na taxonomii. Všimněme si také důležitosti nástrojů, které dovedou překládat mezi různými slovníky a spojují tak způsoby, jakými je pojem v každém z nich nahlížen. Někdy se tomu říká mapování metadat – překladače převádějí pojmy mezi dvěma specifickými slovníky nebo metadatovými formáty. Existují také metaslovníky, které spojují několik slovníků nebo terminologií; příkladem je *Unified Medical Language System* (UMLS), který spojuje zdravotnické slovníky (Robinson, 2010). Napřed se však krátce podíváme na poněkud nepřesně používaný termín ontologie.

Ontologie

V informačních vědách neexistuje jediný obecně přijímaný význam pojmu „ontologie“. Slovo pochází z řeckého *ontos*, čili to, co existuje, a používá se ve filozofii ve významu studia toho, jaké druhy věcí mohou existovat a jak se dají popsat. Někdy se používá k neformálnímu popisu obecného souboru určitých předmětů zájmu − ontologie nějaké oblasti v tomto smyslu bude znamenat hlavní pojmy této oblasti.
Termín se ujal v počítačové vědě, kde znamená formální popis domény znalostí pomocí termínů v rámci této domény a relací mezi nimi. V tomto smyslu by se mnoho řízených slovníků − zejména klasifikačních schémat, taxonomií a tezaurů − dalo považovat za ontologie, i když poněkud jednoduché, a také se tak často popisují, zvláště v kontextu sémantického webu.

Někdy se termín vyhrazuje pro slovníky s bohatou množinou relací, větší než asociativní relace typu synonymum/hierarchie, které jsou běžné u rešeršních slovníků. Například biomedicínské ontologie používají relace „je obsažen v“, „je sousední k“, „předchází mu“, „transformace čeho“ nebo „má účastníka“ (Smith a kol., 2005).

Systematické slovníky: klasifikace a taxonomie

Klasifikace, kategorizace a taxonomie jsou formy organizace znalostí, které mají ukázat vztahy mezi pojmy (obecně hierarchické vztahy) seskupením termínů představujících podobné významy. Proto se jim říká „systematické“ slovníky – jsou vytvořené podle nějakého „systému“. Podrobnosti o klasifikacích, jak praktických, tak teoretických, najdete u Broughtona (2005), Huntera (2009) a Bowkera a Starové (2000), velmi přehledný a stručný výčet zásad, včetně aplikací ve webovém prostředí podává Slavic (2011).

Teorie klasifikace může být velmi složitá (viz například Langridge, 1992; Beghtol, 2010; Bowker, 2005; Bowker a Star, 2000; Hjørland a Pedersen, 2005; a Hjørland, 2008b), ale některé pragmatické body se dají formulovat jednoduše:

- Klasifikace znázorňuje vztahy mezi pojmy, zejména, i když ne výlučně, hierarchické vztahy.
- Klasifikace je proces kategorizace pojmů do vzájemně disjunktních množin pomocí racionálních zásad dělení.
- Konkrétní jednotky se dají zřídka klasifikovat absolutně, spíše na bázi celkové podobnosti.

Aby byla klasifikace přiměřená a použitelná:

- musí se vztahovat na podobné věci,
- výsledkem musejí být množiny podobného charakteru a podobných velikostí,
- musí pro rozdělování aplikovat konsistentní kritéria,
- musí aplikovat v dané chvíli jediné kritérium.

U formálně vytvářených klasifikací se tyto zásady vždy dodržují, ale mohou se dál členit na jednodušší taxonomie a kategorizace.

Všechny klasifikace mají stejnou notaci − numerický či alfanumerický kód, který odráží strukturu klasifikace a ukazuje vztah mezi jejími komponentami. Například v Deweyho desetinném třídění je číslo 425 přiřazeno „Gramatice standardní angličtiny“. Podle struktury desetinné notace se jedná o část třídy 420 „Angličtina a stará angličtina“, která sama je součástí třídy 400 „Jazyk“. Klasifikační notace má tu výhodu, že není závislá na jazyku, kniha na určité téma bude mít tutéž notaci v jakékoliv knihovně na světě, která tuto klasifikaci používá.

Existuje vzájemný vztah mezi systematickými a alfabetickými slovníky. Alfabetický indexační slovník, který uvádí širší i užší termíny podrobně, se dá zobrazit jako hierarchická klasifikace, zatímco „horní termíny“ tezauru nebo soubor předmětových hesel se dají použít jako množina širších pojmů nějaké taxonomie.

Klasifikace je proces podobný indexaci, je však obvyklé, aby byla přiřazena pouze jedna klasifikační notace. Není to bezpodmínečně nutné, pokud je cílem klasifikace poskytnout jediné místo pro fyzickou lokaci výtisku. U digitalizovaného materiálu je možno použít tolik klasifikačních znaků, kolik je třeba.

Nejjednodušší formou systematického slovníku je kategorizace, což je jednoduchá forma klasifikace, s omezenou strukturou a detaily. „Široké“ kategorizace se tak nazývají proto, že do každé kategorie zahrnou široké pojmy předmětu. Jsou užitečné pro prohlížení a pro fyzické uspořádání fondu.
Používají se v různých situacích: v knihovnách zejména veřejných či školních, kde se vyžaduje jednoduchá struktura s usnadněnou možností prohlížení nebo tam, kde beletrie tvoří značnou část fondu, v knihkupectvích (kamenných i internetových) a jako dodatečný nástroj vyhledávání v některých počítačových databázích, zejména u humanitních a sociálních věd. Také je můžeme najít v mnoha webových slovnících a portálech.
Jedná se o jednoduchou formu klasifikace bez hierarchické struktury nebo s velmi malou strukturou málokdy jdoucí více než jednu úroveň do hloubky a často porušující zásadu jednoho pravidla pro rozdělení. Například v knihkupectví mohou mít obecnou kategorii pro VAŘENÍ, rozdělenou na VEGETARIÁNSKÉ VAŘENÍ, ČÍNSKÁ KUCHYNĚ, VAŘENÍ V MIKROVLNNÉ TROUBĚ atd. Není pravděpodobné, že by se vyskytlo dělení do dalších úrovní (např. ČÍNSKÁ VEGETARIÁNSKÁ KUCHYNĚ atd.), a také se nedá očekávat, že by si majitel obchodu dělal starosti s nekonzistentním rozdělením podle ingrediencí, regionů, kuchyňských nástrojů atp. Takové systémy fungují, protože se jedná o malé měřítko, zákazníci v obchodě přehlédnou značnou část knih a kategorií jedním pohledem, a jsou nastaveny tak, aby vyhovovaly potřebám uživatelů a povaze matriálu. Stejné úvahy ospravedlňují jejich používání při přístupu k digitálnímu materiálu s jednoduchým navigováním.

Složitější jsou taxonomie − klasifikace vytvořené pro konkrétní prostředí nebo množinu informací. Obvykle dost věrně kopírují místní podmínky a přizpůsobují se lokální „kultuře“. Většinou jsou modifikovány a rozšiřovány častěji než ostatní informační nástroje. Přehled taxonomií tohoto druhu podává Lambe (2007). Tento termín se ovšem také dosud používá ve starším smyslu: vědecká klasifikace nějakého aspektu přirozeného světa. Taxonomie obvykle vytvářejí lidé, ale dají se také generovat automaticky. Jsou populárním nástrojem po organizování digitálních informačních zdrojů, vždy podporují prohlížení a někdy také vyhledávání.
Může se stát, že taxonomie „poruší pravidla“ klasifikace, například tím, že umožní, aby se nějaký pojem vyskytnul na různých úrovních, což pomáhá uživatelům při prohlížení. Mohou představovat popis tématu na vysoké úrovni − například „podniková taxonomie“ je jistým způsobem vyjádření zájmů organizace, aby se například mohl organizovat materiál na intranetu. Tento druh taxonomie může odkazovat na tezaurus pro podrobnější terminologii. Taxonomie mohou také obsahovat mnoho specifických příkladů (názvy míst, jména lidí či názvy oddělení, stejně jako hesla obecných pojmů) a mohou být více podobné tezauru než knihovní klasifikace. Také mohou obsahovat mnoho popisných informací o pojmech a jednotkách a být tak jakýmsi nástrojem poskytujícím informace. Taxonomie se typicky používají k umožnění přístupu k různým formám materiálů (např. databázím, dokumentům, e-mailům, lidem) a tedy k podpoře programů managementu znalostí.
Složitější a všeobsáhlé jsou nejstarší zavedené slovníky tohoto druhu − enumerativní klasifikace, navržené pro fyzické uspořádání knihovních materiálů a v poslední době používané pro věcné vyhledávání digitálních informací. Enumerativní klasifikace mají za cíl uvést kompletně, vyčíslit

neboli enumerovat, všechny aspekty znalosti ve svém záběru. Jsou bez výjimky hierarchicky uspořádané a znalosti hierarchicky rozdělují, například Deweyho klasifikace se dělí na deset hlavních tříd a každá třída má deset divizí, přičemž každá divize se dělí na deset sekcí s dalším dělením podle potřeby, její struktura na úrovni divizí (s hesly kvůli přehlednosti zjednodušenými) je vidět na Obrázku 6.2. Jedná se o typický druh enumerativní klasifikace.

Velmi dobře se hodí pro uspořádávání velkých objemů materiálů, obzvláště když se požaduje fyzické uspořádání s místem pro každou jednotku, a proto se hojně používají pro knihovní klasifikace. Jejich nevýhodou je, že pracují s velmi podrobnými popisy předmětu a s jednotkami zahrnujícími několik pojmů, byl také kritizován rozsah pokrytí všech znalostí (Zins a Santos, 2011). Nehodí se rovněž pro rychle se měnící součásti oborů, protože nemohou být často revidovány. Témata jako například „Internet“ nebo „AIDS/HIV“, ke kterým se objevilo značné množství literatury, způsobovaly enumerativním klasifikacím, které pro ně původně neměly žádné místo, problémy. Protože se ve velké míře využívají mezinárodně, jejich revize jsou obecně v rukou mezinárodních komisí, které revidují konkrétní sekce a podsekce v ne příliš častých intervalech.

Nicméně enumerativní klasifikace jsou stále hlavním nástrojem pro věcný popis používaný v katalogových záznamech knihoven a používají se k organizaci internetových zdrojů na některých webových portálech.

Nejznámější a nejpoužívanější příklady jsou: Deweyho desetinné třídění (DDT), které vyvinul americký knihovník Melville Dewey a publikoval je poprvé v roce 1876, Mezinárodní desetinné třídění (MDT), sestavené v roce 1905 jako rozšíření Deweyho schématu dokumentaristy Paulem Otletem a Henri La Fontainem, o nichž se hovořilo v předchozích kapitolách, a klasifikace knihovny Kongresu (LCC), kterou vytvořila knihovna Kongresu a začala ji používat v roce 1901. Všechny tyto klasifikace se průběžně revidují, např. DDT je v současnosti používáno ve svém 22. vydání, přesto je dosud patrna doba jejich vzniku – například Deweyho hlavní struktura, uvedená výše, odráží jak západní svět 19. století, tak prostředí svobodných umění, ve kterém vznikala. I když byly všechny obecně používané enumerativní klasifikace v detailech aktualizovány, podstatnější změny se netěší přílišné oblibě, vzhledem k následným dramatickým změnám, které by postihly velké knihovny používající klasifikaci i k fyzickému rozmístění materiálů. Podrobnější popisy Deweyho klasifikace uvádějí Bowman (2005), Chan (2007, 13. kapitola) a Satija (2007).

Deweyho třídění a zejména MDT, které je navrženo tak, aby se vypořádalo s detailní analýzou technických materiálů lépe než Deweyho, získaly v posledních vydáních „syntetické“ nebo „čísla vytvářející“ možnosti. Místo uvádění seznamů (enumerací) notací pro všechny možné pojmy, se dají notace vytvářet. Tuto schopnost měly už od začátku ve formě tabulek poddivizí pro typy materiálů, časová období a geografické oblasti, které se daly přiřadit k číslům tříd. To se dá dále vylepšit kombinováním témat. Uveďme jednoduchý příklad. Chceme-li v Deweyho klasifikaci klasifikovat knihu o „zemědělském výzkumu v Japonsku“, napřed musíme určit, že hlavní pojem je „zemědělství“, které se má zúžit pojmy „výzkum“ a „Japonsko“, přičemž poslední dva pojmy se vezmou z tabulek obecných pojmů, které se dají aplikovat v mnoha případech. Protože zemědělství má notaci 630, výzkum 072 a Japonsko je 052, můžeme vytvořit následující číslo třídy:

630.72052 zemědělský výzkum v Japonsku

MDT jde ještě dál tím, že celé sekce klasifikace jsou „syntetické“ a umožňují, aby se čísla tříd vytvářela podle potřeby. Takto získávají enumerativní klasifikace něco z povahy fasetových klasifikací probíraných dále. Popisy klasifikace MDT a jejího posledního vývoje najdete u Mcllwaineho (1997) a Slavice, Cordeira a Riesthuise (2008).
„Nejčistším“ příkladem velké enumerativní klasifikace je klasifikace Library of Congress (LCC) mající omezené syntetické možnosti. Nehledě na to a na fakt, že byla navržena pro jednu specifickou národní knihovnu, je velmi populární u vědeckých knihoven po celém světě, zatímco klasifikace, kterou používá National Library of Medicine, což je v podstatě podmnožina LCC, se hojně používá ve zdravotnických knihovnách. Ve Velké Británii se obecně používá lokální varianta zvaná Wessexská klasifikace. LCC poskytuje velmi detailní, někdy idiosynkratické seznamy oborů založené na enumeraci v rámci 19 hlavních tříd, například H označuje sociální vědy a R lékařství a jsou v nich třeba takovéto záznamy:

HQ9261 náprava a kultivování dospělých vězňů
R601-602 potraviny a nabídka potravin ve vztahu k veřejnému zdraví

Podrobný popis LCC podává Chan (2007, 14. kapitola).

Analyticko-syntetické klasifikace, takzvané fasetové klasifikace, jsou určeny pro klasifikaci komplexních materiálů na vysoké úrovni předmětové specifikace. Terminologie je seskupena do příslušných pojmů fasetovou analýzou (odtud název analytická), ze kterých se konstruuje klasifikace pro jakoukoliv jednotku (odtud název syntetická), což znamená, že tyto klasifikace si mohou poradit s novými pojmy tak, jak to nedovedou enumerativní schémata (i když, jak bylo řečeno výše, enumerativní schémata získávají syntetické schopnosti).
Fasetové klasifikace vymyslel indický knihovník S. R. Ranganathan ve 30. letech minulého století. Ranganathanova dvojtečková klasifikace (podle interpunkčního znaménka, které charakterizuje její notaci), první univerzální klasifikace tohoto typu, se kromě Indie používá jen málo; přehled tohoto schématu uvádějí Satija a Singh (2010). Ranganathanovy myšlenky však vedly k vytvoření mnoha fasetových klasifikačních schémat ve specifických oborových oblastech zejména ve vědě, technice a sociálních vědách. Tento typ schématu byl velmi populární během 50. a 60. let minulého století při klasifikaci specializovaných nebo technických materiálů, zvláště v systémech používajících technik automatizované dokumentace a později počítačových aplikací.
Ztratily však velmi rychle oblibu kvůli své složitosti a potížím při používání − jejich notace jsou pro náhodného uživatele velmi nepohodlné, nehodí se dobře k fyzickému uspořádání materiálů a jsou dnes poměrně málo používané. Jediné větší schéma tohoto druhu, kromě dvojtečkové klasifikace, které se stále vyvíjí, je druhé vydání Blissova třídění (BC2), první vydání bylo enumerativním schématem a bylo kompletně přepracováno. Považuje se sice za vlivné a teoreticky platné schéma, ale opět je málo používané (Broughton, 2010).
Přejdeme teď k druhému hlavnímu typu řízeného slovníku − abecednímu.

Abecední slovníky: předmětová hesla a tezaury
Existuje několik druhů abecedních řízených slovníků, ve kterých jsou termíny (slova nebo fráze) seřazeny podle abecedy. Nejsou mezi nimi žádné dobře definované rozdíly. Obecně se liší tím,

kolik informací je u každého termínu uvedeno, včetně jeho definice a vzájemných vztahů s ostatními termíny.
Seznamy klíčových slov jsou nejjednodušší formou abecedního řízeného slovníku. Často se skládají prostě jenom ze seznamu „schválených termínů“, někdy uvádějí synonyma. Jedná se o jednoduchou a nenákladnou formu, která se snadno vytváří a používá, ale její užitečnost je značně omezená a dá se použít pouze u velmi malých souborů a uživatelů s nepříliš náročnými potřebami.
Předmětová hesla jsou seznamy termínů (často značně dlouhé, definují-li složité pojmy), které se používají pro indexaci, vyhledávání a někdy i pro prohlížení. Všechny obecně obsahují synonyma a někdy také hierarchické odkazy a odkazy typu „VIZ TAKÉ“. Komplexní předmětová hesla mohou být velmi podobná tezaurům. Na druhé straně, jednodušší formy jsou jen o málo víc než seznamy klíčových slov. Nejběžněji se používají v situacích, kdy má každý záznam pouze jedno nebo několik málo hesel, například u knihovních databází a bibliografií.
Nejznámější terminologie tohoto druhu, *The Library of Congress Subject Headings* (LCSH), poskytuje předmětovou indexaci v mnoha knihovních databází a stále více se uplatňuje v digitálním prostředí. Jedná se o velmi rozsáhlý slovník poprvé uveřejněný v roce 1914, který má přes 270 000 termínů pokrývajících všechny věcné oblasti. Příklady termínů jsou:

Halloweenské kuchařky,
Hlemýždi jako přenašeči nemocí,
Ženy s nadváhou v umění,
Elektronické rezervní fondy v knihovnách,
Virová onemocnění u dětí,
Práce duchovních s generací populační exploze,
Lety do vesmíru na poštovních známkách.

Uvedené příklady ukazují, jak tato hesla spolu propojují několik pojmů. Podrobnější přehled LCSH, uvádí Broughton (2012) a možností jeho aplikace u digitálních zdrojů Walsh (2011) a Yi a Chan (2010).

Tezaury mají zvláštní důležitost, protože jsou propracovanou formou terminologie, která může být velmi vhodná pro poskytování efektivního přístupu k digitálním informacím. Jsou také jedinou selekční terminologií definovanou národními a mezinárodními standardy, i když ne všechny slovníky nazývané tezaury se takovými předpisy řídí. Obšírný výklad o tezaurech uvádí Broughton (2006b) a základní principy s podrobnými datovými údaji Aitchison, Gilchrist a Bawden (2000).
Tezaury jsou seznamy termínů s uvedenými vzájemnými vztahy mezi nimi. Používají se různé vztahy, ale standardní množinou (tedy doslova množinou definovanou příslušnými mezinárodními standardy ISO 2788 pro jednojazyčné tezaury) je:

SY – synonymum,
BT – širší termín,
NT – užší termín,
RT – asociovaný termín.

Další užitečnou relací je základní termín vrcholový deskriptor (Top Term) nebo nadpisová položka identifikující hierarchii, ve které se termín vyskytuje.
Jeden z množiny synonym je „preferovaný termín“, což vytváří nesymetrickou relaci užij/ekvivalent. Všechny ostatní relace jsou obecně symetrické. Standard také vyžaduje poznámky o rozsahu, poznámky s definicemi nebo vysvětlením deskriptorů a/nebo předepisující jejich použití při indexaci. Příklad deskriptoru v tezauru je uveden na Obrázku 6.3 na další stránce.
Tezaury se obecně používají jak pro indexaci, tak pro vyhledávání, ale je možno je použít i ve formě „indexačního tezauru“ (poskytuje navíc termíny umožňující volné vyhledávání v textu) nebo „vyhledávacího tezauru“ (navrhuje rešeršérovi navíc další termíny, když se dotazuje ve fulltextové databázi).
Proces vytváření tezauru zahrnuje obsahovou analýzu oblasti, podobně jako vytváření fasetové klasifikace. Tezaurus a klasifikace se vskutku mohou považovat za dvě „tváře“ stejného schématu. Obvykleji se struktura klasifikace použije jednoduše jako kostra pro odvození deskriptorů s jejich vzájemnými vztahy, které budou tezaurus tvořit.
Teď, když jsme probrali metody popisu zdrojů a jejich věcného obsahu, můžeme tuto kapitolu uzavřít tím, že se podíváme na dva dlouho zavedené způsoby organizace a řízení informací − referování (tvorbu abstraktů a resumé) dlouhých článků a indexaci dokumentů a fondů.

Referování
Referát nebo abstrakt je „krátká, ale přesná reprezentace dokumentu“ (Lancaster, 2003), nebo formálněji podle příslušného standardu ISO: „zkrácená přesná reprezentace obsahu dokumentu, bez přidané interpretace nebo kritiky a bez rozlišení toho, kdo abstrakt napsal.“
Referování má dlouhou historii, resumé článků se už dlouho používají k udržení kontaktu s vědeckou, zejména lékařskou a odbornou literaturou. Podle Borka a Berniera (1975) se dají jeho počátky vysledovat až do klasického období s prvními prokazatelně referenčními časopisy a oddíly pro abstrakty v běžných časopisech už v 17. století. Jejich používání se velmi rozšířilo v 19. století spolu s vytvořením oborově specializovaných referenčních a indexačních služeb, např. *Index Medicus* a *Chemical Abstracts*. Přehledy kromě knihy od Borka and Berniera najdete u Chowdhuryho (2010, 8. kapitola) a Koltaye (2010), Alonso a Fernandez (2010) uvádějí konceptuální model pro studium abstraktů a referování.
Abstrakty mohou být odlišné v mnoha aspektech, a tak mohou být různě kategorizovány. Podstatným rozdílem je, zdali jsou informativní − s dostatkem informací, aby mohly nahradit originál, nebo indikativní, pouze s tolika informacemi, aby se mohl čtenář rozhodnout, zdali má pro něj článek význam. Mohou se také lišit:

- délkou, od „mnohomluvných“ a „doslovných“ po strohé a „telegrafické“,
- rozsahem kritiky a interpretace originálu,
- mírou, do jaké jsou tendenční nebo zaměřené na konkrétní oblast zájmu či typ uživatelů, nebo na druhé straně jak moc usilují o nestrannost a objektivnost.

Psaní abstraktů se řídí mezinárodním standardem ISO 214 (1976) *Abstrakty pro publikace a dokumentaci*. Vydavatelé a producenti databází mají rovněž de facto standardy; viz například Montesi a Owen (2007). Existuje obecná tendence k tomu, aby byly abstrakty „strukturovanější“

s konzistentním souborem přítomných prvků. Stále existují spory o nejefektivnějším způsobu, jak učinit referáty užitečnější; viz například Hartley a Betts (2008, 2009), Zhang a Liu (2011) a Ripple a kol. (2011). Obzvláště nyní, kdy se na abstrakty spoléhá kvůli informačnímu přetížení stále více lidí, to má mimořádný význam (Nicholas, Huntington a Jamali, 2007). Tento stav je současně zdrojem znepokojení, protože podle studií typicky 20 % referátů obsahuje závažné nepřesnosti − většinou prezentují předmět hlavního dokumentu v nepřiměřeně pozitivním světle.

Po mnoho let je cílem vyvinout automatické referování. Je tomu tak od doby, kdy se objevil první výzkum na toto téma od H. P. Luhna koncem 50. let minulého století. Dosud bylo vynaloženo značné úsilí a použito mnoha přístupů k designu systémů automatického referování; přehled podává Chowdhury (2010), nicméně většina praktického referování v informačních systémech a službách je převážně intelektuálním úkolem.

Indexace a tagování

Rejstříkem se obvykle rozumí systematické uspořádání záznamů určených k tomu, aby usnadňovaly uživatelům lokalizaci informací v dokumentu nebo souboru dokumentů. Proces vytváření takových rejstříků se nazývá indexace a ti, kteří je dělají, jsou indexátoři. Řídí se mezinárodními standardy ISO 5963 (1985) *Metody analýzy dokumentů, určování jejich obsahu a výběru lexikálních jednotek selekčního jazyka* a ISO 999 (1996) *Informace a dokumentace – Zásady zpracování, uspořádání a grafické úpravy rejstříků*. Klasickým textem o indexačním procesu je Lancaster (2003), přehled teoretického a historického základu najdete u Weinberga (2009).

Indexací se obvykle rozumí přiřazování několika termínů z abecedního slovníku na rozdíl od klasifikace, ve které se používá systematický slovník. Není-li použit žádný řízený slovník, proces se nazývá „volná indexace termínů“. I když dovede zachytit velmi přesnou a aktuální terminologii, používání izolovaných klíčových slov znamená, že se nemohou vizualizovat žádné sémantické relace mezi termíny. Značnou pozornost získalo v poslední době volné tvoření klíčových slov, jako je „folksonomie“ nebo „sociální tagování“, kde uživatelé volně indexují internetové materiály, například webové stránky, fotografie, video sekvence a katalogové záznamy pro knihu (Ding a kol., 2009; Mai, 2011; Park, 2011; Voorbij, 2012). Není však jasné, jaká je jeho dlouhodobá užitečnost a jak ho provázat nebo jím doplnit tradiční intelektuální indexaci. Někteří informační profesionálové radí hledat způsoby, jak kombinovat tagování s řízenými slovníkovými strukturami; viz například Sauperl (2010) pro UDC a Yi a Chan (2009) pro LCSH .

Existuje mnoho druhů rejstříků − „rejstříky na poslední stránce“, rejstříky obsahů čísla nebo ročníku časopisu, kumulativní rejstříky k časopisům, novinám a magazínům, databázové rejstříky a rejstříky stránek na internetu nebo na intranetech. Rejstřík se může vztahovat k souboru jednotek, např. při indexaci databáze, intranetu nebo internetového slovníku, nebo na jedinou jednotku, např. obsah knihy, zprávy nebo ročníku časopisu. V případě souborů, rejstřík navede uživatele na konkrétní jednotku v rámci souboru. V případě jediné jednotky rejstřík odkazuje na stránku nebo kapitolu v rámci jednotky.

Hlavní zásady pro tvorbu rejstříku, tj. „indexační proces“, jsou většinu stejné bez ohledu na typ rejstříku. Indexace spočívá v prvé řadě na rozhodnutí o čem indexovaná jednotka je (obsahová analýza) a potom, jakými výrazy nejlépe tento obsah reprezentovat (volba termínu). Indexátor

musí rozpoznat abstraktní pojmy a potom je přiřadit k příhodným termínům. Zde spočívá rozdíl od jednoduché automatické fulltextové indexace, při které se termíny vybírají z dokumentu na základě konkordance (vybrána jsou všechna slova), statistické analýzy (vyberou se termíny, které se vyskytují s relativně největší četností) nebo poziční analýzy (jsou vybrány termíny vyskytující se v titulu, abstraktu, prvním odstavci atd.). Tato poslední analýza může být provedena velmi snadno počítačem, protože se zpracovávají vlastně pouze textové řetězce, bez nutnosti identifikace podstatných pojmů. Ale právě kvůli konceptuální analýze je indexování záležitostí lidského intelektu. Důvodem pro další používání lidské indexace místo levnější automatické je, že díky lidské analýze je rejstřík užitečnější a použitelnější. Není však obzvlášť konzistentní, mnoho studií prokázalo, že nejlepší stupeň konzistence, kterého mohou lidští indexátoři dosáhnout, nepřevyšuje 50 %.
Lidská indexace má kromě konceptuální analýzy řadu značných výhod před fulltextovým vyhledáváním nebo vytvářením konkordancí počítačem, což je v podstatě totéž. Pro lidského indexátora nepředstavují problém:

homografy – slova, která se stejně píší, ale mají různé významy,
synonyma – různá slova s týmž významem,
inference – je zamlčen podmět,
rozdíly mezi významnou a triviální „přechodnou“ zmínkou nějakého slova.

Automatické indexační systémy jsou sice stále výkonnější, ale se všemi výše uvedenými aspekty mají problémy. Existuje řada počítačových programů pro automatické vytváření rejstříků nebo na pomoc lidským indexátorům; například Cindex, Macrex nebo Sky Index (Coates, 2009).
Intelektuální indexaci konkrétního dokumentu lze považovat za proces skládající se ze tří částí:

- pochopení obsahu – indexátor rozhodne, „o čem“ dokument je,
- analýza obsahu – indexátor rozhodne, které pojmy se mají indexovat a do jaké hloubky,
- přeložení pojmů do jazyka rejstříku – indexátor zvolí nejvhodnější termíny z používaného abecedního slovníku (např. tezaurus nebo seznam předmětových hesel).

Dvě obecně přijímané zásady indexace jsou:

- zahrnout všechny pojmy, o kterých se lze domnívat, že budou uživatele rejstříku zajímat,
- indexovat na nejspecifičtější úrovni, kterou umožní indexační slovník.

Jedná se o dost obecné myšlenky, které indexátor musí nějak interpretovat. Dvě hlavní kritéria pro indexaci jsou:

- úplnost – rozsah, ve kterém jsou zahrnuty všechny možné pojmy,
- hloubka – stupeň specifičnosti s jakou jsou pojmy popisovány.

Indexace, která je úplná i hluboká se obvykle považuje za „zlatý standard“, ale za určitých okolností mohou být vhodné i ostatní tři možnosti – hluboká, ale ne úplná atd.

Shrnutí

Nástroje pro organizování informací se mění pomalu, některé z těch, které se používají dnes, více než sto let od doby zavedení, by Melville Dewey a Anthony Panizzi docela dobře poznali. Dramatické změny technických prostředků, kterými se vytvářejí a rozšiřují dokumenty, se nesetkaly s odpovídajícími racionálními prostředky, kterými jsou zpracovány. Budou-li vynalezeny automatizované metody, které by tuto práci zastaly, a zdali se nakonec folksonomie a podobné prostředky ukážou jako životaschopné, je otázkou.

- Organizování informací je stále jádrem informační vědy a jeho důležitost se v novém informačním prostředí ještě zvětšuje.
- Aby bylo možno zpracovat nové formy dokumentů, objevily se nové formy popisných metadat, věcný popis se ale změnil velmi málo.
- Je potřeba najít rovnováhu mezi expertními vstupy lidí, automatizovanými procesy a sociálním tagováním.
- Přes značný technický pokrok jsou teorie a pojmy formulované před dlouhou dobou stále důležité.

Další čtení

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge MA: MIT Press.

− Zevrubné pojednání o základní problematice.

Antoniou, G. & Harmelen, F. (c2004). *A semantic Web primer*. Cambridge MA: MIT Press.

− Srozumitelná analýza nových forem organizování informací.

Literatura

Aitchison, J., Gilchrist, A. and Bawden, D. (2000) *Thesaurus construction and use: a practical manual* (4th edition), London: Aslib.

Allinson, J. (2012) OpenART: open metadata for art research at the Tate, *Bulletin of the American Society for Information Science and Technology*, 38(3), 43–8.

Alonso, M. I. and Fernandez, L. M. (2010) Perspectives of studies on document abstracting: towards an integrated view of models and theoretical approaches, *Journal of Documentation*, 66(4), 563–84.

Anhalt, J. and Stewart, R. A. (2012) RDA simplified, *Cataloging and Classification Quarterly*, 50(1), 33–42.
Antoniou, G. and van Harmelen, F. (2008) *A semantic web primer* (2nd edn), Cambridge MA: MIT Press.

Ayre, C., Smith, I. A. and Cleeve, M. (2006) Electronic library glossaries: jargonbusting essentials or wasted resource?, *Electronic Library*, 24(2), 126–34.

Beghtol, C. (2010) Classification theory, *Encyclopedia of Library and Information Sciences* (3rd edn), Abingdon: Taylor & Francis, 1045–60.

Borko, H. and Bernier, C. L. (1975) *Abstracting concepts and methods*, New York NY: Academic Press.

Bowker, G. C. (2005) *Memory practices in the sciences*, Cambridge MA: MIT Press.

Bowker, G. C. and Star, S. L. (2000) *Sorting things out: classification and its consequences*, Cambridge MA: MIT Press.

Bowman, J. (2003) *Essential cataloguing*, London: Facet Publishing.

Bowman, J. (2005) *Essential Dewey*, London: Facet Publishing.

Bowman, J. H. (2006) The development of description in cataloguing prior to ISBD, *Aslib Proceedings*, 58(1/2), 34–48.

Broughton, V. (2005) *Essential classification*, London: Facet Publishing.

Broughton, V. (2006a) The need for a faceted classification as the basis of all methods of information retrieval, *Aslib Proceedings*, 58(1), 49–72.

Broughton, V. (2006b) *Essential thesaurus construction*, London: Facet Publishing.

Broughton, V. (2010) *Bliss Bibliographic Classification Second Edition*, *Encyclopedia of Library and Information Sciences*, Abingdon: Taylor & Francis, 1:1, 650–9.

Broughton, V. (2012) *Essential Library of Congress Subject Headings*, London: Facet Publishing.

Buckland, M.K. (2008) Reference library service in the digital environment, *Library and Information Science Research*, 30(2), 81–5.

Cassell, K. A. and Hiremath, U. (2011) *Reference services in the 21st century* (2nd edn revised), London: Facet Publishing.

Chan, L. M. (2007) *Cataloguing and classification: an introduction* (3rd edn), Lanham MD: Scarecrow Press.

Chowdhury, G. G. (2010) *Introduction to modern information retrieval* (3rd edn), London: Facet Publishing.

Chowdhury, G. G. and Chowdhury, S. (2007) *Organizing information: from the shelf to the web*, London: Facet Publishing.

Coates, S. (2009) Software solutions, *Indexer*, 27(4), 168–72.

Coyle, K. and Hillman, D. (2007) Resource Description and Access (RDA). Cataloguing Rules for the 20th Century [sic], *D-LIB Magazine*, 13(1/2), Jan/Feb 2007, available from: http://www.dlib.org/dlib/january07/coyle/01coyle.html.

Dimic, B., Milsavljevic, B. and Surla, D. (2010) XML schema for UNIMARC and MARC21, *Electronic Library*, 28(2), 245–262.

Ding, Y., Jacob, E. K., Zhang, Z., Foo, S., Yan, E., George, N. L. and Guo, L. (2009) Perspectives on social tagging, *Journal of the American Society for Information Science and Technology*, 60(12), 2388–2401.

Guerrini, M. (2009) In praise of the unfinished: the IFLA Statement of International Cataloguing Principles 2009, *Cataloguing and Classification Quarterly*, 47(8), 722–40.

Hartley, J. and Betts, L. (2008) Revising and polishing a structured abstract: is it worth the time and effort? *Journal of the American Society for Information Science and Technology*, 59(12), 1870–7.

Hartley, J. and Betts, L. (2009) Common weaknesses in traditional abstracts in the social sciences, *Journal of the American Society for Information Science and Technology*, 60(10), 2010–18.

Haynes, D. (2004) *Metadata for information management and retrieval*, London: Facet Publishing.

Heidorn, P. B. (2011) Biodiversity informatics, *Bulletin of the American Society for Information Science and Technology*, 37(6), 38–44.

Hider, P. (2012) *Information resource description: creating and managing metadata*, London: Facet Publishing.

Hitchens, H. (2005) *Dr Johnson's Dictionary – the extraordinary story of the book that defined the world*, London: John Murray.

HjØřrland, B. (2003) Fundamentals of knowledge organization, *Knowledge Organization*, 30(2), 87–111 .

HjØrland, B. (2008a) What is knowledge organization (KO)?, *Knowledge Organization*, 35(2/3), 86–101.

HjØrland, B. (2008b) Core classification theory: a reply to Szostak, *Journal of Documentation*, 64(3), 333–42.

HjØrland, B. (2009) Concept theory, *Journal of the American Society for Information Science and Technology*, 60(8), 1519–36.

HjØrland, B. (2011) The importance of theories of knowledge: indexing and information retrieval as an example, *Journal of the American Society for Information Science and Technology*, 62(1), 72–7.

HjØrland, B. and Pedersen, K. N. (2005) A substantive theory of classification for information retrieval, *Journal of Documentation*, 61(5), 582–97.

HjØrland, B. and Shaw, R. (2010) Concepts: classes and colligation, *Bulletin of the American Society for Information Science and Technology*, 36(3), 2–4, available from http://www.asis.org/Bulletin/Feb10/Bulletin_FebMar10_Final.pdf.

Hüllen, W. (2004) *A history of Roget's Thesaurus: origins, development and design*, Oxford: Oxford University Press.

Hunter, E. J. (2009) *Classification made simple* (3rd edn), Aldershot: Ashgate.
IFLA (1998) Functional Requirements for Bibliographic Records, produced by the IFLA, Study Group on the Functional Requirements for Bibliographic Records, Munich: K. G. Saur: available from http://www.ifla.org/VII/s13/frbr/frbr.pdf.

Koltay, T. (2010) *Abstracts and abstracting: a genre and skills for the 21st century*,Oxford: Chandos.

La Barre, K. (2010) Facet analysis, *Annual Review of Information Science and Technology*, 44, 243–86.

Lambe, P. (2007) *Organizing knowledge: taxonomies, knowledge and organisational effectiveness*, Oxford: Chandos.

Lancaster, F. W. (2003) *Indexing and abstracting in theory and practice* (3rd edn),London: Facet Publishing .

Langridge, D. W. (1992) *Classification: its kinds, systems, elements, and applications*, London: Bowker-Saur.

Leigh, G. J. (2011) *Principles of Chemical Nomenclature 2011: a guide to IUPAC recommendations*, London: Royal Society of Chemistry.

Mai, J-E. (2011) Folksonomies and the new order: authority in the digital disorder, *Knowledge Organization*, 38(2), 114–22.

McIlwaine, I. C. (1997) The Universal Decimal Classification: some factors concerning its origins, development and influence, *Journal of the American Society for Information Science*, 48(4), 331–9.

Mersky, R. M. (2004) The evolution and impact of legal dictionaries, *Legal Reference Services Quarterly*, 23(1), 19–35.

Miller, S. J. (2011) *Metadata for digital collections: a how-to-do-it manual*, London: Facet Publishing.

Montesi, M. and Owen, J. M. (2007) Revision of author abstracts: how it is carried out by LISA editors, *Aslib Proceedings*, 5991, 26–45.

Mugglestone, L. (2005) *Lost for words: the hidden history of the Oxford English Dictionary*, New Haven CT: Yale University Press.

Mugglestone, L. (2011) *Dictionaries: a very short introduction*, Oxford: Oxford University Press.

Nicholas, D., Huntington, P. and Jamali, H. R. (2007) The use, users, and role of abstracts in the digital scholarly environment. *Journal of Academic Librarianship,* 33(4), 446–53.

Oliver, C. (2010) *Introducing RDA: a guide to the basics*, London: Facet Publishing.

Park, H. (2011) A conceptual framework to study folksonomic interaction, *Knowledge organization*, 38(6), 515–29.

Patton, G. E. (ed.) (2009) *Functional requirements for authority data: a conceptual model*, Munich: K. G. Saur.

Pisanski, J. and Žumer, M. (2010) Mental models of the bibliographic universe. Part 1: mental models of descriptions, *Journal of Documentation*, 66(5), 643–67.

Ripple, A. M., Mork, J. G., Knecht, L. S. and Humphries, B. L. (2011) A retrospective cohort study of structured abstracts in MEDLINE, 1992–2006, *Journal of the Medical Library Association*, 99(2), 160–2.

Robinson, L. (2010) *Understanding healthcare information*, London: Facet Publishing.

Salaba, A., Zeng, M. L. and Žumer, M. (eds) (2011) *Functional requirements for subject authority data (FRSAD): a conceptual model*, Munich: de Gruyter.

Satija, M. P. (2007) *The theory and practice of the Dewey Decimal Classification system*, Oxford: Chandos.

Satija, M.P. and Singh, J. (2010) Colon Classification (CC), *Encyclopedia of Library and Information Sciences* (3rd edn), Abingdon: Taylor & Francis, 1:1, 1158–68.

Šauperl, A. (2010) UDC and folksonomies, *Knowledge Organization*, 37(4), 307–17.

Slavic, A. (2011) Classification revisited: a web of knowledge, in Foster, A. and Rafferty, P. (eds), *Innovations in information retrieval*, London: Facet Publishing, pp 23–48.

Slavic, A., Cordeiro, M. I. and Riesthuis, G. (2008) Maintenance of the Universal Decimal Classification: overview of the past and preparation for the future, *International Cataloguing and Bibliographical Control*, 37(2), 23–9.

Smiraglia, R. P. (2001) *The nature of a work: implications for the organization of knowledge*, Lanham MD: Scarecrow Press.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L. and Rosse, C. (2005) Relations in biomedical ontologies,*Genone Biology*, 6(5):r46 [online] available at http://genomebiology.com/content/pdf/gb-2005-6-5-r46.pdf.

Svenonious, E. (2000) *The intellectual foundation of information organization*, Cambridge MA: MIT Press.

Tackabery, M. K. (2005) Defining glossaries, *Technical Communication*, 52(4), 427–33.

Taylor, A. G. (2004) *The organization of information* (2nd edn), Santa Barbara CA: Libraries Unlimited.

Tennis, J. T. (2008) Epistemology, theory and methodology in knowledge organization: towards a classification, metatheory and research framework, *KnowledgeOrganization*, 35(2/3), 102–12.

Tillett, B. B. and Cristán, A. L (2009) *IFLA cataloguing principles: the statement of international cataloguing principles (ICP) and its glossary in 20 languages*, Munich: K. G. Saur. Also available online at http://www.ifla.org/publications/statement-ofinternational- cataloguing-principles.

Voorbij, H. (2012) The value of LibraryThing tags for academic libraries, *OnlineInformation Review*, 36(2) [online EarlyCite file.].

Walsh, J. (2011) The use of Library of Congress Subject Headings in digital collections, *Library Review*, 60(4), 328–43.

Weinberg, B. H. (2009) Indexing: history and theory, *Encyclopedia of Library andInformation Sciences* (3rd edn), Abingdon: Taylor & Francis, 1:1, 2277–90.

Welsh, A. and Batley, S. (2012) *Practical cataloguing: AACR, RDA and MARC21*,London: Facet Publishing.

Yi, K. and Chan, M. L. (2009) Linking folksonomy to Library of Congress subject headings: an exploratory study, *Journal of Documentation*, 65(6), 872–900.

Yi, K. and Chan, L. M. (2010) Revisiting the syntactical and structural analysis of Library of Congress Subject Headings for the digital environment, *Journal of the American Society for Information Science and Technology*, 61(4), 677–87.

Zeng, M. L. and Qin, J. (2008) *Metadata*, London: Facet Publishing.

Zhang, C.and Liu, X. (2011) Review of James Hartley's research on structured abstracts, *Journal of Information Science*, 37(6), 570–6.

Zhang, Y. and Salaba, A. (2009) *Implementing FRBR in libraries: key issues and future directions*, New York NY: Neal-Schuman.

Zins, C. and Santos, P. (2011) Mapping the knowledge covered by library classification schemes, *Journal of the American Society for Information Science and Technology*, 62(5), 877–901.