

System M-CAST v českém kontextu

Marie Balíková / Národní knihovna ČR / e-mail: marie.balikova@nkp.cz

Abstrakt:

Vícejazyčný systém agregace informací (M-CAST, Multilingual Content Aggregation System) umožňuje tvorbu digitálních knihoven prostřednictvím agregace dat dostupných v různých formátech a z různých zdrojů. Systém M-CAST je založen na výsledcích projektu TRUST – Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (Vícejazyčný sémantický a kognitivní mechanismus vyhledávání textů využívající sémantické technologie, IST-1999-56416), který byl financován z Pátého rámcového programu EU pro rozvoj vědy a výzkumu Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies. Vyhledávací stroj TRUST umožňoval vyhledávání ve čtyřech jazycích (francouzštině, italštině, polštině a portugalštině). Systém je nyní obohacen o dva další jazyky: angličtinu a češtinu. Jazykové zdroje systému TRUST jsou aktualizovány a v systému nadále využity. Systém vyhledávání je obohacen o možnost zúžení dotazu pomocí standardního mezinárodního desetinného třídění (MDT), celosvětově používaného v knihovních systémech. V příspěvku je nastíněna možnost budoucího využití systému M-CAST v českém prostředí.

Abstract:

The Multilingual Content Aggregation System (M-CAST) will allow for the development of Digital Libraries by aggregating digital data available in different formats and locations. M-CAST will be based on the outcome of the TRUST – Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (IST-1999-56416) project financed by the 5th European Union Framework Programme for RTD. The TRUST search engine is developed for 4 languages (French, Italian, Polish and Portuguese). Two other languages modules will be added: English and Czech. The TRUST language resources will be used and updated. Searching system will enable to narrow the search using the standard Universal Decimal Classification (UDC) applied in library cataloguing systems worldwide. A potential application of the M-CAST system within the Czech context is discussed.

Úvodní poznámka

Příspěvek je přepracovanou a zkrácenou verzí studie Systém M-CAST (Multilingual Content Aggregation System – Vícejazyčný systém agregace informací) publikované ve sborníku Dotazování v přirozeném jazyce.

Příspěvek je rozdělen do několika základních tematických celků. Úvodní kapitola obsahuje základní charakteristiku systému M-CAST. Ve druhé kapitole se můžeme seznámit s architekturou systému, ve třetí je charakterizován proces indexace. Čtvrtá kapitola je věnována problematice vyhledávání, pátá kapitola přináší informace o testování a hodnocení systému M-CAST. V závěrečných kapitolách (6 a 7) jsou nastíněny možnosti aplikace systému v českém prostředí a možnosti jeho dalšího vývoje.

1 Charakteristika systému

Systém M-CAST je vícejazyčným indexovacím a vyhledávacím nástrojem umožňujícím tvůrcům obsahu integrovat a prohledávat rozsáhlé soubory textů včetně multimédií, jako

jsou internetové knihovny, informační zdroje nakladatelství a tiskových agentur nebo databáze vědeckých informací poskytující informační služby široké odborné i laické veřejnosti.

Představuje sofistikovaný dialogový (nehlasový) systém s prvky umělé inteligence založený na bázi sémantických technologií. Používá metodu dotazování v přirozeném jazyce, která je součástí technologie zpracování přirozeného jazyka a uživateli nabízí možnost klást/formulovat dotaz celou větou.

Zpracování přirozeného jazyka, jehož cílem je přesná a jednoznačná sémantická reprezentace textu, se odehrává ve všech jazykových rovinách: v morfologické, syntaktické, sémantické i pragmatické. Úspěšnost procesu vyhledávání je přímo závislá na kvalitě sémantické reprezentace textu, tokenizaci výchozích nestrukturovaných textů, kvalitní morfologické analýze a desambiguaci, tj. zjednoznačnění prvků přirozeného jazyka.

1.1 Systém M-CAST vícejazyčný

Systém M-CAST umožňuje v současné době sémantické vyhledávání v šesti jazycích: ve francouzštině, portugalštině, italštině, polštině, angličtině a češtině; nezávislost struktury formálního jazyka na konkrétních jazycích tím vystupuje do popředí.

Čeština je ve srovnání s ostatními jazyky, zejména s angličtinou, komplikovaným jazykem s bohatou morfologickou strukturou. Procesy jako např. lematizace, tj. převádění výrazů na základní slovníkový tvar, a derivace – metoda sloužící k tvorbě všech odvozenin ze základního slovníkového tvaru, dále pak morfologická, syntaktická i sémantická desambiguace jsou mnohem náročnější než v jiných jazycích.

1.2 Systém na bázi sémantických technologií

Systém M-CAST se řadí do rodiny interaktivních dialogových vyhledávacích systémů, jejichž konečným cílem je v podstatě opuštění umělých selekčních jazyků jakožto vyhledávacích nástrojů. Schopnost uživatele komunikovat s vyhledávacím systémem formou celých vět odpovídá současným trendům ve vývoji interaktivních vyhledávacích systémů na bázi sémantických technologií. Vyhledávání je jednoduché a rychlé. Uživatel nemusí znát žádný formální dotazovací jazyk, navíc existují dotazy, které se v přirozeném jazyce formulují snadno, jejich konstrukce je však v dotazovacích jazycích obtížná („negativní“ dotazy nebo dotazy s obecnou kvantifikací).¹

1.3 Dotazování v přirozeném jazyce

Systém M-CAST nabízí netradiční způsob vyhledávání v databázích primárních (ale i sekundárních) informačních zdrojů. Vyhledáváme-li v těchto databázích tradičním, „klasickým“ způsobem, zadaný výraz je vyhledán jako textový řetězec. Výsledkem dotazu je seznam dokumentů, ve kterých se zadané výrazy vyskytují. Nevýhodou tohoto postupu je, že uživatel je mnohdy přesycen množstvím odkazů na dokumenty, byť různým způsobem zhodnocených, které musí konzultovat, aby získal požadovanou informaci. Při netradičním způsobu vyhledávání (dotazování v přirozeném jazyce) je vyhledávaný výraz, kterým je v tomto případě celá věta, analyzován a obohacen o další sémantické informace ve strojem srozumitelné podobě. Výsledkem vyhledávání je v tomto případě jednoznačná konkrétní odpověď, nebo úryvek textu obsahující konkrétní odpověď a možnost zasazení těchto odpovědí do širšího kontextu (vizualizace zdrojové stránky).

1.4 Uživatelé systému

Systém M-CAST jako ostatní vyhledávací systémy na bázi sémantických technologií je určen široké veřejnosti i specialistům. Ukazuje se, že intenzivní využití sémantických technologií v procesu vyhledávání je perspektivní metodou jak pro analýzu dotazů a indexovaných textů, tak pro extrakci odpovědí.

¹ viz Strossa 2004

1.5 Mezinárodní kooperace

Systém M-CAST je výsledkem mezinárodní kooperace: na vývoji systému se podílely firmy Infovide-Matrix S.A. a TiP sp. z o.o. (Polsko), Synapse Développement SARL (Francie), Priberam Informática Lda. (Portugalsko), Expert System S.p.A (Itálie) a Vysoká škola ekonomická v Praze. Systém byl testován v Polské internetové knihovně a Národní knihovně České republiky.

2 Architektura systému M-CAST

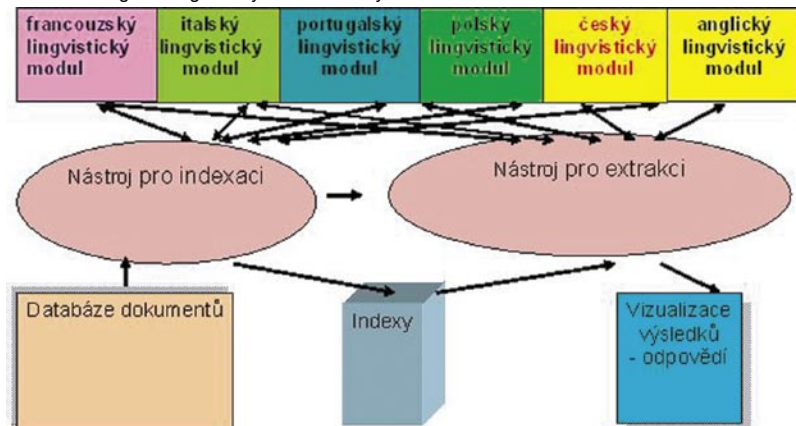
Pro kvalitu indexovacího a vyhledávacího systému založeného na sémantických technologiích jsou důležité tři komponenty: pro funkčnost systému je rozhodující velikost a reprezentativnost databáze indexovaných dokumentů, funkčnost a výkonnost jednotlivých komponentů indexovacího a vyhledávacího stroje M-CAST a funkčnost portálu M-CAP. Architektura systému je na všech stupních modulární. Po technické stránce vychází architektura systému M-CAST z tradiční třívrstvé architektury a obsahuje vrstvu klientskou, tj. browser, vrstvu aplikační, tj. web server a část business logic, a vrstvu datovou, tj. lingvistický procesor, který se skládá ze základních procesních prvků, tj. lingvistických modulů. Třívrstvý model podporuje vyšší úroveň stability; klient pracuje pouze s uživatelským rozhraním, datové a aplikační služby jsou od sebe odděleny do samostatných logických modulů. Jde o síťovou architekturu, kde komunikace mezi jednotlivými vrstvami je umožněna pomocí rozhraní webových služeb. Lingvistický procesor je dostupný pomocí rozhraní webových služby a může být používán jako vzdálený flexibilní zdroj. Portál M-CAST pracuje na platformách J2EE a Tapestry. Aplikace těchto technologií otevírá celou řadu možností integrace s jinými informačními technologiemi, usnadňuje další rozvoj systému a zmenšuje nároky na jeho údržbu.

2.1 Lingvistické moduly

Lingvistické moduly pro jednotlivé jazyky byly vytvářeny na sobě nezávisle; reflektují potřeby jednotlivých národních jazyků, sdílejí však základní obecné principy aplikované při automatizovaném zpracování přirozeného jazyka.

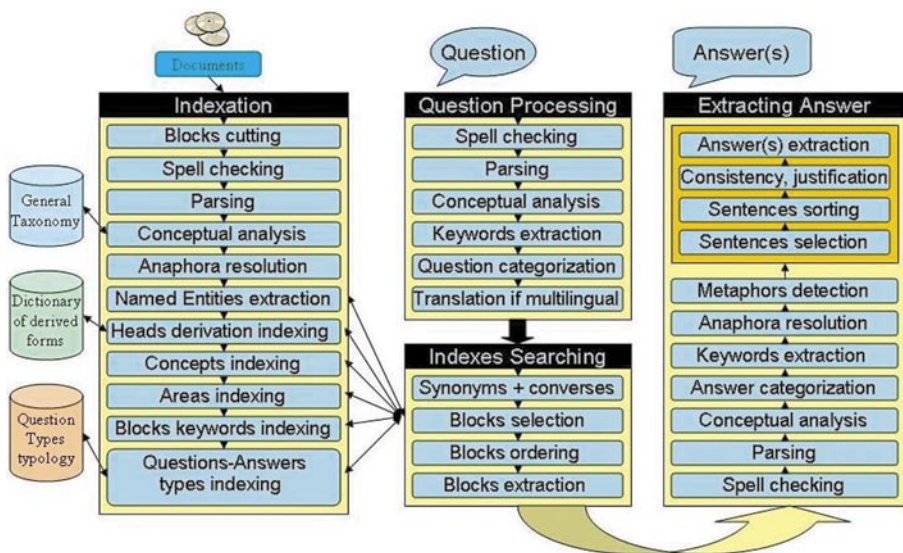
Kvůli podstatným rozdílům mezi jednotlivými jazyky byl zvolen otevřený design, který umožňuje připojení dalšího jazyka pomocí oddělených zásuvných modulů (plug-ins). Lingvistický procesor používá lingvistické moduly k vyhledání jazykově neutrální reprezentace dotazu. Tato jazykově neutrální reprezentace dotazu je pak použita při vlastním procesu vyhledávání v rámci lingvistického procesoru. Každý lingvistický modul používá odpovídající lingvistické zdroje.

Obr. č. 1: Integrace lingvistických modulů v systému M-CAST



Složitost (maximální ²) struktury lingvistických modulů je patrná z následujícího schématu pro francouzský lingvistický modul vyvinutý francouzským partnerem, firmou Synapse Développement:

Obr. č. 2: Francouzský lingvistický modul



2.1.1 Český lingvistický modul v projektu M-CAST

Český lingvistický modul byl vyvíjen paralelně s modulem polským. Na jeho vývoji se podílely polská firma TiP (garant polského modulu), portugalská Priberam Informática, jejíž programové nástroje byly použity pro zpracování obou nově připojených jazyků a Vysoká škola ekonomická v Praze. Základní lingvistickou datovou strukturu podporující funkce systému M-CAST představují formální definice předem vyčleněných kategorií dotazů a potenciálních odpovědí na ně v indexovaných textech. V systému M-CAST se rozlišuje 86 sémantických typů dotazů definovaných firmou Synapse Développement. Definice pro jejich rozpoznávání v češtině byly formulovány pomocí nástroje SintaGest (Priberam Informática).

Nutným předpokladem fungování vzorců pro rozpoznávání typů dotazů a odpovědí je morfologická analýza indexovaných textů a pokládaných dotazů, která se provádí pomocí morfologického analyzátoru.

Morfologický analyzátor češtiny používaný v českém lingvistickém modulu systému M-CAST a vyvinutý týmem doktorandů a studentů pod vedením Dr. P. Strossy je založen na slovníku, ve kterém je každému slovu přiřazen jeden ze vzorů ohýbání, a systému tabulek definujících jednotlivé vzory ohýbání. Z těchto dat je kompilována pracovní datová struktura efektivně uchovávajících a rozpoznávajících všechny tvary všech slov. Tabulky definující vzory mají odlišnou strukturu podle slovních druhů.³

² ve skutečnosti ne všechny lingvistické moduly uplatňují toto schéma v celé jeho komplexnosti

³ viz Strossa 2007, str. 43 a 45

2.2 Lingvistický procesor

Lingvistický procesor představuje technicky nejpokročilejší součástí systému M-CAST. Obsahuje modul pro zpracování dotazů a modul vyhledávací a jeho úkolem je zpracovat dotazy z portálu M-CAST.

2.2.1 Základní komponenty procesoru:

- SearchEngine.dll – dotazovací a vyhledávací stroj
- IndexEngine.dll – indexovací stroj
- DetectLanguage.dll – detektor jazyků
- Parsers.dll – konvertor různých formátů dokumentů – systém rozeznává velké množství formátů (.html, .xml, .txt, .doc, .dbx, .pdf, .ps atd.), čímž umožňuje indexaci nesmírného množství textů
- Databáze metadat dokumentů – je nedílnou součástí lingvistického procesoru, obsahuje metainformace o dokumentech

2.3 Webové služby systému M-CAST

Funkčnost lingvistického procesoru (LP) systému M-CAST je zajištěna pomocí technologie webové služby. Používají se dvě kategorie rozhraní: první rozhraní souvisí s indexovacím modulem LP. Umožňuje indexovat dokumenty vyhledané ve vzdálených repozitářích a zpřístupnit je efektivním způsobem v průběhu vyhledávání. Druhé rozhraní umožňuje zpracování dotazů přímo v indexovaných dokumentech v modulech LP. Rozhraní jsou definována jazykem WSDL.

2.4 Integrační vrstva Portálu M-CAST

Integrační vrstva systému M-CAST zajišťuje metadata, sběr dat (harvesting) a předání těchto údajů LP k indexování pomocí rozhraní webové služby.

V tomto modulu (v této vrstvě) probíhají dva procesy:

- Proces indexace – proces je iniciován administrátory systému M-CAST, případně je spouštěn automatizovaně Schedulerem. Proces indexace zahrnuje přenesení rejstříku dokumentů, které jsou velikostí významné, do LP. M-CAST portál obsahuje externí rozhraní, které zajistí přenos dat mezi externí databází a modulem LP. Toto rozhraní je dostupné pomocí webové služby a administrativní aplikace.

- Automatický sběr dat – tento proces přenáší pravidelně OIA indexy a stahuje zdroje popsané těmito indexy pomocí FTP a HTTP protokolu; modul pro sběr dat úzce kooperuje s programem Resource Manager a programem Scheduler.

Procesy jsou řízeny programem Scheduler, který je v podstatě srdcem systému M-CAST. Tato služba iniciuje a zajišťuje indexační proces v rámci systému. Program Scheduler organizuje pořadí jednotlivých úkolů, které musejí být realizovány v rámci systému M-CAST.

2.5 Prezentační vrstva systému M-CAST

Prezentační vrstva systému M-CAST je modul odpovídající za uživatelské rozhraní pro kladení dotazů a za komunikaci se zbývajícími moduly. Architektura modulu je složitá, protože modul musí být schopen zajistit několik různých procesů souvisejících se službami systému M-CAST:

- Proces vyhledávání – tento proces zahrnuje běžnou interakci uživatele se systémem. Prezentační vrstvy M-CASTu vytvářejí webové rozhraní, které umožňuje běžné dotazování. Uživatel M-CASTu má možnost volby dotazu a systém M-CAST realizuje dotaz pomocí vyhledávacího modulu za přispění programu Resource Manager a lingvistického procesoru prostřednictvím rozhraní webové služby.

- Vyhledávání ve více instancích systému M-CAST – v této verzi může uživatel M-CASTu prohledávat paralelně několik instancí M-CAST systému, např. instanci Ná-

rodní knihovny ČR a instanci Polské internetové knihovny (PBI). V tomto případě je uživatelův dotaz simultánně zpracováván v několika (předvolených) instancích systému M-CAST.

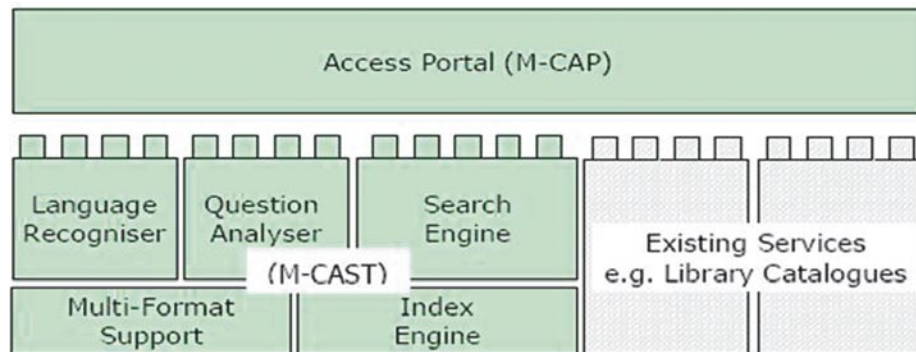
Schopnost portálu M-CAST vyhledávat ve více instancích poskytuje v podstatě neomezenou škálovatelnost (schopnost distribuovaného systému využívat dodatečné hardwarové zdroje pro uskutečnění většího počtu operací s daty).

2.6 Portál M-CAST/M-CAP

Portál M-CAST/M-CAP je schopen kombinovat různé aplikace a informační zdroje do jediné ucelené prezentace; uživatelé v různých rolích mohou vidět odlišný obsah dle svých přístupových oprávnění nebo svého profilu a mohou si obsah přizpůsobit.

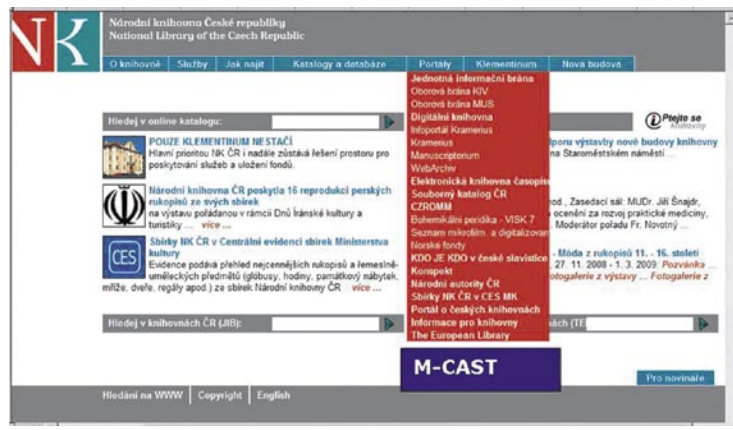
Portál M-CAST/M-CAP je modulární; je koncipován tak, aby umožnil integraci stávajících vyhledávacích nástrojů používaných v dané instituci. Existují dva způsoby integrace portálu M-CAST/M-CAP do procesu vyhledávání dané instituce. Jednak lze obohatit stávající vyhledávací možnosti o rešeršní strategii dotazování v přirozeném jazyce a zvolit portál M-CAST/M-CAP jako základní nástroj.

Obr. č. 3: Portál M-CAST/M-CAP jako základní nástroj



Je však možný i obrácený postup: Instituce zahrne portál M-CAST/M-CAP jako součást stávajícího portálu.

Obr. č. 4: Ukázka možného zapojení systému M-CAST do stávajícího portálu instituce (ilustrativní ukázka)



3 Databáze indexovaných dokumentů

Indexovací a vyhledávací stroj M-CAST je nyní koncipován jako nástroj pro extrahování odpovědí z dokumentů a korpusů umístěných na pevném disku; v budoucnu se počítá i s extrahováním odpovědí z internetu – z webových stránek nebo prostřednictvím klasických webových vyhledávačů (Google, MSN, AOL atd.).

Jako u jiných systémů na bázi sémantických technologií, tak i zde hraje důležitou roli velikost a reprezentativnost databáze indexovaných dokumentů, jejichž indexovaný obsah je uložen v interní databázi, ve které následně vyhledávají uživatelé. Tato velikost a reprezentativnost je do jisté míry měřítkem množství a kvality informací, které lze ve vyhledávacím systému najít. Systém M-CAST neindexuje všechny tištěné či elektronické dokumenty dostupné v daném oboru/daných oborech, protože informační zdroje zařazené do databáze M-CAST podléhají výběru podle předem stanovených kritérií.

3.1 Databáze indexovaných dokumentů zpřístupněné pro testování systému M-CAST

Pro testovací fázi systému M-CAST zpřístupnila Národní knihovna databáze, které jsou součástí lokálního repozitáře, tedy instance M-CASTu používané Národní knihovnou. Jde o aplikace pracující s digitálními objekty:

- Kramerius (kramerius.nkp.cz), která obsahuje sbírku digitalizovaných periodik a monografií; vyhledává se v plných textech, výsledkem hledání může být každá strana dokumentu, slouží k testování českého modulu;
- Memoria (www.manuscriptorium.com) obsahující sbírku historických dokumentů; vyhledává se v popisných údajích, výsledkem hledání jsou metadatové záznamy, slouží k testování českého modulu;
- Databáze v Alephu – ALEPH (sigma.nkp.cz), tj. elektronický katalog Národní knihovny; vyhledává se v popisných údajích, výsledkem hledání jsou metadatové záznamy, slouží k testování českého modulu.

Externím zdrojem pro testování byla Evropská ústava (europa.eu/constitution/), která obsahuje texty ve všech jazycích M-CASTu, slouží tedy k testování vícejazyčného modulu.

Obr. č. 5: Ukázka instance systému M-CAST v Národní knihovně

The screenshot displays the M-CAST search interface. At the top, the logo of the National Library of the Czech Republic (NK) is visible. The search bar contains the query "Kdy nastane extáze?". Below the search bar, there are buttons for "Hledat", "Nastavení", "Pokročilá vyhledávání", "Rozdělit", and "Návod". A checkbox labeled "vyhledej ve vzdálených repozitářích" is checked. The search results are displayed in a list format, starting with "1 Psyche". Under "Psyche", there is a sub-result "1 „Extáze nastane, když božské působení má takovou sílu, že každé spojení smyslu se zevním světem je úplně anebo skoro úplně přerušeno." followed by a list of related items with their respective percentages: "Zdroj - 100%", "2 listopadu 1918 do 29. května 1926 nastane na světě jakýsi oddech nebo jakýsi klid, ale ne jenom na zemi, nýbrž v celém vesmíru. Zdroj - 14%", "3 Polom nastane osobní vřída Kristova. Zdroj - 11%", "4 se zase do př. střešního (pokojikal a setrval zde až do mystické smrti Terezie, jež nastala asi o 1. hod. Zdroj - 11%", "5 Zde pak nastává podobnost s auto matičkým psaním a spirítistickou extází. Zdroj - 2%".

Instance M-CASTu používaná v Národní knihovně může být propojena i s dalšími instancemi systému, v současné době s instancí používanou v Polské internetové knihovně. Uživatel Národní knihovny může tedy vyhledávat i ve vzdáleném repozitáři, ve kterém jsou uloženy dokumenty zpřístupňované Polskou internetovou knihovnou.

3.2 Indexace v systému M-CAST

Proces indexace probíhá ve všech lingvistických modulech podle stejných pravidel, extrahovaná data jsou stejné kategorie, zpracování těchto dat je tedy nezávislé na původním jazyce.

3.2.1 Indexace dokumentů

Základním předpokladem pro úspěšné indexování dokumentů je tokenizace, kdy je text rozložen na základní selekční jednotky a v textu jsou identifikována slova, mezeře, interpunkce a začátky a konce vět; dále stemming (lematizace), kdy se odstraňuje zakončení slova a ponechává se kmen/kořen (slovní základ), resp. (při lematizaci) je určena pro každý slovní tvar jeho základní podoba; neméně důležitá je derivace, kdy se k základnímu tvaru slova generují jeho inflexní tvary. V textech se provádí morfologická desambiguace slovních tvarů, nevýznamová a nespecifická slova jsou pomocí negativního slovníku (slovníku stop slov) odstraněna; při indexaci textů se uplatňuje také stejná typologie dotazů a odpovědí, používá se tentýž analyzátor dotazů, aplikuje se ontologie TRUST apod. Texty jsou konvertovány do Unicode a rozděleny do textových kilobytových bloků, dochází tak k redukci velikosti indexů. Každý textový blok je podroben morfologické, syntaktické a sémantické analýze.

3.2.2 Indexace typů dotazů

Indexace typů dotazů používaná v systému M-CAST je bezpochyby nejoriginálnější prvkem systému Qristal, který vyvinul francouzský partner, firma Synapse Développement z Toulouse (dále Synapse Développement). Při analýze bloků, které se indexují, se označují eventuální odpovědi, například název funkce určité osoby („pekař“, „ministr“, „ředitel kanceláře“ ...), datum narození („narozen 28. dubna 1958“), kauzální vazby („kvůli množství sněhu“, „vzhledem k náledí“ ...), nebo příčinné vazby („což způsobilo velké zmatky“, „což umožnilo zvládnutí – řízení provozu“ ...); blok je tudíž indexován tak, aby mohl dodat odpověď označeného typu.

3.2.3 Indexy systému M-CAST

Na základě získaných výsledků je budováno 8 různých indexů:

- Index základů slov, v případě homonymních a polysémních slov jejich významů;
- Index vlastních jmen;
- Index idiomů;
- Index pojmenovaných entit;
- Index konceptů, tj. uzlů ontologie TRUST;
- Index jednotlivých pojmů ontologie TRUST;
- Index typů otázek a odpovědí;
- Index klíčových slov z textu.

3.3 Extrakce odpovědí

Uživatelův dotaz je podroben syntaktické a sémantické analýze. Je určen typ dotazu. Výsledek sémantické analýzy dotazu může být negativně ovlivněn tím, že prvky dotazu vytvářejí omezený kontext, neboť dotaz je nutně na rozdíl od dokumentů výrazně kratší. Na základě sémantické analýzy dotazu jsou podle váhy stanoveny významově důležitá klíčová slova, pivots. Při vyhledávání se používají tyto výrazy obohacené o synonyma,

o odpovídající koncepty, a přiřazené k typu otázky. Po analýze dotazu jsou prohledávány všechny indexy a jsou vybrány textové bloky, které nejvíce odpovídají parametrům dotazu; z nich jsou vybrány jednotlivé relevantní odpovědi, u nichž je stanovena váha na základě statisticko-lingvistických metod a jejich pořadí.

4 Vyhledávání v systému M-CAST

Systém podporuje dva typy vyhledávání: jednoduché a pokročilé. Pro každý typ vyhledávání je určen speciální formulář.

4.1 Jednoduché vyhledávání

Při jednoduchém vyhledávání se zapíše do vyhledávacího boxu dotaz, kterým může být otázka v přirozeném jazyce, např. „Kdy byly předány insignie Karlově univerzitě?“, nebo skupina selekčních termínů, např. „život“; „láska“, případně fráze, např. „hodně muziky za málo peněz“. Systém M-CAST odpoví vytvořením stránky s výsledky, tj. nabídne přesnou odpověď na dotaz a seznam úryvků odpovědí. Nejvíce relevantní odpovědi jsou umístěny jako první.

4.2 Pokročilé vyhledávání

Kromě jednoduchého vyhledávání umožňuje M-CAST provádět pokročilé vyhledávání, při kterém je možné zúžit „oblast vyhledávání“ podle tří úrovní kategorií klasifikačního systému MDT.

Obr. č. 6: Zúžení dotazu klasifikačním systémem MDT

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápoje

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novinyře

M-CAST život Jednoduché vyhledávání
Hledat Nastavení
Rezdělit Nápoje

Kategorie MDT:

6 - Aplikované vědy. Lékařství. Technika
61 - Lékařské vědy

NKCP

Žádná odpověď pro život

1 - 2 z 2 nalezených odpovědí (0 min 07 sec)

1 **hnutí pro život ČR - potrat všechno změní**

1 hnutí pro život ČR

Ždoci - 100%

2 **Senior - Relaxace (Z. : 2003 : Lysá nad Labem, Česko), Výstaviště Lysá nad Labem - senior - relaxace výstava nejen pro dříve narozené : výstava o zdravém životním stylu : 10.-13. dubna 2003 výstaviště lysá nad labem**

1 senior - relaxace výstava nejen pro dříve narozené : výstava o zdravém životním stylu : 10.-13. dubna 2003 výstaviště lysá nad labem

Ždoci - 100%

1 - 2 z 2 nalezených odpovědí (0 min 07 sec)

Ptejte se knihovny | Hledání na WWW | Copyright

4.3 Metoda dotazování v přirozeném jazyce aplikovaná v systému M-CAST

Používá se v modulu jednoduchého vyhledávání.

Z pozice uživatele zpracovává systém M-CAST dotazy položené v přirozeném jazyce a nabízí přesné a jednoznačné odpovědi podpořené úryvky odpovědí, které jsou extrahovány z obsáhlé databáze indexovaných dokumentů. Jednou z nejdůležitějších podmínek úspěšnosti vyhledávání pomocí této metody je soubor dobře formulovaných otázek. Všechny informace, které jsou předmětem dotazu, musejí být obsaženy v databázi indexovaných dokumentů.

4.3.1 Délka dotazů

Dotazy používané při aplikaci metody dotazování v přirozeném jazyce jsou obvykle krátké, skládají se ze tří, čtyř slov. Dotaz formulovaný jako celá věta vede však automaticky k jeho prodlužování. Dlouhá otázka obsahující více klíčových významových prvků, „pivotů“, může mít za následek, že při vyhledávání dokumentů jsou relevantní dokumenty odfiltrovány a nabídnuty dokumenty méně relevantní, obsahující více klíčových slov, avšak nerelevantních pro daný dotaz. Např. dotaz „Který umělecký soubor vystoupí na zahájení výstavy Velká Morava v Berlíně?“ je příliš dlouhý, obsahuje 6 významových prvků; přímá odpověď, ani úryvek obsahující přímou odpověď na tento dotaz nebyly získány.

4.3.2 Odpovědi v systému M-CAST

Systém M-CAST generuje na uživatelův dotaz celý soubor, blok odpovědí, který se skládá z:

- přímé odpovědi
- úryvku obsahujícího/podporujícího přímou odpověď
- možnosti vizualizace zdrojového dokumentu

4.3.2.1 Přímá odpověď

Přímé odpovědi jsou většinou jmenné entity (jméno, místo, chronologický údaj, jmenné a slovesné fráze). Např. přímá odpověď na otázku „Co držel Zeus v pravé ruce?“ je „bohyni vítězství“.

4.3.2.2 Úryvek odpovědi

Systém M-CAST generuje spolu s přímou odpovědí i odpovídající úryvek zdrojového dokumentu, který zasazuje přímou odpověď do potřebného minimálního kontextu. Úryvek obsahující přímou odpověď na výše uvedenou otázku je tedy: „V pravé ruce držel Zeus bohyni vítězství, jak se právě k němu sklání, chtíc hlavu jeho věncem ozdobiti; v levé držel žezlo.“

4.3.2.3 Vizualizace zdrojového dokumentu

Poslední část bloku odpovědi představuje potenciální vizualizace zdrojového dokumentu, tedy hypertextový odkaz vedoucí k příslušné stránce zdrojového dokumentu.

Obr. č. 7: Soubor odpovědí na dotaz Kdo je Niké

The screenshot shows the M-CAST search interface with the query "Kdo je Niké?". The search results are displayed in a list format. The first result is "Cimhanzi Týma - Mythologie čili báješlvi Řekův a Římanův". The second result is "Plzeňské noviny, Český lev". To the right, there is a snippet from a historical document titled "TYCHĚ lat. FORTUNA." which includes illustrations of the goddess Tyche and Fortuna.

M-CAST
 vyhledej ve vzdálených repozitářích

NKP
 Odpověď/Odpovědi:
 bohaté vzhledem
 pro Kdo je Niké?

1- 10 z 22 nalezených odpovědí (0 min 02 sec)

1 Cimhanzi Týma - Mythologie čili báješlvi Řekův a Římanův

1 Niké jest bohyně vzhledu a zdoru včetejch jak ve výce tak v umění výtvarném, v kústém i božském počínání.
 Zdroj: 100%

2 Za to, že Niké a její ostarší bratři a sestry, Zéos (řevivost), Krátos (Ška) a Dio (pókoi) byli první, kteří se na výzvěti boha Zéos do toje dostavě, aby jmu ve válce s Týdru nájemcové byli, vykážal jm Zéus Olym za stáje jech ošča.
 Zdroj: 100%

3 Niké vysvězováča ce jako parna, v starších obcích bez křákel, pozděj okřákená, brzy kyně vželi, brzy se nad rín vžáši a věneč mu na hlavu klade.
 Zdroj: 100%

4 I NIKÉ ut.
 Zdroj: 100%

2 Plzeňské noviny, Český lev

1 loc. NIKÉ Arnej, jakož Tůbec ty, kád máj prá. vo myšlvost samostatně provozovat, aneb se v

TYCHĚ lat. FORTUNA.
 Tyché, deusa Zeva a Thetidy, jest bohyně vzhledu a štěstí, a jakživa domovně se divěti, že sáhala má velkj vliv na pšobení jeho a štěstí prave. Římanci si také "sňhoda" říkali, a říci prvotně bohyně Tyché lat. Fortuna. O bohyni této se myslilo, že ona má "sňhodu" tyto, a proto jí upobívali, jak děti v rukou vesle života anebo stoji na ková, kteráž jest znamením včetejch ošča; jako Makhónaj dárlyna štěstí nosí Amakhón ruk, ruk bojmění.
 Ořava byla jak v Řecku tak a Římanci včetejch předětim ale vřlyty jí znanenci.
 Tyche. Fortuna.

NIKÉ lat. VICTORIA.
 Niké jest bohyně vítězství a zdoru včetejch jak ve válce tak aneb vřvarně, v kústém i božském počínání. Ona brala proto máj ve všech válkách a bitváchých oděch, při velkých řák a slavnostech řeckých i římských, ano i na mnohých výkonech

Obr. č. 8: Odpověď na dotaz Co tropili staří Čechové

The screenshot shows the M-CAST search interface with the query "Co tropili staří Čechové?". The search results are displayed in a list format. The first result is "Zibrť Čeněk - Staročeské výroční obyčeje, pověry, slavnosti a zábavy prstonárodní pěkod o nich vypravují písemné památky až po nás věk". The second result is "Že tropili staří Čechové o novém roce žerty a kratochvíle, po-řvzuje Resacius (r)".

M-CAST
 vyhledej ve vzdálených repozitářích

NKP
 Odpověď/Odpovědi:
 novém roce žerty
 pro Co tropili staří Čechové?

1- 10 z 45 nalezených odpovědí (0 min 04 sec)

1 Zibrť Čeněk - Staročeské výroční obyčeje, pověry, slavnosti a zábavy prstonárodní pěkod o nich vypravují písemné památky až po nás věk
 Příspěvek ke kulturním dějinám českým.

1 Že tropili staří Čechové o novém roce žerty a kratochvíle, po-řvzuje Resacius (r).
 Zdroj: 100%

4.4 Vícejazyčné vyhledávání v systému M-CAST

Připojit nový jazyk, tedy vytvořit nový jazykový modul, je v systému M-CAST poměrně jednoduché; architektura systému zodpovídání dotazů je navržena tak, aby byla nezávislá na konkrétním jazyce. Ve všech jazykových modulech se používají stejné softwarové nástroje, např. SintaGest. Musí být upraveny a importovány jen jazykové zdroje nového jazyka (lexikon, thesaurus, ontologie, formální definice otázek), aby souhlasily s používanými nástroji pro zpracování přirozeného jazyka. Pro vícejazyčné vyhledávání v systému M-CAST je jako propojovací jazyk použita angličtina.

Již bylo uvedeno, že obě testovací pracoviště neměla k dispozici rozsáhlou sbírku identických dokumentů ve všech jazycích M-CASTu. Bylo tedy zvoleno náhradní řešení a tato funkčnost systému M-CAST byla testována na textu Evropské ústavy. Pro tyto účely byl ve všech jazycích M-CASTu definováno 49 otázek, jako např.:

Questions in English	Otázky v češtině
What is „Europe Direct“ ? For how many years are elected the members of the European Parliament ? What is the percentage for a qualified majority ? What is the capital of the European Investment Bank ? What is „distrainment-description“ ? What is a „heavy goods vehicle“ ? When did Portugal join the European Community ?	Co je „Europe Direct“? Na kolik let jsou voleni členové Evropského parlamentu? Jaké procento tvoří kvalifikovanou většinu? Jaký kapitál se přiděluje Evropské investiční bance? Co se rozumí „dokumentačně-zajišťovacím řízením“? Co se rozumí „těžkým nákladním vozidlem“? Kdy přistoupila Portugalská republika k Evrop. společenstvím?

Obr. č. 9: Otázka v angličtině, odpověď v polštině

NK Národní knihovna České republiky
National Library of the Czech Republic

M-CAST [what is the anthem of European Union?]
 vyhledej ve vzdálených repozitářích
 Hledat Rozdělit
 Dobrořeč
 Vymazání
 Nastavení
 Nápověda

NKP
 Odpověď / Odpovědi:
 Ody do radości
 pro What is the anthem of European Union?
 1- 10 z 68 nalezených odpovědí (0 min 05 sec) 1 2 3 4 5 6 7 Další

1 TRAKTÁT USTANAWIAJĄCY KONSTYTUCJĘ DLA EUROPY
 1 Hymn Unii pochodzi z „Ody do radości” z TK Symfoni Ludwiga van Beethovena.
 Zdroj: 100%

Obr. č. 10: Otázka ve francouzštině, odpověď v češtině

NK Národní knihovna České republiky
National Library of the Czech Republic

M-CAST [Quel est le capital de la Banque centrale européenne?]
 vyhledej ve vzdálených repozitářích
 Hledat Rozdělit
 Dobrořeč
 Vymazání
 Nastavení
 Nápověda

NKP
 Odpověď / Odpovědi:
 162 652 727 400 EUR
 pro Quel est le capital de la Banque centrale européenne ?
 1- 10 z 94 nalezených odpovědí (0 min 02 sec) 1 2 3 4 5 6 Další

1 SMLOUVA O Ústavě pro Evropu
 1 Bance se přiděluje základní kapitál ve výši 163 653 727 000 EUR, který spolčí desítky států v těchto částkách.
 Zdroj: 100%

5 Testování a hodnocení systému M-CAST

Cílem procesu testování a hodnocení systému bylo ověřit jeho funkčnost v plném provozu, prověřit odezvu systému a praktickými podněty a zkušenostmi přispět k odstranění všech případných nedostatků. Z hlediska uživatele bylo nutné se zaměřit také na proces vyhledávání jako takový a ověřit, je-li vyhledávání v systému M-CAST komplikované, nebo naopak snadné a intuitivní i pro méně zkušeného či nezkušeného uživatele.

5.1 Odezva systému

Jedním z důležitých kritérií hodnocení kvality systému je dosažená odezva. Průměrná odezva prototypu systému M-CAST byla v době testování 4,9 sekundy.

5.2 Správnost odpovědí

Národní knihovna v letech 2004-2006 nevladila rozsáhlé soubory vícejazyčných dat soudobých dokumentů, proto byla fáze testování v Národní knihovně zaměřena na ověření funkčnosti systému M-CAST při vyhledávání informací ve sbírce indexovaných historických textů, a to v českém jazyce. Při testování bylo nutné vyřešit problémy spojené s dotazováním v přirozeném jazyce aplikovaném při dotazování ve sbírce historických dokumentů, protože vzorové otázky pro M-CAP Národní knihovny byly původně definovány pro vyhledávání ve sbírkách současných textů. Bylo tedy nutné překonat problémy s dobově podmíněným pravopisem, s historickým slovníkem a složitostí historické syntaxe. Výsledky testování byly výrazně ovlivněny kvalitou OCR textů. Původním cílem databáze Kramerius byla záchrana a zpřístupnění bohemikálních dokumentů tištěných na kyselém papíru, jejichž existence je ohrožena rozpadem (křehnutím) papírového nosiče. Špatný fyzický stav především starších novin a časopisů ovlivnil kvalitu výsledných obrazových souborů a výstup konverze OCR. Horší kvalita obrazových souborů se pak mnohdy negativně promítla do chybovosti při rozpoznávání během procesu konverze OCR do textové podoby. OCR texty obsažené v databázi Kramerius byly indexovány pro potřeby M-CASTu bez předchozích úprav: nebyly odstraněny nepesnosti při segmentaci textu a chybějící nerozpoznané znaky, proto mnohdy některá slova, případně fráze nebyly identifikovány správně a docházelo tak k významovým posunům, případně výsledek dotazu nedával smysl. Proces hodnocení byl ztížen tím, že systém M-CAST nabízí soubor/blok odpovědí, přičemž je všeobecně známo, že hodnocení komplexu odpovědí generovaných systémem je obtížnější než hodnocení jednotlivých typů odpovědí. Proto byla pro potřeby hodnocení systému vypracována tato speciální kategorizace odpovědí:

5.2.1 Přímá odpověď

- správná, tj. přesná krátká odpověď
- nesprávná, tj. chybná krátká odpověď
- nepřesná, tj. krátká odpověď obsahující méně či více informací, než vyžadoval dotaz
- žádná odpověď

5.2.2 Úryvek odpovědi

- správná odpověď obsažená v 1.–5. úryvku
- správná odpověď obsažená v úryvcích na dalších pozicích
- nesprávná + žádná odpověď, tj. úryvek obsahoval nesprávnou odpověď, případně systém negeneroval žádný úryvek jako odpověď na položený dotaz

5.2.3 Výsledná statistika

Přímá odpověď

správná	10 %	206
nesprávná	6 %	124
nepřesná	31 %	638
žádná	53 %	1092
celkem	100 %	2060

Úryvek odpovědi

správná odpověď, 1.-5. úryvek	73 %	1520
správná odpověď, ostatní	15 %	315
nesprávná + žádná odpověď	12 %	225
celkem	100 %	2060

6 Možnosti aplikace systému M-CAST

6.1 M-CAST v digitálních knihovnách

Jak jsme již uvedli, základní metodou vyhledávání je dotaz v přirozeném jazyce, dotaz celou větou. Tento typ vyhledávání je dominantní v rozsáhlých sbírkách digitálních či digitalizovaných dokumentů, tedy v digitálních knihovnách. Digitální knihovna bývá definována jako integrovaný systém zahrnující soubor elektronických informačních zdrojů a služeb umožňující získávání, zpracovávání, vyhledávání a využívání informací v tomto systému uložených. Základní funkcí digitální knihovny je poskytnout uživatelům možnost jednoduchého přístupu k digitálním anebo digitalizovaným dokumentům, případně i k sekundárním informacím o tištěných primárních zdrojích.

Předpokládáme, že zdokonalená verze vyhledávacího systému M-CAST se uplatní i při procesu vyhledávání ve vznikající Národní digitální knihovně, jejímž cílem bude zpřístupnění digitálních a digitalizovaných knihovních dokumentů a která bude v budoucnu fungovat v širším kontextu České digitální knihovny.⁴ Následuje ukázka potenciální aplikace systému M-CAST v databázi Manuscriptorium.⁵

Obr. č. 11: Odpověď na dotaz Kdo pínáší dvodám dřevo z lesa

The screenshot shows the M-CAST search interface. At the top, there is a search bar with the text "Kronika tak řečeného Dalimila" and a search button labeled "Hledat". Below the search bar, there are tabs for "Document Summary", "Bibliographic Description", "Browsing", "Bookmarks", and "Help". The "Bibliographic Description" tab is active, showing a detailed entry for a manuscript. The entry includes a thumbnail image of a manuscript page, a title "Kronika tak řečeného Dalimila", and various fields such as "Form", "Support", "Extent", "Collection", "Layout", and "Description". The "Description" field contains a detailed text description of the manuscript's content and history.

6.2 M-CAST v hybridních knihovnách

Hybridní knihovna integruje klasickou knihovnu představanou především tištěnými dokumenty a digitální knihovnu, obvykle s cílem zkvalitnění služeb uživatelům. Typickou ukázkou hybridní knihovny je v současné době Národní knihovna, která zpřístupňuje tradiční i elektronické dokumenty a byla jedním z testovacích pracovišť systému

⁴ viz <http://www.ndk.cz/narodni-dk>

⁵ obrázek zveřejněn se souhlasem správce databáze Manuscriptorium PhDr. Z. Uhlíře

M-CAST. V průběhu vývoje a testování systému jsme se snažili rozšířit možnosti vyhledávání systému o prvky a postupy, které jsou typické pro hybridní knihovny. Výsledkem je výše zmíněná integrace klasifikačního systému MDT do procesu vyhledávání a možnost klást dotaz pomocí souboru selekčních termínů. Za účelem dalšího testování systému M-CAST aplikovaného v hybridních knihovnách využívá Národní knihovna výsledky jiného projektu, tj. Obohacení bibliografických záznamů o obsahy popisovaných dokumentů (projekt TOC – Table Of Content), který byl zahájen po schválení novely autorského zákona. Od poloviny r. 2006 existuje možnost obohatit bibliografické záznamy u odborné a populárně-naučné produkce o obsahy popisovaných dokumentů.

6.2.1 Systém M-CAST a obsahové údaje (projekt TOC – Table Of Content)

Dotaz je formulován formou fráze (souboru selekčních termínů) z obsahových údajů „králové a věštki“. Výsledkem dotazu je úryvek odpovědi obsahující danou frázi:

Obr. č. 12: Úryvek odpovědi na dotaz frází Králové a věštki

The screenshot shows the M-CAST search interface. At the top left is the logo of the National Library of the Czech Republic (NKP). The search bar contains the text "králové a věštki". Below the search bar, there is a checkbox for "vyhledej ve vzdálených repozitářích" which is checked. To the right of the search bar are buttons for "Hledat", "Rozdělit", "Pokročilý vyhledávání", "Nastavení", and "Nápověda". Below the search bar, there is a message "Žádná odpověď pro králové a věštki" and a list of search results. The first result is "1 Krámer, Claus, 1957. - Keltské umění léčit". The snippet of the book review is as follows:

1 V Merinových stopách – Římané, čarodějnice a šamani – Návrat dubových kněžů – Stopy druidů na Ierškém ledu – Tajné vřobení pro věštky – Tehdy stejné aktuální jako dnes – hledání smyslu života – Odkaz druidů – Druidové šamani z Východu – Směsí puntíků – Králové a věštki – Umění přežít – Moudřejší ustoupí – Zachránci kultury – Indiáni Evropy – Začátek všeho šamanismus – Šamanismus v praxi – O elfech a silových zvířatech – Kontakt s duchy ochránců – Zpráva o autorově zážitku – cesta k vlastním silovým zvířetím – Celostní léčitelství Keřů – Druidové jako všestranní génové – Náboženství balzám pro duši – Umirání a smrt neznamenají konec – Maří moudrosti – Druidství dnes – Praktici a specialisté – Zdravý a nemocný tak to vidí Kelťové – Na začátku je diagnóza – Mentální léčitelství Keřů – Zdraví začíná v hlavě – Psychická sebeochrana – Ochranný kruh – Stát oběma nohama na zem uzením – Spojení se vším dech – Meditace cesta k druidovi v nás – Magické léčitelství Keřů – Slova, která léčí – Léčivá síla magických znamení – Ochranná síla amuletů a talismanů – Rostlinné léčitelství Keřů – Poklad keltajských léčivých bylin – Jmelí obilná rostlina Keřů – Léčivá síla stromů – Rostliny bohů – Co kdyby druidové znali se zas vrazi – S Měsícem k vnější harmonii – Léčivá síla tónů – Léčivá síla světla a barvy – Jíst jako Kelťové – Léčivá síla kamenů – Léčivá síla vůní a vykuřovadel – Léčící spřátelená duše Anam Cara Duchovní léčba – Vědomé snění – Magické kamenné kruhy a léčivá energie Země – Od léčitelky k čarodějnici – Keltství ke koupi Kelťové přírodní léčitelství šamanismus Celts healing shamanism populárně-naučné publikace popular works

At the bottom of the snippet, there is a source link "Zdroj" and a progress indicator "100%".

Při vizualizaci zdrojového dokumentu získáme bibliografický záznam a prokliknutím na specifický formát „Obsah“ získáme údaje obsahu dokumentu, ve kterých je uvedena fráze.

Obr. č. 13: Odpověď na dotaz frází Králové a věšci

Záznam 1 z 1

Hlavní záhlaví: [Krámer, Claus, 1957-](#)

Název: [Keltské umění léčit / Claus Krämer](#)

Vydání: Vyd. 1.

Nakl. údaje: Frýdek-Místek : Alpress, 2006

Popis (rozsah): 212 s. : il. ; 21 cm

EXEMPLÁŘE

ISBN: 80-7362-215-7 (váz.)

Edice: Klokan

Název originálu: [Heilkunst der Kelten](#)

Poznámka: Přeloženo z němčiny

Poznámka: Obsahuje bibliografické odkazy

Předmět. heslo

- [Keltové](#)
- [přírodní léčitelství](#)
- [šamanismus](#)

Forma, žánr: * populárně-naučné publikace

Skupina Konspektu: [615.8 - Fyzioterapie, Psychoterapie, Altern](#)

MDT: [615.83+615.89](#)
[133.4-2-382](#)
[\(-15\)](#)
[\(0.062\)](#)

Předm. hesla-angl.: * Celts * healing * shamanism

Forma, žánr-angl.: * popular works

Zobrazit: [Obsah](#)

Obsah

V Merlinových stopách	12
Římané, čarodějnice a šamani	13
Návrat dubových kralů	15
Stopy druidů na tenkém ledu	16
Tajné vědění pro všechny	16
Tedy stejně aktuální jako dnes – hledání smyslu života	18
Odkaz druidů	19
Druidové – šamani z Východu	
Směť peneků	26
Králové a věšci	28
Umění přežít	29
Moudřejší ustoupí	33
Zachránit kultury	35
Indiáni Evropy	36
Začátek všeho – šamanismus	38
Šamanismus v praxi	42
O otělech a sílových zvířatech	49
Kontakt s duchy ochránců	50
Závěra o autorově zážitku –	

6.2.2 M-CAST v hybridních veřejných knihovnách

Většina veřejných knihoven v současné době buduje a zpřístupňuje databáze regionálních osobností, událostí apod., které jsou ve většině případů budovány jako databáze plnotextové. Aplikace systému M-CAST v těchto databázích by umožnila klást uživatelům těchto knihoven dotazy v přirozeném jazyce.

Příklad potenciálního využití vyhledávacího systému M-CAST

dotazy: *Kde studoval Bohumír Jaroněk?*

Kde působila rodina Jaroněků?

Obr. č. 14: Odpověď na dotaz Kde studoval Bohumír Jaroněk (ilustrativní ukázka)

http://www.mvk.cz - MASARYKOVA VEŘEJNÁ KNIHOVNA VSETÍN - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Deutsch English

Čtvrtek 5. července 2007 Svátek má Cyril a Metoděj

REGIONÁLNÍ BIBLIOGRAFIE

Regionální bibliografie

BOHUMÍR JARONĚK

malíř, grafik

23.04.1866 Zlín

18.01.1933 Zlín

Bohumír Jaroněk se narodil ve Zlíně v rodině barvíře. Studoval na Státní odborné uměleckopřmyslové škole pro zpracování dřeva ve Valašském Meziříčí. Dobrý řemeslný základ rozvíjel pak během svých neobyčejných cest. Pracovat začal v ateliéru fotografa Šteřelského v Budapešti, neboť na umělecká studia v Paříži nebo v Mnichově, kam se stahovala většina adeptoů umělecké dráhy, neměl prostředky. Pobyt v Budapešti mu umožnil několikrát působení v předním káhirském salonu, po této životní etapě začal cestovat.

V roce 1896 se vrátil na Moravu a ve Valašském Meziříčí začal pracovat v ateliéru svého staršího bratra Aloise, zdeoblič porcelán, a své sestry Julie, která tkala gobelíny. Od roku 1909 působila rodinná dílna Jaroněků v Rožnově pod Radhoštěm. Tam se sourozenci Jaroněkovi ještě více přiblížili ke kořenům valašské lidové kultury a využili snahu zachovat a zachraňovat památky lidové architektury. To vedlo k založení Valašského muzea v přírodě v roce 1925.

Bohumír Jaroněk se zabýval především grafikou, ale zhotovoval i temperové návrhy nebo skicpy pro řešení gobelínů.

V dřevorytech zachoval onu jednodušost, s jakou příroda ve spojení s lidovou architekturou tvoří jeden pohledový i citový celek. Celé Jaroněkovy dílo charakterizuje stylizované zjednodušené kresba, doplněná jen několika málo barvami, odvozenými z přírody.

Bohumír Jaroněk zemřel 18. ledna 1933 a je pochován na Valašském Slavném v Rožnově pod Radhoštěm.

Výběrová bibliografie z fondu MVK Vsetín

REGIONÁLNÍ BIBLIOGRAFIE

- [Barusch, Josef](#)
- [Božnerová, Marie](#)
- [Bubela, Karel](#)
- [Čalibová, Zlata](#)
- [Čadáková, Božena](#)
- [Dočkalová - Kobzáhová, Jana](#)
- [Frýč, Tomáš](#)
- [Hainý, Zdeněk](#)
- [Hálek, Josef](#)
- [Hlavica, Emil](#)
- [Hlavica, František](#)
- [Hlavica, Rudolf](#)
- [Hofman, Karel](#)
- [Chromek, Antonín](#)
- [Jáček, Ivan](#)
- [Jaroněk, Bohumír](#)
- [Jaroněk, Alois](#)
- [Kříška, Jaroslav](#)
- [Kusný, Vojtěch](#)
- [Kalendář, Zdeněk](#)
- [Kaleta, Josef](#)
- [Karašiat, Jan](#)
- [Kobelha, Karel](#)
- [Kobzák, Jan](#)
- [Kramolík, Cenek](#)
- [Křenek, Jiří](#)
- [Kříška, Petr](#)
- [Křížan, Jiří](#)
- [Machala, Miroslav](#)
- [Majer, Luděk](#)
- [Maláková, Zuzana](#)

6.3 M-CAST v paměťových institucích

Také další paměťové instituce jako archivy, galerie a muzea se podílejí na zpřístupnění národního kulturního dědictví, a to zpřístupněním informací na úrovni celých kolekcí či zpřístupněním informací o jednotlivých exponátech/artefaktech.

Pro záznamy popisovaných objektů, zvláště archivních a muzeálních, je typická kombinace strukturovaných selekčních prvků/přístupových bodů a rozsáhlého volného textu, který obsahuje další informace nezbytné pro badatelskou činnost.

Zpřístupnění má být efektivní, rychlé, komfortní a uživatelsky vstřícné. Efektivitu, rychlost a komfortní přístup zajišťují na jedné straně standardizované selekční prvky/přístupové body, tj. atributy, podle kterých můžeme informaci vyhledat. V níže uvedené ukázce záznamu o Janu Dočkálkovi zveřejněném na portálu Archives Canada jsou standardizované selekční prvky označeny jako „subjects“.⁶

V popisných částech záznamu (ADMINISTRATIVE HISTORY/BIOGRAPHICAL SKETCH a COPE AND CONTENT), které obsahují informace v podobě volného textu, se za komfortní a uživatelsky vstřícný přístup považuje možnost klást otázky v přirozeném jazyce, tedy ptát se celou větou.

Příklad potenciálního využití vyhledávacího systému M-CAST

dotazy: *Who was the leading officer in the Czechoslovak Red Cross in England?*

When did the communists come to power in Czechoslovakia?

When did Dockalek join the British Army?

Obr. č. 15: Odpověď na dotaz Who was the leading officer in the Czechoslovak Red Cross in England (ilustrativní ukázka)

ARCHIVESCANADA.ca
Canadian Archives / Archives canadiennes
Information Network / Réseau d'information archivistique

your gateway to Canada's past

search archives

CAIN No. 200924
TITLE: Jan Dockalek fonds [textual record, graphic material]
 PHYSICAL DESCRIPTION: 7.5 cm of textual records.
 5 photographs: b&w and col.
 DATES: 1941-1971.

ADMINISTRATIVE HISTORY/BIOGRAPHICAL SKETCH: Jan Dockalek, editor, poet, dramatist, born in Kromeraz, Czechoslovakia, was a newspaper reporter in his homeland. In 1940 he fled to Turkey where he joined the Czechoslovak Contingent of the British Army in the Middle East. He served with the Contingent until after Rommel's defeat in Africa, when the unit was taken to England. Here he became a leading officer in the Czechoslovak Red Cross. He remained in England until 1946 when he was named delegate of the Czechoslovak Red Cross with the Canadian Red Cross, with headquarters in Toronto. When the communists came to power in Czechoslovakia in 1948, he requested and was granted asylum in Canada. He was active in the Czech community in Canada and passed away on January 1, 1980.

SCOPE AND CONTENT: Fonds consists of a book of poetry by Dockalek, entitled *Slezske Stry*, 1941, and English translation of his play "The Miners", n.d., 2 issues of *Zpravodaj* (Daily News of the Czechoslovak Contingent in the Middle East), 1941, two issues of *Bobrik*, 1946, 1947, as well as clippings of a history of Canada by Dockalek, which appeared in *Nase Hlasy* during the early 1970s, and the program of the 30th Anniversary Celebration of the liberation of Czechoslovakia, held in Dresden, Ontario, in 1948.

Fonds also consists of photographs, 1946-1965, depicting Eva and Otto Pavlenek in Czech national dress; Hana and Eva Dockalek and Professor Venceslav Hlavaty of the Czechoslovak Society for Arts and Science; Czech language school and Sokol Gymnastic Organization of Masaryk Memorial Institute, Toronto, Ont.; Moravian Slovensk Dance Group, 1964 or 1965.

REPOSITORY: National Archives of Canada
 RESTRICTIONS ON ACCESS: Textual Records Open
 Graphic (Photo) Open TERMS GOVERNING USE AND REPRODUCTION:
 Photographs: No donor restrictions. Various copyrights on items less than 50 years old. Copyright expired on items more than 50 years old. Credit: Name of photographer / National Archives of Canada / Copy negative no.

FINDING AID AVAILABLE: No finding aid
 ADDITIONAL INFORMATION: Received in 1976 from Jan Dockalek of Toronto, Ont.

Také záznamy českých archivů vynikají svým rozsahem. I zde by se mohla uplatnit zdokonalená verze systému M-CAST, zvláště v popisných částech, jako jsou dějiny, funkce, zaměření, činnost, „obecný kontext“ aj.

Příklad potenciálního využití vyhledávacího systému M-CAST ⁷ dotazy:

Kdy bylo zřízeno ministerstvo železnic v Předlitavsku?

Co bylo úkolem ministerstva železnic?

Kdy zaniklo ministerstvo železnic v Předlitavsku?

Jaká byla struktura ministerstva?

Které sekce tvořily ministerstvo v roce 1912?

Z kterých sekcí sestávalo ministerstvo železnic v roce 1912?

Co bylo důležité pro hospodářský rozvoj koncem 19. století?

Které kompetence vykonávalo ministerstvo?

Obr. č. 16: odpověď na dotaz Kdy bylo zřízeno ministerstvo železnic v Předlitavsku (ilustrativní ukázka)

Autorizovaná forma jména	Ministerstvo železnic Vídeň
Dějiny	Vzhledem k výstavbě nových státních železničních tratí, budování místních drah a zestátnování železnic bylo vyčleněním agendy železnic z Ministerstva obchodu k 15.1.1896 zřízeno samostatné ministerstvo. Současně byl vyhlášen nový organizační statut státní železniční správy rakouské části monarchie (Předlitavska). Zaniklo rozpadem Rakousko-uherské monarchie ke dni 15.11.1918.
Funkce, zaměření, činnost	Úkolem ministerstva bylo nejvyšší jednotné vedení a správa státních železnic v Předlitavsku, jejich výstavba a provoz drah státem na účet vlastníka, jakož i paroplavba a trajekty na Bodamském jezeře. Vykonávalo též státní dohled nad železnicemi Předlitavska. Hlavními kompetencemi ministerstva jako vrchnostenského orgánu byly: <ol style="list-style-type: none"> 1) mezinárodní vztahy, účast na kongresech 2) železniční legislativa 3) zestátnování železnic
Vnitřní struktury	Ministerstvo mělo obvyklou strukturu nejvyšších rakouských úřadů. Členilo se na presidium a sekce, nižší organizační útvary byly departmenty (průběžně číslované skrz sekce) a samostatné postavení měla hlavní pokladna státních drah. Zpočátku to byly sekce I. pro záležitosti právně-administrativní a místní drah, II. pro věci finanční a statistické, III. pro záležitosti komerční a IV. pro věci technické. Postupně byly přidávány sekce a měněna jejich působnost, jejich čísla se však razantně nezvyšovala – k římské číslici se přidalo písmeno, obdobně tomu bylo u departmentů. V roce 1912 ministerstvo sestávalo z následujících sekcí: Ia – záležitosti železničních koncesí a státního dozoru nad drahami nižšího řádu; Ib – záležitosti železniční legislativy a státního dozoru na hlavní dráhy, právní záležitosti; IIa – administrativní záležitosti; IIb – záležitosti finančních a přepravních příjmů, železniční statistika; III – komerční záležitosti a cestovní ruch; IV – stavební záležitosti; V – doprava a strojní záležitosti.

⁶ viz http://www.archivescanada.ca/english/search/RouteRqst.asp?sessionKey=1143478259019_206_191_57_199&r=2&i=NA&i=0&v=0&coll=1&lvl=1&t=Dobson,+Will,+1914-1993

⁷ ukázka záznamu publikována se svolením autora Ing. Miroslava Kunta (NA)

Národní knihovna ČR spravuje a zpřístupňuje nejenom sbírky knihovních fondů, ale i významnou kolekci řemeslně-uměleckých předmětů (glóbusy, hodiny, památkový nábytek, mříže, dveře, regály apod.), které jsou evidovány v Centrální evidenci sbírek Ministerstva kultury. I při vyhledávání informací obsažených v popisu jednotlivých kolekcí a jednotlivých exponátů lze potenciálně využít zdokonalený systém M-CAST.

dotazy:

Která sbírka byla nejstarším veřejným muzeem v českých zemích?

Kdo byl prvním prefektem sbírky historických glóbusů v Klementinské koleji?

Kdo sestavil řadu hodin a mechanických automatů?

Kdy bylo matematické muzeum v Klementinu zrušeno?

Jak se nazývají nejkrásnější hodiny v matematickém muzeu v Klementinu?

Kde se nacházejí exponáty sbírky památkového nábytku Národní knihovny?

Kolik slunečních hodin se nachází na budovách Klementina?

Obr. č. 17: Odpověď na dotaz Která sbírka byla nejstarším veřejným muzeem v českých zemích (ilustrativní ukázka)

centrální evidenci sbírek

Úvod Co je CES O sbírkách Vyhledávání Sbírka dne Kontakty English

Sbírka Národní knihovny ČR

Oborová část sbírky - podsbíрка:

Vlastník:
Česká republika

Správce:
Národní knihovna ČR, Klementinum 190, 110 00 Praha 1
 www:
<http://www.nkp.cz/>

Počet ev. čísel:
13 (vyřazeno: 0)

Archiválie:
Ne

Kult. památky:
Ano

Území:
Benátky, Norimberk, Amsterdam, Praha a střední Evropa

Období:
Sbírka historických glóbusů byla součástí sbírek původního matematického muzea klementinské koleje Museum mathematicum Collegii Clementini, kterou utřídil Kašpar Pflieger po r. 1723. Tato sbírka byla nejstarším veřejným muzeem v českých zemích a vyjadřovala zájem zdejších jezuitů o matematicko-astronomická studia. Astronomie navíc patřila do učebního plánu jezuitské univerzity. Sbírka obsahovala vědecké přístroje užívané

Tip: Pro zobrazení ve větším rozlišení klikněte na zvolený obrázek

7 Závěr

Výsledky projektu M-CAST prokázaly možnosti uplatnění systému dotazů v přirozeném jazyce v prostředí digitálních i hybridních knihoven, a nově i v dalších paměťových institucích. Technologie zpracování přirozeného jazyka (natural language processing) se úspěšně uplatňují v oblasti analýzy dotazů, indexování dokumentů a extrakce otázek i ve vícejazyčném prostředí. Budoucím záměrem je rozvíjet a zlepšovat systém M-CAST v několika směrech. V současné době dokáže systém zodpovědět zhruba 70 % faktografických otázek a 30–40 % nefaktografických otázek, a to ve francouzštině a portugalštině. Nyní je potřeba soustředit se na to, aby i ostatní jazyky dosáhly stejného procenta zodpovězených faktografických otázek, a současně zvýšit výrazně poměr zodpovězených nefaktografických otázek ve všech jazycích. V neposlední řadě je třeba zapojit do systému další jazyky (uvažuje se o němčině), včetně jazyků nelatinkového písma (arabština, čínština). Projekt je v současné době formálně ukončen; nyní se hledají možnosti další kooperace a především financování. V Národní knihovně práce na projektu pokračují. V současné době se Národní knihovna zaměřuje na vytváření předpokladů pro aplikaci systému v hybridních knihovnách všech typů: probíhá výzkum dalších možností integrace klasifikačního systému MDT, připojují se údaje obsahů v rámci projektu

TOC. Dosavadní vývoj systému M-CAST ve srovnání s podobnými projekty ukazuje, že zvolená řešení jsou správná a perspektivní. Tímto však byla realizována první fáze: další vývoj vedoucí ke standardnímu a rutinnímu využití systému je teprve před námi.

Literatura

AMARAL, Carlos; LAURENT, Dominique. *Implementation of a QA system in a real context* : [Paper Presented at TEL-ME-MOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] [online]. [cit. 2007-10-19]. Dostupný z WWW:

<http://knihovnam.nkp.cz/docs/telmemor/subject/TellMeMore_version_2.ppt?PHPSESSID=e93ac950899817d81e359e42355fb5ae>.

BALÍKOVÁ, Marie. *M-CAST in libraries* : [Paper Presented at TEL-ME-MOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] [online]. [cit. 2007-10-19]. Dostupný z WWW: <http://knihovnam.nkp.cz/docs/telmemor/subject/M-CAST_in_libraries.ppt?PHPSESSID=e93ac950899817d81e359e42355fb5ae>.

CZERNIEJEWSKI, Borys. *Multilingual Content Aggregation System based on TRUST Search Engine (M-CAST)* : [Paper Presented at TEL-ME-MOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] [online]. [cit. 2007-10-19]. Dostupný z WWW: <http://knihovnam.nkp.cz/docs/telmemor/subject/M-CAST-project_presentation-final.ppt?PHPSESSID=e93ac950899817d81e359e42355fb5ae>.

LISEK, Sebastian. *P2P networks for distributed queries* : [Paper Presented at TEL-ME-MOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] [online]. [cit. 2007-10-19]. Dostupný z WWW: <<http://knihovnam.nkp.cz/docs/telmemor/subject/M-CAST-P2P.ppt?PHPSESSID=e93ac950899817d81e359e42355fb5ae>>.

STROSSA, Petr. *Information Query Formulation in a Slavonic Language and its Automatic Processing : Experience from Polish and Czech in comparison to Western European Languages* : [Paper Presented at TEL-ME-MOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] [online]. [cit. 2007-10-19]. Dostupný z WWW: <<http://knihovnam.nkp.cz/docs/telmemor/subject/TEL-ME-MOR-PS.ppt?PHPSESSID=e93ac950899817d81e359e42355fb5ae>>.

PALA, Karel. *Počítačové zpracování přirozeného jazyka* [online]. Brno, 2000. [cit. 2007-10-19]. Dostupný z WWW: <http://nlp.fi.muni.cz/poc_lingv/pala_zprac.pdf>.

BOLDIŠ, Petr. Vyhledávače : současné problémy a trendy vývoje. *Knihovna plus* [online]. 2005, č. 1 [cit. 2007-11-19]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovnaplus51/boldis.htm>>.

STROSSA, Petr. Komunikace mezi člověkem a počítačem v přirozeném jazyce. *Science WORLD* [online]. Praha, 2004 [cit. 2007-11-19]. Dostupný z WWW: <<http://scienceworld.cz/sw.nsf/ID/10D74E2E7ED7559EC1256F32005ACF20?OpenDocument>>.

BALÍKOVÁ, Marie; STROSSA, Petr; VŘEŠŤÁLOVÁ, Dana. *Dotazování v přirozeném jazyce : zkušenosti s aplikací prototypu systému M-CAST v českém prostředí*. Praha : Národní knihovna České republiky, 2007. 78 s. ISBN 978-80-7050-537-3.