

NTK

50°6'14.083"N, 14°23'26.365"E

Národní technická knihovna
National Technical Library



Pravidla pro správu a aktualizaci Polytematického strukturovaného hesláře (PSH)

2012

ÚVOD

Polytematický strukturovaný heslář (PSH) je spravován referátem PSH v Národní technické knihovně (NTK) a jeho údržba probíhá v prostředí softwaru Aleph. Oproti verzi PSH 2.1 z ledna 2009, která obsahovala 13 570 hesel, verze PSH 3.0 obsahuje o 329 preferovaných hesel (deskriptorů) více, přesně 13 899 hesel¹. Kromě preferovaných hesel byla do hesláře přidána také celá řada nepreferovaných hesel (nedeskriptorů), jejichž počet několikanásobně převýšil počet nově vzniklých hesel PSH. Nová verze PSH tedy obsahuje nové deskriptory v češtině a k nim odpovídající hesla v angličtině, stejně jako nové nedeskriptory v češtině a nové nedeskriptory v angličtině. Verze PSH 3.0 se od svých předchozích verzí liší především způsobem, kterým byla aktualizace PSH provedena.

V letech 1993-1995 byla vypracována koncepce tvorby a obsahu Polytematického strukturovaného hesláře. Za pomoci odborníků byla v roce 1995 vytvořena zkušební verze PSH 0, která se stala základem testování ve Státní technické knihovně a několika dalších spolupracujících knihovnách. První oficiální verze PSH 1.0 byla představena v dubnu 1997. Za zmínku jistě stojí fakt, že PSH byl knihovnám a komerčním firmám poskytován za úhradu. Celý heslář v té době obsahoval přibližně 16 000 hesel ve 42 tematických řadách². Od té doby postupně vznikalo několik verzí, které vycházely z jednotlivých aktualizací hesláře. Poslední aktualizace proběhla v roce 2010 - 2011 a přinesla zatím poslední verzi PSH 3.0, která nově podléhá licenci Creative Commons Uvedte autora -Zachovejte licenci 3.0 Česko.

Správa a aktualizace PSH je úzce provázána se zásadami pro vytváření PSH. Z toho důvodu je nutné v procesu správy a aktualizace PSH vycházet i z dokumentu *Zásady pro vytváření Polytematického strukturovaného hesláře (PSH)*.

210 mm

PROČ AKTUALIZOVAT

Systémy organizace znalostí, jejichž zástupcem je také PSH, vyžadují pravidelnou aktualizaci. Rychlé zastarávání termínů je typickým problémem všech zástupců systémů organizace znalostí, které mají pevnou strukturu i obsah (a tedy i PSH). Tvorba takových systémů se proto stává nikdy nekončící činností.

Při aktualizacích PSH by se mělo přihlížet k účelu, pro který byl vytvořen. V tomto případě bylo cílem tvůrců vybudovat nástroj blízký přirozenému jazyku, jenž by doplnil nebo částečně nahradil stávající klasifikační systémy zejména v knihovnách s polytematickými fondy a ve finále také zkvalitnil vyhledávání v knihovních databázích. Za účelem zvýšení kvality vyhledávání se v hesláři kromě preferovaných termínů vyskytují také termíny nepreferované a asociované. Hesla PSH jsou uspořádána v rámci hierarchie, která nutně nekopíruje strukturu vědních oborů.

Systémy organizace znalostí vyžadují neustálou údržbu z mnoha různých důvodů. Těmi důvody jsou kupříkladu zpětná vazba uživatelů, nový obsah přinášející nová témata, některá témata

¹ Údaj z ledna 2011.

² Původních 42 tematických řad bylo rozšířeno do 44.

už se dále nevyvíjí, objevují se nové trendy nové populární výrazy, tzv. buzzwords a nové terminologie. Aktualizace taxonomie může představovat:

- tvorbu nového hesla,
- rozdělení existujícího hesla,
- smazání celého hesla,
- slučování hesel (smazání preferovaného znění hesla a předělání na nepreferované znění hesla),
- změnu formulace hesla,
- přiřazování nových vztahů ke stávajícím heslům,
- mazání existujících vztahů u stávajících hesel,
- změny existujících vztahů u stávajících hesel,
- přiřazování nepreferovaných znění ke stávajícím heslům,
- přiřazování poznámek ke stávajícím heslům,
- přiřazování dalších nových informací, atributů, kategorií apod. ke stávajícím heslům,
- přesun celé větve taxonomie na jiné místo v její hierarchii.

Intelektuální přístup k údržbě hesláře zahrnuje dva základní kroky - výběr vhodných termínů a jejich zařazení do struktury hesláře. Druhý krok je pak náročnější a vyžaduje hlubší znalost struktury PSH.

Klíčovým kritériem aktualizace PSH je zachování kontinuity. V hesláři by neměly být prováděny žádné výrazné změny, preferovány jsou pouze změny dílčího charakteru. Nedoporučuje se proto zbytečně měnit hierarchickou strukturu a vztahy mezi hesly a přesouvat pojmy mezi tematickými řadami. Stabilita je totiž důležitější než snaha o vylepšení detailů. Ne všechna hesla mohou být do nové verze přijata. Zejména je třeba dbát na skutečnost, že hesla příliš specifická a hesla patřící do hlubších vrstev logické struktury nemají v PSH místo. V případě nezbytnosti lze tato hesla zařadit do skupiny odkazů. V zásadě tedy platí, že při zařazování nových hesel do struktury by měla být dodržena úroveň obecnosti. Při aktualizaci PSH je téměř vyloučeno vyřazovat deskriptory použité ve starších verzích. V hesláři je možné provádět pouze tyto typy úprav - zařazování nových hesel (deskriptorů), která jsou výsledkem pokroku lidského poznání a tvorba vylučovacích a přidružovacích odkazů. Vylučovací odkazy se používají pro možné vyloučení synonym nebo významově příliš podřadných hesel. Přidružovací odkazy vyjadřují asociativní vztahy mezi pojmy tak, aby pomocí odkazů umožnily nalézt obsahově příbuzné dokumenty.

1. AKTUALIZACE PSH

V otázce časové frekvence aktualizace se střetávají dva protichůdné požadavky; na jedné straně by řízený slovník měl být stabilní, na straně druhé by ale měl být aktualizován dle potřeby. Aktualizace PSH probíhá průběžně. Nová hesla PSH jsou vytvářena v průběhu celého roku, stejně tak úpravy stávajících hesel či vzájemných vztahů. V prosinci, tedy na konci každého kalendářního roku, je vždy vyhlášena nová verze PSH, která je distribuována partnerským knihovnám a institucím.

Aktuální změny v PSH se ihned projevují v autoritní databázi Aleph. PSH Manager je aktualizován každý den. Balíček PSH/SKOS je aktuální každý den.

Aktualizace PSH probíhá na základě zpětné vazby oddělení věcného popisu, analýzy uživatelských dotazů (analýza logu vyhledávání v katalogu NTK), návrhů zaslaných prostřednictvím webového formuláře a na základě návrhů hesel akademických pracovníků Univerzity Pardubice.

Proces aktualizace hesláře je optimalizován několika způsoby. Jedním z nich je export a následná analýza klíčových slov použitých při věcném popisu knih v NTK (obsah pole 653 v bibliografickém záznamu ve formátu MARC 21). Pole 653 je v NTK používáno pouze doplňkově, a to v případech, kdy danou skutečnost nelze dostatečně přesně vyjádřit prostřednictvím hesel PSH (jedná se o podrobnější úroveň zpracování tématu) nebo se v hesláři odpovídající heslo nevyskytuje (jedná se o nový termín).

Doplňkově může být provedena analýza a následný výběr nových hesel PSH z pole 653 v bibliografickém záznamu ve formátu MARC 21, které je používáno při věcném popisu knih v NTK. Toto pole je v NTK používáno pouze doplňkově, a to v případech, kdy danou skutečnost nelze dostatečně přesně vyjádřit prostřednictvím hesel PSH (jedná se o podrobnější úroveň zpracování tématu) nebo se v hesláři odpovídající heslo nevyskytuje (jedná se o nový termín).

1.1. Návrhy z oddělení věcného popisu

1.2. Analýza uživatelských dotazů

Aktualizace PSH vychází také z analýzy dotazů položených uživateli při vyhledávání v katalogu NTK.

Cílem algoritmu jsou termíny zadávané při vyhledávání ve všech polích, v poli předmětovém a v heslech PSH. Získané termíny se periodicky přidávají do výstupního souboru. Při vyšší frekvenci některého z nich je heslo pracovníky referátu PSH individuálně posuzováno jako kandidát na nové heslo PSH. Pracovníci hodnotí, zda heslář PSH obsahuje podobný termín a kolik a jaké dokumenty OPAC po položení dotazu nalezne. Zhruba třetina z 1653 uživatelských dotazů hodnocených v první testovací analýze představovala vyhledávání podle klíčových slov a předmětových hesel. Mezi těmito uživatelskými dotazy bylo objeveno 12 nových deskriptorů a 49 nedeskriptorů. Ostatní dvě třetiny uživatelských dotazů představovaly zejména hledání dokumentu podle názvu, autora (o osobě), či nepřesně zadané dotazy.

Neočekávaným výsledkem analýzy bylo zjištění, že v bibliografické bázi v poli 630, které je v NTK je vyhrazeno pro produkty (typickými příklady jsou např. hesla “Škoda Octavia” nebo “Windows 7”), se vyskytují termíny, jež neoznačují žádné konkrétní výrobky, produkty ani služby (např. FTP, SEO, SOA) a které je tudíž nutné zahrnout do hesláře PSH. Analýza tak dala podnět k hledání nových hesel v poli 630 a k revizi tohoto pole. Ačkoliv hlavním cílem analýzy bylo obohacení hesláře PSH, tak kromě nových hesel byly identifikovány i dokumenty, o které mají čtenáři zájem, které však ve fondu NTK chybí. Informace o takových dokumentech jsou proto předávány oddělení akvizice v NTK.

Analýza probíhá přes rozhraní, do kterého se nahrají vybrané soubory ze služby Google Analytics. Nahraná hesla jsou porovnána s hesly PSH a poté jsou srovnána se všemi zpracovanými hesly. Výsledkem je tabulka s příslušným vyhledávacím termínem a frekvencí.

Hodnoty tabulky jsou následně importovány do excelového souboru, ve kterém jsou dále hodnoceny. Hodnotí se ta hesla, jejichž frekvence přesahuje míru čtyř. V případě, že měsíční soubor obsahuje méně než dvě stě termínů (při frekvenci čtyři), nahrají se do souboru i termíny s frekvencí tři.

1.3. Návrhy na nová hesla - formulář

Odborná veřejnost a uživatelé mají možnost navrhnout nová hesla ať už v českém či anglickém jazyce prostřednictvím webového formuláře na stránkách NTK. Navržená hesla jsou dále odbornými pracovníky hodnocena a srovnávána s hesly v PSH. Inspiraci pro umístění hesel do PSH je možné hledat např. v systému Library of Congress Subject Headings.

1.4. Návrhy hesel a Univerzita Pardubice

PSH v současnosti neslouží pouze jako prostředek pro věcnou selekci bibliografických záznamů, ale jeho potenciál je rozvíjen také v oblasti indexace a zpřístupňování webových dokumentů. Jeho užití překračuje i hranice indexace dokumentů samotných, ať už tradičních nebo právě těch webových, a tak byl implementován i na Univerzitě v Pardubicích, kde slouží k indexaci odborných profilů vědecko-výzkumných pracovníků univerzity. Nový způsob využití PSH na této univerzitě přinesl nečekané spojení, které má pro NTK mnoho výhod. Do PSH jsou navrhována nová hesla PSH akademickými pracovníky univerzity, tedy odborníky z jednotlivých oborů. Zpětnou vazbu poskytují také ve formě návrhů na změny v PSH či poukazují na zastaralé termíny a nesrovnalosti ve struktuře. Spolupráce, která tímto jedinečným využitím PSH vznikla, je obohacením a přínosem pro obsah, ale i strukturu celého hesláře. Návrhy zasílají akademičtí pracovníci prostřednictvím webového formuláře pro návrhy na nová hesla PSH. Využívat jej mohou jak katalogizátoři oddělení věcného zpracování v NTK, tak externí uživatelé, např. zájemci o problematiku selekčních jazyků, ale i zákazníci NTK, kteří se prostřednictvím formuláře stávají tvůrci samotného obsahu PSH. Po vyplnění příslušných polí je tabulka odeslána pracovníkům Referátu PSH, kteří získané návrhy posuzují a zvažují jejich zařazení do struktury PSH. Ti dále ve spolupráci s katalogizátory NTK u získaných návrhů zvažují účelnost hesel, popř. kontrolují jejich duplicitu. Příkladem požadavku na vytvoření nového hesla PSH zasláno prostřednictvím webového formuláře je "urgentní lékařství", synonymum - "urgentní medicína", a také v angličtině "emergency medicine", "programming language C Sharp". Dalšími příklady byly návrhy na hesla - "gender history", "romistika" nebo "fotoakustické metody".

210 mm

1.5. Reindexace bibliografické databáze

V Národní technické knihovně se PSH používá od roku 1993 (indexovány jsou záznamy všech dokumentů publikovaných od r. 1990 a výběrově starší publikace). PSH implementován nejprve do systému.

V souvislosti s implementací nové verze je vždy nutné rozhodnout, jestli záznamy indexovány předchozí verzí hesláře zůstanou v původním stavu nebo budou novou verzí reindexovány.

Obecně je při rozhodování o provedení reindexace nutno vzít do úvahy následující varianty:

V rámci přípravy překladu českých hesel PSH bylo nutné nejprve vybrat relevantní zdroje pro jejich ověření. K tomuto účelu byly zvoleny tři hlavní zdroje - databáze národních autorit NK ČR, klasifikační web LCSH (jehož předplatné má NTK zajištěné) a Wikipedie. Doplňkově byl využit také vyhledávač Google, který byl použit zejména v případech, kdy bylo nutné rozhodnout mezi dvěma synonymními výrazy. V takovém případě byl porovnán výsledek výsledného počtu nalezení záznamů. pozn. : geologický slovník

Z praxe vyplynulo, že

V případě PSH hovoříme o aktualizacích na dvou úrovních. První z nich představuje aktualizaci (tedy tvorbu nových deskriptorů, nedeskriptorů, případně zrušení starších hesel apod.) autoritní báze NTK. Druhá rovina je reprezentována aktualizací prohlížení PSH včetně SKOS balíčku ke stažení.

2. PSH MANAGER ONLINE

PSH Manager byl vytvořen dle potřeb, návrhů a přání pracovníků věcné katalogizace v NTK za účelem zefektivnění práce s heslářem při indexaci dokumentů. PSH Manager je proto nástroj primárně určený indexátorům, kteří používají PSH k indexaci dokumentů a správcům PSH, kteří PSH Manager využívají při aktualizaci hesláře.

Výhodou PSH Manager je možnost užití funkcí našeptávače (český i anglický), která uživateli při vyplňování nabízí hledaný výraz odpovídající již napsané části textu, dále zobrazení historie prohlížení (3 poslední hesla) a funkce automatického odkazování na odpovídající hesla v portálu české Wikipedie³. Odkazy mohou indexátorům pomáhat při orientaci v terminologické oblasti, stejně jako mohou být užitečné při verifikaci konkrétních hesel.

Denně probíhá kontrola nových hesel a vždy jednou za týden se aktualizuje celá databáze. PSH Manager je denně aktualizován, pracovníkům referátu PSH mohou zpouštět i ruční aktualizace. V rozhraní PSH Manager je k dispozici několik ikon. Vyhledávání lze pomocí anglické vložky přepnout do anglické verze. První ikona na pravo od vyhledávacího okna přenesení uživatele po jejím stisknutí do katalogu NTK, kde se zobrazí ty dokumenty, které jsou označeny příslušným heslem právě zobrazeným v prostředí PSH Manager. Tlačítko se symbolem plus představuje odkaz na webový formulář pro návrh nového hesla PSH, k dispozici je tak pro případ, že by uživatel nenašel heslo v PSH, ale chtěl by podat návrh na jeho vytvoření. Tlačítko MARC zobrazí příslušné heslo v katalogu NTK v autoritním formátu MARC 21. V případech, kdy se uživatel potřebuje vrátit ke svému vyhledávání, je k dispozici funkce zobrazení historie prohlížení, která zaznamenává tři poslední vyhledávaná hesla. V neposlední řadě jsou v rozhraní zobrazena také loga PSH a NTK, která odkazují na stránky o PSH (pro informace o hesláři) a stránky NTK. V pravém rohu je rovněž umístěno datum a čas s údaji o poslední aktualizaci. V rámci rozhraní

PSH Manager je možné také stáhnout celý, vždy aktualizovaný heslář ve formátu SKOS.

ZÁVĚR

³ cs.wikipedia.org

V letošním roce bychom chtěli provést zkušební analýzu logu vyhledávání v katalogu některé z partnerských knihoven, tj. v knihovnách které využívají PSH a které mají jiné profilové zaměření fondu než NTK. Cílem by bylo obohatit PSH v oblastech, které nejsou, vzhledem k technickému zaměření NTK, tolik rozvíjeny.

V této souvislosti bude zhodnocena možnost analýzy logu vyhledávání v katalozích těchto knihoven za účelem rozšiřovat a aktualizovat heslář také v oblastech, které nespádají do profilového zaměření NTK (např. Ústřední knihovna Filozoficko-přírodovědecké fakulty Slezské univerzity v Opavě nebo knihovna Západočeského muzea v Plzni).

Správa a aktualizace PSH je úzce provázána se zásadami pro tvorbu PSH. Z toho důvodu je nutné v procesu správy a aktualizace PSH vycházet i z dokumentu *Zásady pro vytváření Polytematického strukturovaného hesláře (PSH)*.