

System M-CAST (Multilingual Content Aggregation System — Vícejazyčný systém agregace informací)

Marie BALÍKOVÁ
Národní knihovna České republiky, Praha
Marie.Balikova@nkp.cz
Petr STROSSA
Vysoká škola ekonomická, Praha
Kizips@vse.cz

INFORUM 2007: 13. konference o profesionálních informačních zdrojích
Praha, 22. - 24.5. 2007

Abstrakt

Vícejazyčný systém agregace informací (M-CAST, Multilingual Content Aggregation System) umožní 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 umí v současnosti vyhledávat 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.

Úvod

M-CAST (Multilingual Content Aggregation System based on TRUST Search Engine) je projekt programu eContent (EDC 22249), jehož cílem bylo vyvinout vícejazyčný vyhledávací systém umožňující 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.

Projekt M-CAST byl financován Evropskou unií v rámci víceletého komunitárního programu na podporu rozvoje a užívání evropského digitálního obsahu v globálních sítích a současně překonávání jazykové různorodosti v informační společnosti.

Konsorcium

Podmínkou pro zahájení prací souvisejících s projektem bylo vytvoření konsorcia participujících subjektů:

Koordinátorem projektu se stala společnost

- Infovide S.A., Varšava, Polsko, v průběhu projektu přejmenována na Infovide-Matrix S.A.

Za vývoj lingvistických modulů, implementaci ontologie *TRUST* v jednotlivých jazycích a implementaci klasifikačního systému MDT v systému *M-CAST* odpovídají:

- TiP sp. z o.o., Katowice, Polsko — polský a český lingvistický modul (v kooperaci s VŠE Praha);
- Synapse Développement SARL, Toulouse, Francie — francouzský lingvistický modul;
- Priberam Informática Lda., Lisabon, Portugalsko — portugalský lingvistický modul zahrnující evropskou i brazilskou portugalštinu);
- Expert System S.p.A., Modena, Itálie — italský a anglický lingvistický modul;
- Vysoká škola ekonomická v Praze, Česko — český lingvistický modul (v kooperaci s TiP sp. z o.o., Katowice).

Systém *M-CAST* je testován ve dvou institucích:

- Polská internetová knihovna, Toruń, Polsko;
- Národní knihovna České republiky, Praha, Česko.

Systém *M-CAST* navazuje na výsledky dvou projektů, které byly financovány z Pátého rámcového programu EU pro rozvoj vědy a výzkumu:

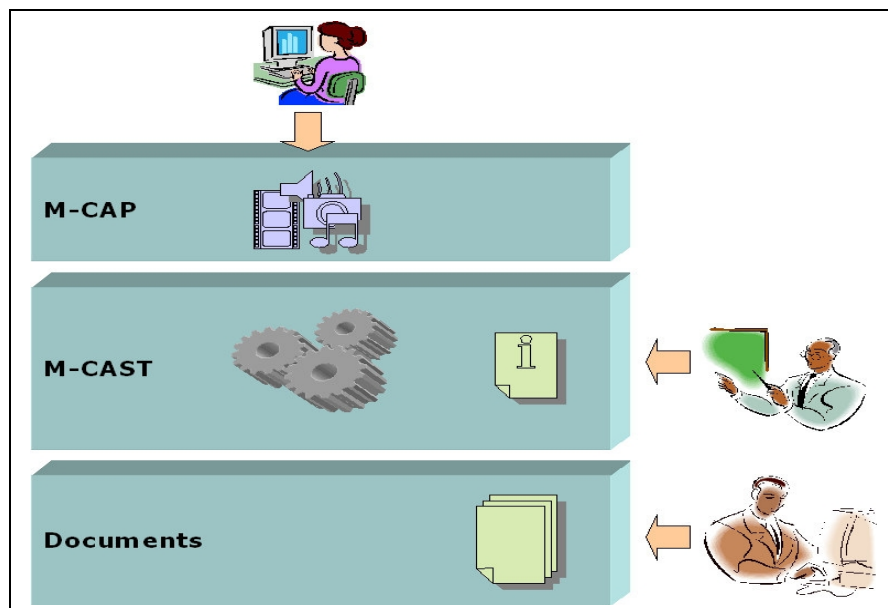
- *TRUST* — Multilingual Semantic and Cognitive Search Engine for Text Retrieval Using Semantic Technologies (Vícejazyčný sémantický a kognitivní mechanismus vyhledávání textů využívající sémantické technologie, IST-1999-56416);
- *ICONS* — Inteligentní systém pro správu obsahu (Intelligent Content Management System, IST-2001-32429).

Projekt *TRUST*

Vícejazyčný vyhledávací stroj *TRUST* umožňoval vyhledávat ve čtyřech jazycích: francouzštině, italštině, polštině a portugalštině. V rámci projektu *M-CAST* je rozšířen o dva další jazyky: angličtinu a češtinu. Jednouživatelská aplikace pro PC je transformována na serverovou aplikaci pro operační systém *UNIX* nebo *Windows*. Jazykové zdroje systému *TRUST* jsou aktualizovány a v systému nadále využity, jsou rozšířeny o další jazykové zdroje pokrývající potřebu nově připojených jazyků. Dosud využívaná jazyková ontologie (taxonomie) *TRUST* tvoří páteř vyhledávání i v projektu *M-CAST*. Možnosti vyhledávání jsou obohaceny o alternativu pokročilého hledání založenou na standardním mezinárodním desetinném třídění (MDT), celosvětově používaném v knihovních systémech.

Portál systému *M-CAST*, zv. také *M-CAP*, je vyvinut podle zásad metodologie návrhu systémů pro správu obsahu založených na znalostech (Knowledge-based Content Management Application Design Methodology). Tuto metodologii vytvořila firma Infovide, S.A. v rámci jiného projektu financovaného z Pátého rámcového programu EU — *ICONS*.

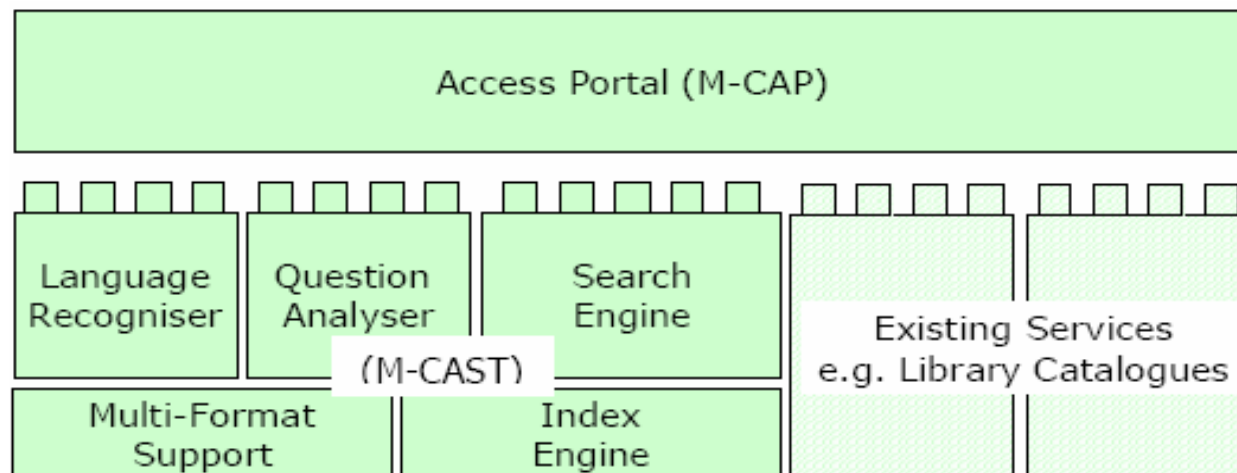
Základní schéma systému M-CAST



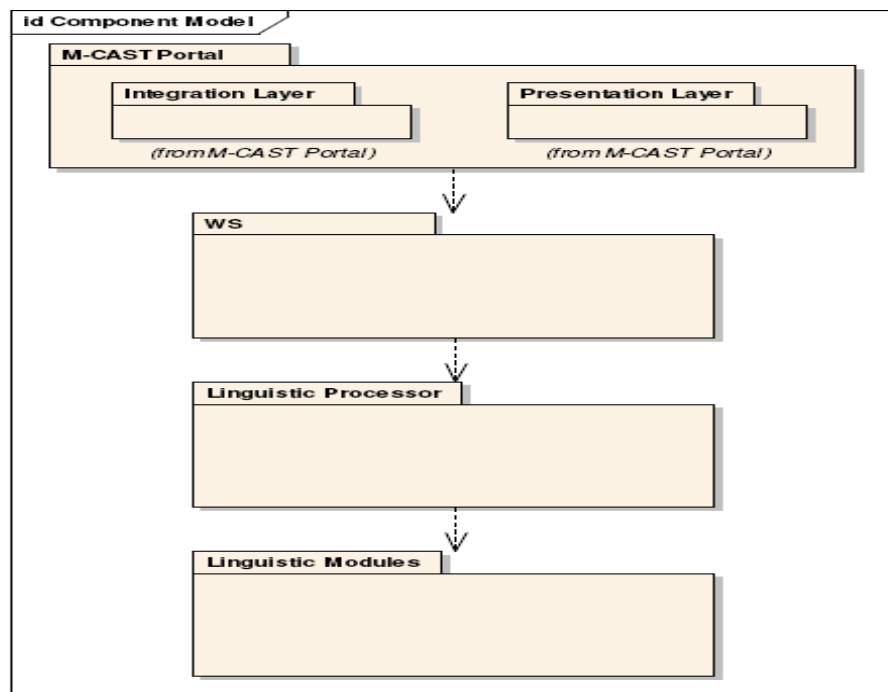
Portál M-CAST/M-CAP

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, uživatelé si mohou obsah přizpůsobit.

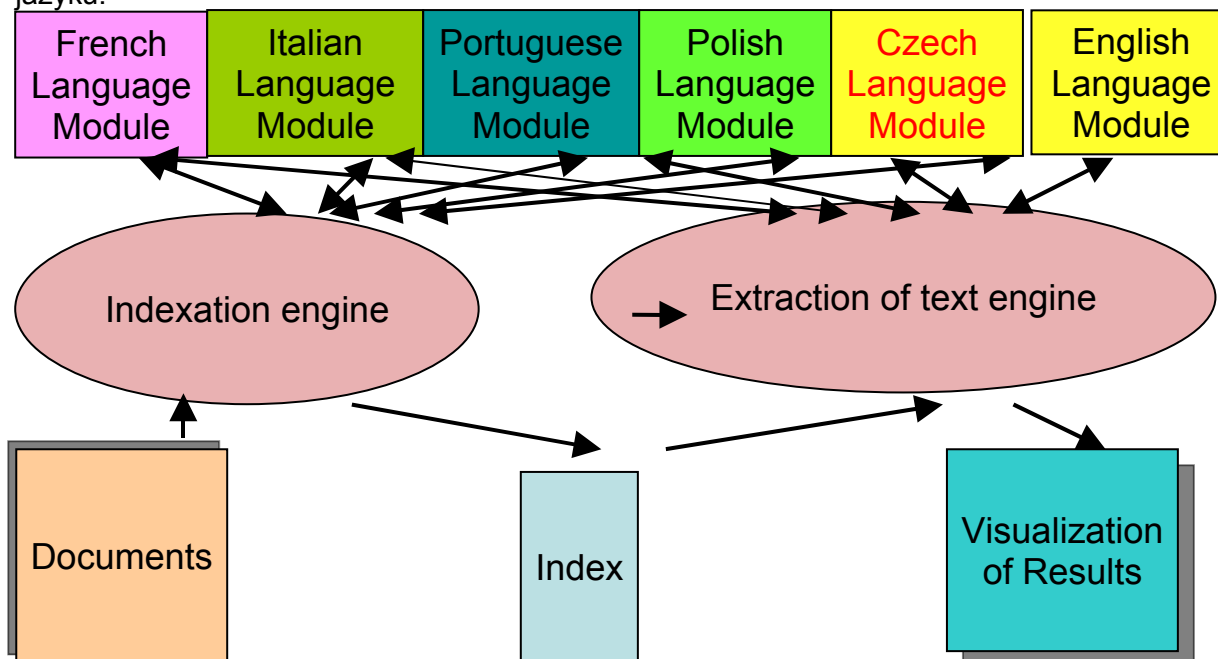
Portál *M-CAST/M-CAP* je koncipován tak, aby umožnil integraci stávajících vyhledávacích nástrojů používaných v dané instituci. Na obrázku vidíme návrh potenciálního využití systému *M-CAST* v knihovně, která chce své stávající vyhledávací možnosti obohatit o rešeršní strategii dotazování v přirozeném jazyce a zvolí portál *M-CAST/M-CAP* jako základní nástroj. Je možný i obrácený postup: instituce zahrne portál *M-CAST/M-CAP* jako součást stávajícího portálu.



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 bussines 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.



Lingvistický procesor představuje nejdůležitější součást systému M-CAST. Obsahuje nástroj pro detekci jazyka, analyzátor dotazů, konvertor formátů dokumentů a lingvistické moduly jednotlivých jazyků.



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.

Indexace dokumentů

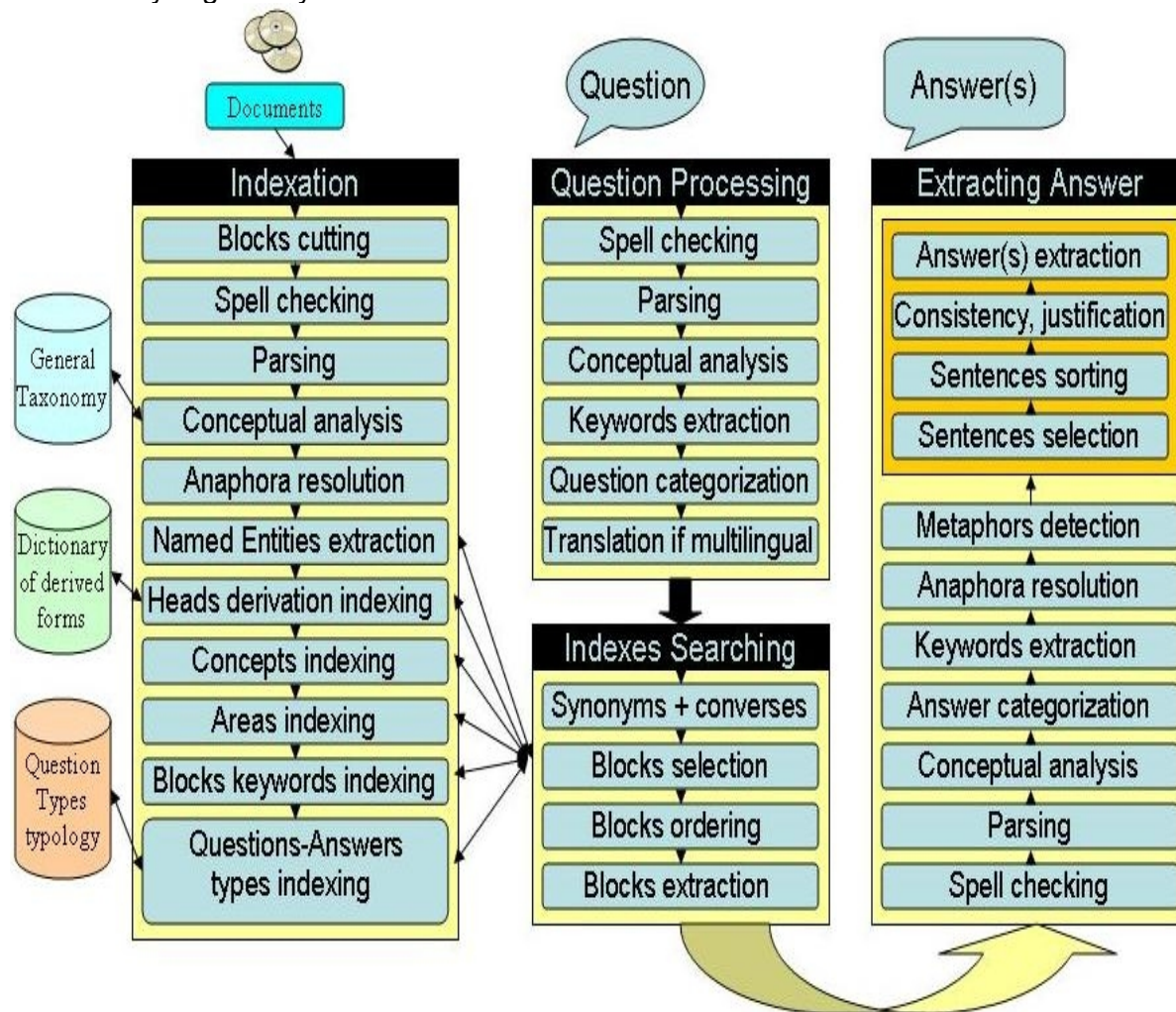
Společné rysy jsou patrné především v oblasti indexování dokumentů. Základním předpokladem pro úspěšné indexování dokumentů je *tokenizace*, kdy je text rozložen na základní selekční jednotky a v textu jsou identifikována slova, mezery, interpunkce a začátky a konce vět, dále *stemming* (*lem(m)atizace*), kdy se odstraňuje zakončení slova a ponechává se kmen / kořen (slovní základ), resp. (při lemmatizaci) je určena pro každý slovní tvar jeho základní podoba. V textech se provádí morfologická disambiguace 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. 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.

Proces indexace je ve všech jazycích stejný, extrahovaná data jsou stejné kategorie, zpracování těchto dat je tedy nezávislé na původním jazyce.

Složitost (maximální¹) struktury lingvistických modulů je patrná z následujícího schématu pro francouzský lingvistický modul:



Extrakce odpovědí

Uživatelský dotaz je podroben syntaktické a sémantické analýze. Je určen typ dotazu. Výsledek sémantické analýzy dotazu může být negativně ovlivněn tím, že kontext dotazu je poměrně malý, dotaz je na rozdíl od dokumentů výrazně kratší.

Na základě sémantické analýzy dotazu jsou podle jejich váhy stanovena významově důležitá klíčová slova, „pivots“. Při vyhledávání se používají tyto výrazy, obohacené o synonyma, 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 statistickolingvistických metod a jejich pořadí.

Pro vícejazyčné vyhledávání v systému *M-CAST* je jako propojovací jazyk použita angličtina.

¹ Ve skutečnosti ne všechny lingvistické moduly uplatňují toto schéma v celé jeho komplexnosti.

Český lingvistický modul v systému M-CAST

Český lingvistický modul byl vyvíjen v těsné návaznosti na modul polský, v úzké spolupráci se dvěma firmami zaměřenými na lingvistické technologie: polskou TiP, garantem polského modulu, a portugalskou Priberam Informática, jejíž programové nástroje jsme se po zvážení různých možností rozhodli použít jako finální řešení pro zpracování polštiny i češtiny. Během práce se potvrdilo, že tyto nástroje skutečně lze použít i pro zpracování slovanských jazyků potřebné v systému M-CAST, přestože byly původně navrženy pro portugalsčinu. Některé zvláštnosti slovanských jazyků si nicméně vyžádaly menší úpravy zvolených nástrojů.

Vzorce pro rozpoznávání typů dotazů a potenciálních odpovědí

Základní lingvistickou datovou strukturu podporující funkce systému M-CAST představují formální definice předem vyčleněných kategorií dotazů a potenciálních odpovědí na ně v indexovaných textech. Rozlišujeme 86 sémantických typů dotazů a odpovědí, ke kterým v rámci vývoje svých technologií a prvků systému M-CAST dospěl francouzský partner, firma Synapse Développement. Definice pro jejich rozpoznávání v češtině formulujeme pomocí nástroje *SintaGest* (čti „sintažešt“) portugalské firmy Priberam Informática. Následující příklad ukazuje jednu takovou definici — pro dotazy a odpovědi týkající se hmotnosti.

Question(WEIGHT)

: Root("jaký")? Dist(0,5) WeightNoun = 20
// Jaká je hmotnost Země?

: Wrd(jak) WeightAdj = 20
// Jak těžký může být slon?

: Wrd(kolik) WeightUnit = 20
// Kolik kg má dospělý kapr?

: Wrd(kolik) Root("vážit") = 20
// Kolik váží kapr?

Answer

: WeightNoun Definition With Pivot Dist(0,5) {Number6 WeightUnit}
= 20
// Váha kapra může dosáhnout až 5 kg.

: Pivot Dist(0,5) Cat(V) Dist(0,5) {Number6 WeightUnit} = 20
// Roční kapr může dosáhnout 5 kg tělesné váhy.

;

Answer(WEIGHT)

: Number6 WeightUnit = 20

;

Blok nadepsaný „Answer(WEIGHT)“ se uplatňuje při předběžném rozpoznávání potenciálních odpovědí na tento typ dotazu ve fázi indexování textů ukládaných do databáze. V tomto konkrétním případě to znamená, že každý text obsahující výraz jako např. „dva kilogramy“ bude označen jako text obsahující potenciální odpověď na dotaz typu WEIGHT (HMOTNOST).

Blok nadepsaný „Question(WEIGHT)“ se uplatňuje při klasifikaci položených dotazů a následném vyhledávání skutečných odpovědí na konkrétní dotaz typu WEIGHT (HMOTNOST — v textech předem vyhodnocených jako potenciálně relevantních k tomuto typu dotazu). Rozpoznávání dotazů daného typu je založeno na třech hlavních kategoriích formálních výrazů: **Root(x)** — zastupuje libovolný tvar, případně odvozeninu slova x, kterou jako takovou rozpozná *morfologický modul*;

Wrd(x) — slovní tvar x (přesně tak, jak je napsán — to obecně zjednodušuje systému práci s neohebnými slovy, např. s předložkami nebo příslovci typu „jak“, ale někdy může být vhodné i pro rozpoznávání určitých ustálených frází, které se v určitých typech kontextů fakticky neohýbají);² **konstanty** pojmenované určitým jménem (jako např. „WeightNoun“) jsou definovány v pomocném souboru — např. takto:

```

Const  WeightNoun = AnyRoot(hmotnost, hmota, "tíha", "váha",
                             "zatížení");
Const  WeightAdj = AnyRoot("těžký", "lehký");
Const  WeightUnit1 = AnyRoot(mikrogram, miligram, centigram,
                             decigram, gram, dekagram, hektogram, kilogram, kilo, cent,          megagram,
                             miligram, tuna, "karát", pond, kilopond, megapond,                libra);
Const  WeightUnit2 = AnyWrd(mg, cg, dg, g, dag, deka, Dg, dkg, hg,          kg, q, Mg, t, p,
                             kp, Mp, lb, "lb.", lbs, "lbs.", cwt,          "cwt.");
Const  WeightUnit = AnyConst(WeightUnit1, WeightUnit2);

```

Výraz „Cat(x)“ představuje (sám o sobě) jakékoli slovo, jemuž byla ve fázi morfologicko-lexikální analýzy přiřazena gramatická kategorie x (konkrétně např. „V“ — sloveso). Pomocný výraz „Dist(a, b)“ vyjadřuje, že mezi jinými dílčími výrazy se může nacházet jistý počet (a až b) jiných slov.

Za oddělovačem „=“ je vždy uvedeno číslo vyjadřující relativní míru, v jaké text pokrytý daným vzorcem odpovídá danému typu dotazu. (To lze využít v případech, kdy určité obraty mohou vyjadřovat různé typy dotazů, ale některé pravděpodobněji.) Za oddělovači „/“ jsou jako komentáře k formálním vzorcům uvedeny konkrétní příklady dotazů, které tyto vzorce pokrývají.

Výrazy uzavřené ve složených závorkách — „{...}“ — v prostředním bloku, nadepsaném jen „Answer“, by měly představovat *přesnou odpověď na daný dotaz* (např. na dotaz typu HMOTNOST je to zpravidla výraz vyjadřující počet a jednotku hmotnosti). Výraz „Definition With Pivot“ v témže bloku zastupuje jakoukoli jmenovou frázi (rozpoznanou opět podle nějakých pravidel v pomocném souboru) *obsahující alespoň jedno plnovýznamové slovo z dotazu*.

SintaGest je vlastně účelový editor a kompilátor definičních souborů naznačeného typu. Pomocí dalších programových nástrojů firem Priberam Informática a TiP jsou tyto soubory spolu s tabulkami definujícími morfologii (tvarosloví) a některými dalšími podpurnými datovými soubory kompilovány do českého lingvistického modulu, spolupracujícího s indexovacím a vyhledávacím strojem systému *M-CAST*.

Problém volného slovosledu v češtině

Čeština patří podle zavedené lingvistické typologie mezi tzv. *jazyky s volným slovosledem*. To sice neznamená, že by mohlo pořadí slov ve větě, která má vyjadřovat určité konkrétní sdělení, být úplně libovolné, nicméně dotaz na určitou věc a stejně tak odpověď na takový dotaz mohou být vyjádřeny určitými slovy v různém uspořádání. Tím se čeština (a jí podobné jazyky, jako např. polština, ruština aj.) dost výrazně liší od západoevropských jazyků typu angličtiny, francouzštiny nebo portugalštiny.

Například otázka po vzdálenosti Brna může být formulována — snad po řadě v klesající „míře obvyklosti“, ale stále gramaticky přípustně — těmito způsoby:

- Jak daleko je do Brna?
- Jak je daleko do Brna?

² Pokud bychom tímto způsobem chtěli například definovat frázi „vzít nohy na ramena“, mohli bychom asi použít formální výraz „Root(vzít) Dist(0,3) Wrd(nohy) Wrd(na) Wrd(ramena)“.

- Jak je do Brna daleko?
- Do Brna je jak daleko?
- Do Brna jak je daleko?
- Do Brna je daleko jak?
- Daleko je do Brna jak?

Tento problém je prozatím řešen dvěma způsoby:

- 1) Některá slova, jako např. tvary sponového slovesa „být“, ale i některá jiná slovesa, která sama o sobě k rozpoznání typu dotazu/odpovědi nepřispívají, nejsou ve vzorcích uváděna vůbec (nahrazujeme je „distančním operátorem“), anebo jsou uváděna paralelně na různých místech jako nepovinné prvky (s doplňujícím operátorem „?“; ve výše uvedené ukázce nepoužitým).
- 2) V souboru definic typů dotazů jsou uváděny některé zásadní slovosledné varianty jako samostatné vzorce.

Morfologická analýza

Nutným předpokladem fungování výše popsaných vzorců pro rozpoznávání typů dotazů a odpovědí je morfologická (tvaroslovná) analýza indexovaných textů a pokládání dotazů. Obecným cílem morfologické analýzy je určit pro každý slovní tvar v každém textu, o jaký slovní druh a tvar jde a jak zní odpovídající základní (slovníkový) tvar. Ve stávajícím provedení vzorců pro rozpoznávání dotazů a odpovědí ovšem téměř nepoužíváme konkrétní hodnoty kategorií jako pád, číslo, osoba ap. Z výsledků morfologické analýzy vlastně využíváme prakticky jen údaje o slovním druhu a základním tvaru (lemmatu) slova. Možnost využití přesnějších morfologických kategorií však zůstává otevřená — použitý formální jazyk rozpoznávacích vzorců to umožňuje, stejně jako morfologický analyzátor, který je součástí lingvistického modulu. (Srovnej ovšem následující oddíl Problém gramatické homonymie v češtině.)

Morfologický analyzátor češtiny 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 uchováající a rozpoznávající všechny tvary všech slov.) Tabulky definující vzory mají pochopitelně odlišnou strukturu podle slovních druhů. U českých podstatných jmen rozlišujeme 2 čísla a 7 pádů, takže např. tabulka definující vzor „žena“ by na první pohled mohla vypadat následovně (symbol „–“ reprezentuje *prázdnou koncovku*):

Pád	Koncovka jednotného čísla	Koncovka množného čísla
1.	a	y
2.	y	–
3.	ě	ám
4.	u	y
5.	o	y
6.	ě	ách
7.	ou	ami

Ve skutečnosti je však už jen skloňování českých podstatných jmen (resp. jeho exaktní popis, který potřebujeme) o něco složitější. U některých slov dochází v některých tvarech k určitým *kmenovým změnám*, které jsou poměrně snadno automaticky realizovatelné, nicméně musí být přesně popsány

v tabulkách definujících vzory. Tak například pro slova jako „babka“ potřebujeme vzor definovaný touto tabulkou:

Pád	Jednotné číslo		Množné číslo	
	kmen	koncovka	kmen	koncovka
1.	0	a	0	y
2.	0	y	+E	–
3.	+M0	e	0	ám
4.	0	u	0	y
5.	0	o	0	y
6.	+M0	e	0	ách
7.	0	ou	0	ami

Symbols v šedě vystínovaných sloupcích této tabulky reprezentují *variantu kmene*, ke které se má připojovat příslušná koncovka: „0“ znamená základní kmen, „+M0“ jeden ze způsobů *měkčení* koncové souhlásky kmene (tak vznikne tvar „babce“, anebo — podle stejného exaktního vzoru — např. „hře“),³ „+E“ vložení písmena e mezi poslední dvě písmena kmene (tak vznikne tvar „babeč“, nebo „her“). Elementárních typů kmenových změn rozlišujeme v tomto pojetí asi deset, některé z nich se můžou vzájemně kombinovat.⁴

Za předpokladu, že chceme *každému* českému podstatnému jménu ve slovníku přiřadit nějaký vzor exaktně popisující jeho skloňování (i kdyby to měl být vzor právě jen pro toto jediné podstatné jméno), potřebujeme celkem přibližně 300 (!) vzorů jen pro skloňování podstatných jmen. Tyto vzory se ovšem ani zdaleka neuplatňují se srovnatelnou frekvencí. Při sestavování slovníku se na jedné straně ukázalo, že 19 nejfrekventovanějších vzorů stačí k popisu skloňování 95 % českých podstatných jmen a 56 vzorů pokrývá 99 % podstatných jmen. Na druhé straně zhruba 150 (tj. polovina) definovaných vzorů popisuje skloňování jen 1–3 podstatných jmen, přibližně 80 vzorů popisuje skloňování jediného podstatného jména.⁵

Vzory popisující ohýbání českých přídavných jmen a sloves v naznačeném pojetí mají ve srovnání se vzory pro podstatná jména výrazně složitější strukturu. Při popisu úplného schématu ohýbání přídavného jména je třeba vycházet z toho, že se tu základní počet 14 tvarů, rozlišitelných u podstatného jména (2 čísla × 7 pádů), nejprve násobí čtyřmi, protože přídavné jméno (princiálně) rozlišuje tvary *4 rodů* (mužského životného, mužského neživotného, ženského a středního), a dále ještě třemi, protože všechny dosud počítané tvary existují ve *3 stupních*! Tím dostáváme *celkem 168 (!)* principiálně rozlišitelných tvarů přídavného jména. Je ovšem pravda, že mnohé z těchto tvarů systematicky splývají (např. rod mužský životný a neživotný se liší jedině ve 4. pádě jednotného čísla a v 1. pádě množného čísla, a rod střední se skoro stejně málo liší od rodu mužského neživotného) a některé tvary jsou vzájemně převeditelné jednoduchými formálními pravidly (např. tvar 3. stupně se

³ Exaktní realizace kmenové změny označené „+M0“ probíhá podle tohoto pravidla: $g, h \rightarrow z$; $ch \rightarrow š$; $k \rightarrow c$; $r \rightarrow ř$.

⁴ Například k tomu, abychom správně nadefinovali tvar „dvoře“ ke slovu „dvůr“, musíme specifikovat, že před koncovkou „e“ se kmen mění hned dvěma způsoby: 1) *krátí se* kmenová samohláska ($ů \rightarrow o$); 2) *měkčí se* koncová souhláska ($r \rightarrow ř$).

⁵ Mezi takové speciální vzory patří mj. např. i vzor pro slovo „člověk“, jehož tvary množného čísla „lidé“, „lidí“ atd. chápeme jako uplatnění ojedinělé kmenové změny formálně definované pravidlem *člověk* → *lid*. Ale nejde zdaleka jen o takové případy, tradičně chápané jako „absolutní nepravdivosti“. Ke vzorům, pro jejichž uplatnění jsme nenašli žádné jiné slovo, patří třeba i vzor pro slovo „vůl“ (varianta tradičního vzoru „pán“ s krácením kmenové samohlásky ve všech tvarech kromě 1. pádu jednotného čísla). Ostatně poněkud překvapivě se ukázalo, že velmi nízkou frekvenci — v řádu několika jednotek — vykazuje i samotný vzor „pán“

v našem exaktním pojetí: krácení kmenové samohlásky jen v 5. pádě jednotného čísla („pane!“) totiž podléhá pouze toto slovo a několik jeho odvozenin („jemnostpán“, „milostpán“, „mocipán“, „zeměpán“ ap.)! Pro zhruba 800 dalších podstatných jmen, která se skloňují, jak se tradičně říká, „podle vzoru pán“, ale bez uvedené anomálie, musel být zaveden jiný vzor.

liší od tvaru 2. stupně vždy jen předponou *nej-*, bez ohledu na rod, číslo nebo pád). Kromě toho typy skloňování a stupňování českého přídavného jména jsou prakticky vzájemně nezávislé. V implementaci skloňování přídavných jmen vycházíme z následujícího schématu, rozlišujícího celkem 19 obecně odlišitelných tvarů (zde reprezentovaných očíslovanými koncovkami — konkrétně pro tzv. *tvrdá přídavná jména*; šedě vystínovaná políčka této tabulky odkazují na jiné tvary, se kterými jsou dané tvary vždy shodné):

	Jednotné číslo				Množné číslo			
	muž. živ.	muž. neživ.	stř.	žen.	muž. živ.	muž. neživ.	žen.	stř.
1. pád	(1) ý		(2) é	(3) á	(4) í	(5) é		(6) á
2. pád	(7) ého			(8) é	(9) ých			
3. pád	(10) ému			(11) é	(12) ým			
4. pád	= (7)	= (1)	= (2)	(13) ou	= (5)			= (6)
5. pád	= (1)		= (2)	= (3)	= (4)	= (5)		= (6)
6. pád	(14) ém			(15) é	(16) ých			
7. pád	(17) ým			(18) ou	(19) ými			

Celkem čtyři tabulky tohoto typu (pro standardní *tvrdé* a *měkké* skloňování přídavných jmen, označené po řadě *A1* a *A2*, a dále pro specifické skloňování *přívlastňovacích přídavných jmen* typu „otcův“ a „matčín“, označené *A3* a *A4*) se kombinují s jinak strukturovanými tabulkami definujícími stupňování — zároveň s odvozováním a stupňováním příslovčí, což se jeví jako výhodné řešení.⁶ Teprve tyto tabulky považujeme za „vlastní vzory ohýbání přídavných jmen“, protože v nich se projevuje mnohem větší variabilita tohoto slovního druhu v češtině. Následující tabulka je příkladem takového vzoru pro asi 2000 českých přídavných jmen typu „divoký“.

Skupina tvarů	Kmen	Přípona	Soubor koncovek
1	0	–	A1
2	+M0	–	
3	+M2 ⁷	ejš	A2
4	+M0	e	–
5	+M2	eji	–

Skupiny tvarů „1“ a „2“ v této tabulce představují dohromady soubor tvarů 1. stupně, přičemž „skupina tvarů 2“ tvoří ve skutečnosti jen jediný tvar, a to 1. pád množného čísla rodu mužského životního (s koncovkou č. 4 v tabulce skloňování přídavných jmen) — ten totiž jako jediný u některých přídavných jmen vyžaduje změnu kmene; konkrétně tak od slova „divoký“ vznikne tvar „divocí“. Tímto trikem se nám podařilo izolovat informace o potřebných kmenových změnách od skloňovacích tabulek *A1–A4*, což zjednodušuje celkový formát popisu ohýbání přídavných jmen.⁸

⁶ Četné úvahy o podstatě fungování gramatiky nejen v češtině, ale ve všech jazycích tradičně rozlišujících slovní druhy „přídavných jmen“ a „příslovčí“ vedou k závěru, že skutečná funkce příslovce odvozeného od přídavného jména není vlastně nic jiného než jakýsi (v češtině) „8. pád přídavného jména“, který nerozlišuje číslo a používá se ve specifických syntaktických vazbách (zatímco přídavným jménem se rozvíjí podstatné jméno, příslovcem se rozvíjí sloveso nebo přídavné jméno).

⁷ Exaktní realizace kmenové změny označené „+M2“ probíhá podle tohoto pravidla: *g, h* → *ž*; *ch* → *š*; *k* → *č*; *r* → *ř*.

⁸ U přívlastňovacích přídavných jmen typu „otcův“ jsme tento problém obešli tak, že celé sekvence *-ův*, *-ova* atd. definujeme technicky jako koncovky. Takové řešení sice není lingvisticky úplně „čisté“, ale přispívá k jednoduchosti datových struktur, na kterých je založen náš lingvistický modul.

Skupina tvarů „3“ představuje tvary 2. stupně („divočejší“ a další odpovídající pády), a zároveň pomocí přípony *nej-* odvozené odpovídající tvary 3. stupně. „Skupina tvarů 4“ je opět ve skutečnosti jediný tvar — příslovce v 1. stupni („divoce“); skupina tvarů „5“ představuje tvary 2. a 3. stupně příslovce („divočejí“, „nejdivočejí“).

Vzorů tohoto typu zahrnuje český morfologický modul systému *M-CAST* celkem asi 50. Z toho ovšem tři nejméně frekventovanější pokrývají ohýbání 78 % všech přídavných jmen, a jen 14 vzorů popisuje ohýbání alespoň 10 různých slov.⁹

Rovněž formát potřebný pro úplný a exaktní popis *časování sloves* vykazuje složitost srovnatelnou se zde uvedeným formátem pro přídavná jména.¹⁰ Zde už jen uvedeme, že v českém lingvistickém modulu systému *M-CAST* je použito zhruba 150 vzorů pro slovesa, přičemž ovšem asi pro 60 % českých sloves by postačovaly dva vzory — „dělat“ a „kupovat“. I další statistika by tu byla podobná jako u podstatných a přídavných jmen.

Otázka „objektivní existence“ a „správnosti“ některých tvarů v české morfologii

V některých tabulkách definujících komplexní vzory ohýbání českých přídavných jmen jsou některé ze skupin tvarů „3“–„5“ prázdné, což znamená, že se přídavná jména ohýbaná podle těchto tabulek nestupňují, resp. neodvozují se od nich příslovce. Tuto možnost však (opět pro maximální jednoduchost celého systému) využíváme jen u kategorií přídavných jmen, kde tato zvláštnost souvisí s jinými morfologickými specifiky. Například pro přídavná jména odvozená od sloves příponou *-ací* platí všeobecně, že se nestupňují a neodvozují se od nich příslovce, a navíc můžeme říci, že normální rodilý uživatel češtiny by ani nevěděl, jak by se od nich takové útvary daly odvodit. Proto takovým přídavným jménům přiřazujeme vzor nepřipouštějící ani stupňování, ani odvozování příslovcí. Na druhé straně přídavná jména odvozená od podstatných jmen příponou *-ový* obecně, čistě morfologicky nahlíženo, lze stupňovat a lze od nich odvozovat příslovce, i když u mnoha z nich to z hlediska sémantického či pragmatického nemá smysl, a proto se to prostě nedělá. (Jaký smysl by měl např. 2. stupeň „morovější“ nebo příslovce „morově“?) Takové jemnosti náš lingvistický modul nerozlišuje, přídavným jménům odvozeným příponou *-ový* je pro jednoduchost přiřazen jediný vzor, který připouští stupňování i odvozování příslovcí. Jinými slovy, náš lingvistický modul rozpoznává všechny tvary, které podle obecných formálních pravidel lze utvořit, i když se v praxi nepoužívají. To nepovažujeme za chybu, protože jako hlavní úkol našeho lingvistického modulu vidíme *rozpoznání slov, která někdo použil* v dobré víře, že mu bude porozuměno.

Uvedený princip, kterým jsme se obecně řídili při návrhu českého lingvistického modulu, se netýká jen vzorů pro přídavná jména, a také se netýká jen otázky existence či neexistence některých tvarů (např. tvarů 2. stupně) jako takových. Ze stejného důvodu naše vzory všeobecně připouštějí i některé *způsoby tvoření* některých tvarů, které např. *Pravidla českého pravopisu* prohlašují za nesprávné, nicméně víme, že je lidé (a nehodláme se zabývat otázkou, kolik je takových lidí a zda „představují vývojový trend“ nebo ne) používají. Speciálně například pro některé skupiny cizích vlastních jmen (jako „Goethe“, „Nietzsche“, „Tódžó“ nebo „Jacques“) jsme zavedli vzory skloňování, ve kterých bylo naším hlavním cílem, aby nechyběla žádná koncovka, kterou by někoho mohlo napadnout u takových jmen použít — bez ohledu na to, co o tom říkají naše příliš často se měnící, rozporuplně zdůvodňované a především nepřilíš ve veřejnosti vžitě normy.

Na druhé straně jsme si výše uvedeným principem zjednodušovali práci všude tam, kde se nám jevílo příliš namáhavé nebo neefektivní kategorizovat slova podle mnoha různých gramaticko-sémantických charakteristik, ze kterých by vyplývala praktická možnost nebo nemožnost výskytu některých tvarů. Například tvary přechodníku přítomného (resp. budoucího?) i minulého formálně připouštíme u všech

⁹ Pro zajímavost: tradiční na školách vyučovaný vzor „mladý“, zde ovšem chápaný komplexně, tedy včetně stupňování „mladší“, „nejmladší“, je v uspořádání podle klesající slovníkové frekvence až na 13. místě.

¹⁰ Je třeba si uvědomit, že do slovesné morfologie v širším pojetí patří vedle elementárních tvarů neurčitého, oznamovacího a rozkazovacího způsobu, minulých a trpných příčestí a přechodníků také (běžně používaná) příčestí činná, jako např. „dělající“ od „dělat“ — ale ta se opět skloňují ve všech rodech jako přídavná jména!

sloves, přestože je zřejmé, že v reálných textech v moderní češtině, pokud se vůbec na nějaké přechodníky narazí, půjde nejspíš jen o přítomné přechodníky nedokonavých sloves („dělaje“ ap.) nebo o minulé přechodníky sloves dokonavých („udělav“ ap.). Formální schopnost rozpoznat i tvary, které neexistují (jako „dělav“ a „udělaje“), nepokládáme za chybu, zejména když je spojená se zmenšením počtu vzorů zhruba o polovinu (zde tím, že principiálně nerozlišujeme vzory pro dokonavá a nedokonavá slovesa).

Problém gramatické homonymie v češtině

K typickým vlastnostem doprovázejícím jiné charakteristiky české morfologie (i morfologie blízké příbuzných jazyků, např. polštiny) patří poměrně velmi častá a široká *homonymie* (*nejednoznačnost*) gramatických tvarů slov. Uvažme například, jaké možné tvaroslovné interpretace má slovo

„kovové“,

pokud ho posuzujeme samostatně, nezávisle na kontextu. Ze zmiňovaných 168 v češtině obecně existujících tvaroslovných interpretací stupňovatelného přídavného jména zde můžeme okamžitě vyloučit všechny tvary 2. a 3. stupně. Ovšem ze zbývajících 56 tvarů se může jednat o tyto:

- 1., 4. nebo 5. pád jednotného čísla středního rodu;
- 1., 4. nebo 5. pád množného čísla mužského rodu neživotného;
- 1., 4. nebo 5. pád množného čísla ženského rodu;
- 2., 3. nebo 6. pád jednotného čísla ženského rodu;
- 4. pád množného čísla mužského rodu životného.

To je celkem 13 skutečně možných interpretací! V této souvislosti se projevil jeden koncepční nedostatek původních verzí portugalských lingvistických technologií ve vztahu ke slovanským jazykům. Představme si, že budeme během procesu analýzy zaznamenávat u každého rozpoznávaného přídavného jména čtyři samostatné morfologické parametry: PÁD, ČÍSLO, ROD a STUPEŇ. Pro naše slovo „kovové“ bude samozřejmě jednoznačně stanovitelná pouze hodnota parametru STUPEŇ („1.“). Podívejme se však pozorněji, co může být zapsáno do zbývajících tří parametrů:

- Za předpokladu, že neznáme hodnoty parametrů ČÍSLO a ROD, může parametr PÁD nabývat hodnot „1.“, „2.“, „3.“, „4.“, „5.“ i „6.“.
- Za předpokladu, že neznáme hodnoty parametrů PÁD a ROD (resp. — s upřesněním pro danou situaci irelevantním — že PÁD může mít jakoukoli hodnotu kromě „7.“, viz výše), může parametr ČÍSLO nabývat hodnot „jednotné“ i „množné“.
- Konečně za předpokladu, že neznáme hodnoty parametrů PÁD a ČÍSLO (resp. pro PÁD víme stále jen, že není „7.“), může parametr ROD nabývat hodnot „mužský životný“, „mužský neživotný“, „ženský“ i „střední“.

Samostatné parametry PÁD, ČÍSLO a ROD tedy v češtině naprosto nejsou schopné vzájemně upřesnit své hodnoty — a celkový počet možných interpretací reprezentovaných touto trojicí parametrů pro slovo „kovové“ vychází nikoli 13, ale $6 \times 2 \times 4 = 48$ — téměř čtyřikrát vyšší, než skutečně je! Takovým způsobem reprezentovanou morfologickou interpretaci slova prakticky nelze použít k následnému testování nějakých prvků syntaktických struktur — např. požadavku, aby celá jmenná fráze byla v 2. pádě: budou-li v textu za sebou následovat slova „kovové lešení“, pak pro slovo „kovové“ i pro slovo „lešení“ hodnota PÁD = „2.“ přichází v úvahu (je mezi možnými), a tedy interpretaci celé fráze „kovové lešení“ jako fráze v 2. pádě nelze formálně vyloučit! Z tohoto důvodu se také zdá, že uvádění požadavků na určitý pád nebo číslo určitých slov ve vzorcích pro rozpoznávání typů dotazů a odpovědí na ně (viz předchozí oddíl Vzorce pro rozpoznávání typů dotazů a potenciálních odpovědí) by prakticky nevedlo k žádnému zlepšení přesnosti rozpoznávání — ledaže

by byl v celém lingvistickém modulu použit jiný způsob vnitřního kódování morfologických informací o slovech, na kterém zatím zahájili společné práce naši polští a portugalské partneri v projektu.

Inherentní gramatická nejednoznačnost českých slovních tvarů ostatně může spolu s volným slovosledem v extrémních případech vést i k objektivní nejednoznačnosti celých vět bez ohledu na to, jak dobře nebo špatně máme zakódované morfologické informace o jednotlivých rozpoznávaných slovech. Oblíbeným příkladem tohoto jevu je věta:

„Ženu holí stroj.“

V této větě můžeme každé ze slov interpretovat jako určitý slovesný tvar, tedy přísudek, a tomu přizpůsobit interpretaci ostatních slov. Gramaticky správná jednoduchá věta v češtině přirozeně nemůže obsahovat dva určité slovesné tvary, ale protože všechna slova této věty jsou „vhodným způsobem“ homonymní, můžeme větu objektivně chápat jako složenou z těchto trojic slov (v základních slovníkových tvarech):

- *hnát(i), hůl, stroj,*
- *žena, holit, stroj,* nebo konečně
- *žena, hůl, strojit.*

Pro první a třetí lexikální trojici navíc přichází v úvahu různé možnosti zapojení podstatného jména *hůl* do struktury věty — protože jeho tvar „holí“ stále ještě připouští dvě gramatické interpretace, které se obě formálně hodí do daného kontextu: 7. pád jednotného čísla a 2. pád množného čísla. Celkově tedy existuje minimálně pět různých interpretací věty „ženu holí stroj“, z nichž některé vypadají v různé míře nesmyslně, ale čistě gramaticky jsou všechny korektní.

Výše uvedený příklad snad může vypadat poněkud vykonstruovaně. Proto uvedeme ještě jeden příklad objektivní homonymie slov v českých větách (i při zohlednění větného kontextu), o jehož realitě nemůže být pochyb. Navíc jde o příklad jednoduchého faktografického dotazu, čili jazykového útvaru, na jehož zpracování je specificky zaměřen systém *M-CAST*. Příklad zní:

„Jaký je plat Petra Hanka?“

Problém této tázací věty spočívá v tom, že ačkoli je člověku na první pohled jasné (a dokonce i počítač může na základě jednoduchých formálních pravidel rozpoznat), že se tazatel zajímá o nějakého člověka jménem *X* a příjmením *Y*, není zdaleka jednoznačné, jak zní základní tvar jména *X* a příjmení *Y*. Uvážíme-li pravidla, podle kterých se v češtině můžou skloňovat osobní jména českého, slovenského, německého, anglického a jihoslovanského původu, spolu s tím, jaká podobně začínající jména opravdu existují, vyjde nám, že základní tvar jména *X* by mohl být *Petr*, *Peter* nebo *Petar*. Podobně pro příjmení *Y* přichází v úvahu základní tvary *Hanek*, *Hank*, *Hanke* a *Hanko* (mimochodem všechny existující v pražském telefonním seznamu).

Ve výše uvedeném příkladu je však zvláště kuriózní, které dvě interpretace jmen *X* a *Y* z hlediska zdravého rozumu *nepřicházejí v úvahu*: totiž že $X = \text{Petra}$ a $Y = \text{Hanka}$ — přestože obě tato jména v češtině existují (a druhé z nich dokonce prokazatelně i jako příjmení)! To se ovšem dá formálně zdůvodnit jedině na základě rozpoznání syntaktické struktury celého dotazu: podmětem věty je slovo „plat“, za ním následuje vlastní jméno jako jeho přívlástek vyjadřující přináležitost („čí plat“), čili toto vlastní jméno je nutně v 2. pádě. Proto musí jít o některé z existujících jmen, jejichž 2. pád může znít „Petra“, resp. „Hanka“. V této poloze jde o další problém, který západoevropské jazyky neznají. Jeho na uživatele „inteligentně působící“ řešení by vyžadovalo:

- 1) předem zahrnout do slovníku všechna existující vlastní jména (což je přístup jistě možný, nicméně odlišný od koncepce všech dřívějších — západoevropských — jazykových modulů systému *M-CAST*, a podle našeho odhadu by to znamenalo několikanásobné zvětšení slovníku);

- 2) zpřesnit vzorce pro rozpoznávání typů dotazů a odpovědí na ně v míře, s jakou v projektech předcházejících projektu *M-CAST* ani v tomto projektu od začátku nikdo nepočítal.¹¹

V každém případě musíme konstatovat, že současná verze českého lingvistického modulu systému *M-CAST* nedokáže ve formulaci typu „jaký je plat Hanka“ rozpoznat, že nemůže jít o osobu jménem *Hanka*. To samozřejmě může uživateli připadat hloupé, a také v konečném důsledku snižovat celkovou přesnost nabízených odpovědí na dotaz. Z uvedených údajů by ovšem zároveň mělo být patrné, že snížení přesnosti samotným tímto detailem ve skutečnosti vůbec nemusí být rozhodující. Přitom úplné odstranění problému homonymie asi zvláště u tvarů vlastních jmen v češtině nikdy nebude možné jinak než pomocí funkce interaktivního upřesnění dotazu („Myslíte jméno: 1. Hanek; 2. Hank; ...“).

Možnosti vyhledávání v systému *M-CAST*

Systém *M-CAST* podporuje dva typy vyhledávání: jednoduché a pokročilé.

Při jednoduchém vyhledávání se zapíše do vyhledávacího boxu dotaz, kterým může být skupina termínů, nebo otázka v přirozeném jazyce. 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.

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

První úroveň:

- 0 Všeobecnosti. Informatika a informační vědy
- 1 Filozofie. Psychologie
- 2 Náboženství. Teologie
- 3 Společenské vědy. Statistika. Politika. Vláda. Ekonomie. Správa. Vojenství. Folkloristika
- 4 Neobsazeno
- 5 Přírodní vědy. Matematika
- 6 Aplikované vědy. Lékařství. Technika
- 7 Umění. Rekreační. Zábava. Hudba. Sport
- 8 Jazyky. Lingvistika. Literatura
- 9 Geografie. Biografie. Dějiny

Kategorie MDT si lze vybrat ze seznamu v prvním řádku kliknutím na rozbalovací šipku vpravo, ve druhém a třetím řádku se po kliknutí na šipku objeví nabídka kategorie nižší úrovně.

¹¹ Bylo by asi skutečně nutné rozpoznávat celou syntaktickou strukturu každého dotazu, nejen jeho prvky charakterizující jednotlivé typy dotazů a odpovědí na ně. Stačí například uvážit tento dotaz: „Jaký plat Hanka dává svým zaměstnancům?“ Zde jde o jméno (resp. příjmení) *Hanka* v základním tvaru, a navíc nikoli o přímé rozvíjení slova *plat* — nejde o plat Hanky, jde o plat zaměstnanců Hanky. V širším pohledu tento problém opět významně souvisí i s problémem volného slovosledu — jde tu mimo jiné i o to, nakolik je uvedená formulace (a) běžná a (b) ekvivalentní formulaci: „Jaký plat dává Hanka svým zaměstnancům?“ Srovnej oddíl Problém volného slovosledu v češtině.

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST život

Jednoduché vyhledávání
Nastavení
Nápověda

Hledat
Rozdělit

Kategorie MDT:

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

NKP

Žádná odpověď pro život

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

1 hnutí pro život ČR - potrat všechno změni

1 hnutí pro život ČR
Zdroj - 100%

2 Senior - relaxace (2. : 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
Zdroj - 100%

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

Ptejte se knihovny Hledání na WWW Copyright

M-CAST v NK ČR

Databáze indexovaných dokumentů

NK ČR nemá k dispozici bohaté reprezentativní sbírky digitálních či digitalizovaných dokumentů, jejichž existence je základním předpokladem a podmínkou úspěšné aplikace metody dotazování v přirozeném jazyce. Proto bylo zvoleno náhradní řešení.

Pro testovací fázi systému M-CAST zpřístupnila NK ČR tyto databáze:

- *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;
- *ALEPH* (sigma.nkp.cz), tj. elektronický katalog Národní knihovny České republiky; vyhledává se v popisných údajích, výsledkem hledání jsou metadatové záznamy, slouží k testování českého modulu;
- *EU Constitution* (europa.eu/constitution/) — Evropská ústava obsahuje texty ve všech jazycích M-CASTu, slouží tedy k testování vícejazyčného modulu.

Jak bylo uvedeno výše, NK ČR v současné době nevlastní rozsáhlé soubory vícejazyčných dat soudobých dokumentů, proto byla tato fáze testování v NK ČR zaměřena na ověření funkčnosti systému M-CAST při vyhledávání informací ve sbírce indexovaných historických textů. 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 NK ČR 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.

Kvalita OCR textů

Navíc 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. Při skenování textu z novin obvykle dochází k problémům při segmentaci znaků, k chybám při rozpoznávání textu z nekvalitní předlohy. OCR texty obsažené v databázi Kramerius byly indexovány pro potřeby *M-CASTu* bez předchozích úprav: nebyly odstraněny nepřesnosti při segmentaci textu a chybějící nerozpoznané znaky, proto mnohdy některá slova, případně fráze nebyly identifikovány správně a docházelo tak k významovým posunům, případně výsledkem dotazu nedával smysl.

Metoda dotazování v přirozeném jazyce aplikovaná v systému M-CAST

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. Dotazy aplikované v systému *M-CAST* mají být faktografické, jednoduché, jasně a přesně formulované. Všechny informace, které jsou předmětem dotazu, musejí být obsaženy v databázi indexovaných dokumentů.

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.

The screenshot shows a web browser window displaying the M-CAST system interface. The page title is "Který umělecký soubor vystoupí na zahájení výstavy Velká Morava v Berlíně?". The interface includes a search bar, a "Hledat" button, and a "Rozdělit" button. Below the search bar, there are search results for "Kulturní noviny" and "Národní listy". Each result includes a snippet of text and a percentage indicating the relevance of the snippet to the query. For example, the first result from "Kulturní noviny" has a relevance of 100% and discusses the opening of the "Velká Morava" exhibition in Berlin. The second result from "Národní listy" has a relevance of 49% and discusses the opening of the exhibition in Brno.

Odpovědi v systému M-CAST

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

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

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".

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, chtějí hlavu jeho věncem ozdobiti; v levé držel žezlo."

Vizualizace zdrojového dokumentu

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

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Co držel Zeus v pravé ruce?

vyhledej ve vzdálených repozitářích

Hledat Rozdělit

[Pokročilé vyhledávání](#)
[Nastavení](#)
[Nápověda](#)

NKP

Odpověď/Odpovědi:
bohyni vítězství
pro **Co držel Zeus v pravé ruce?**

1 - 10 z 64 nalezených odpovědí (0 min 16 sec) 1 2 3 4 5 6 7 Další>

1 Címrhanzi Tůma - Mythologie čili bájesloví Řekův a Římanův

- 1 V **pravé ruce držel Zeus** bohyni vítězství, jak se právě k němu sklání, chtějí hlavu jeho věncem ozdobit; v levé **držel** žezlo.
[Zdroj](#) - 100%
- 2 Jelikož byla Victoria sym- bolem vítězství, uvádí ji myto loggie v neustálém spojení s Jo- vem a Minervou, a proto také ji spatřujeme vyobrazenou, jak **Zeus** Olympický a Pal las Parte nos na **rukou** svých jí **drží**.
[Zdroj](#) - 64%
- 3 ROK vyobrazuje se 1. jako bohyně, která věncem z kvítí hlavu mající ověncenou v **pravé ruce drží** obilí a srp, v levé roh hojnosti.
[Zdroj](#) - 52%
- 4 Diana na oblacích sedíc **drží** luk v **ruce**, na zádech má toulec se šípy a na hlavě půlměsíc; vedle ní leží honící pes.
[Zdroj](#) - 40%
- 5 Hermes **drží** hůl „caduceus“ v **ruce**; za ním spatřujeme zastřené postavy lidské, stíny duší naznačující.
[Zdroj](#) - 40%
- 6 **Zeus** vznáší se v oblacích, u nohou jeho sedí orel s rozpatými křídly ma v drápech blesky **drže**.
[Zdroj](#) - 40%
- 7 Saturnu, starý okřídlený kmet dlouhým vousem zarostlý, **drží** přesýpací hodiny v **ruce**.

Ptejte se knihovny Hledání na WWW Copyright

http://kramerius.nkp.cz - Kramerius - Dokument stránky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Abeceda
Výběr monografie
Monografie
Stránka
31 / 243
PDF Vytvořit
rozsah stran: (max. 20)
31 - 243

LIZARDTECH

— 26 —

Neméně slaven byl **Jupiter** v Římě, maje hlavní chrám na **Kapitolu** — odtud **Capitolinus**. — Jemu na počest konány slavnosti **ludi Romani** (hry římské) v měsíci září, a **ludi Magni** (hry veliké) a j. V oběť dávání mu byli mladí, jhem ještě nedotknutí býci, barvy bílé. Při zvláštních slavnostech obětovány Jupiterovi **lékatomby**, t. j. 100 kusů obětí.

Obraz. Zeus vyobrazoval se jakožto statečný, vážný muž s korunou na hlavě. V jedné ruce má žezlo, v druhé blesky. Sedí na trůnu. U jeho nohou dle zamlouvaný jeho pták, orel, který v zobanu také blesky drží. — Nejdůstojnější podobizna Zeova byla v Olympii od Pheidia zhotovená ze zlata a slonové kosti. Trůn a plášť Zeovův típily se zlatem a jinými lesklými kovy, drahokamy a živými barvami. Sedící bůh byl asi 25 metrů vysoký a dotýkal se hlavou skoro až stropu chrámového. Hlava byla ozdobena olivovým věncem, jakožto symbolem odměny za vítězství při hrách olympických. Obléčel byl lidem řecké mužné, velebnosti božskou však zvýšené krásy. Svrchní část těla byla nahá, ostatní tělo přikrýval řínský plášť, který dosahoval až k nohám sandálí zlatými opatřeným, jež spočívaly na stolici zlatými lvičky držené. V pravé ruce držel Zeus bohyni vítězství, jak se právě k němu sklání, chtějí hlavu jeho věncem ozdobit; v levé držel žezlo. Za kreslem vystřihala ona tajemná pausa se lvím tělem Sphinx. Socha tato byla tak dokonalá a důstojná, že básník jeden při pohledu na ni se tázal: „Vystoupil Pheidias do Olympu, aby Zeva viděl, anebo sestoupil Zeus s nebe, aby Pheidiovi tvář svou ukázal?“ Každý Řek považoval to za velikou nehodu svého života, nebylo-li mu dopřáno vzácnou sochu tuto vlastním očima zřítí. Ze stromů posvátný mu byl dub a mezi ptáky orel.

Jupiter Ammon. Zeus Dodónský.

©2003-2006 NK ČR
developed by Qbizim

NK ČR je jedním z testovacích pracovišť systému M-CAST. Naším úkolem bylo ověřit funkcionalitu systému M-CAST v plném provozu, prověřit odezvu systému, provést různé zátěžové testy a svými praktickými zkušenostmi přispět k odstranění všech případných nedostatků. Na testování se podílelo 40 účastníků: všichni měli zkušenosti s vyhledáváním v online katalozích, faktografických databázích, v prostředí internet, avšak pouze 5 účastníků mělo dílčí zkušenosti s kladením dotazů v přirozeném

jazyce. Měli jsme tak možnost ověřit, je-li vyhledávání v systému M-CAST komplikované, nebo naopak snadné a intuitivní, jak proklamovali naši kolegové, producenti systému. Na základě získaných zkušeností můžeme potvrdit, že vyhledávání v systému M-CAST je intuitivní a snadné i pro méně zkušeného či nezkušeného uživatele.

Hodnocení — evaluace systému M-CAST

Odezva systému

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

Správnost odpovědí

Proces hodnocení byl ztížen tím, že systém *M-CAST* nabízí soubor/blok odpovědí, přičemž je všeobecně známo, že evaluace komplexu odpovědí generovaných systémem je obtížnější než evaluace jednotlivých typů odpovědí. Proto byla pro potřeby hodnocení systému vypracována tato speciální kategorizace odpovědí:

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

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

Optimální počet odpovědí

V průběhu testování se ukázalo, že optimální počet získaných odpovědí je 5. Někteří účastníci, obeznámení s vyhledávaným tématem, by se spokojili se 2 odpověďmi, zatímco větší počet odpovědí požadovali uživatelé, kteří nebyli podrobněji seznámeni s předmětem svého zájmu.

Frekvence výskytu správných přímých odpovědí i úryvků odpovědí byla ovlivněna menší kvalitou OCR textů (vysvětlení výše).

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

Závěr

Výsledky projektu *M-CAST* jsou v souladu s cíli evropského programu *eContent* v oblasti vícejazyčného vyhledávání. Prokázaly možnosti uplatnění systému dotazů v přirozeném jazyce v prostředí hybridních i digitálních knihoven. 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).

Bibliografie

Amaral, Carlos. Laurent, Dominique (2006) *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] Available at

http://knihovnam.nkp.cz/sekce.php3?page=07_Pro/08_TEL_ME_MOR/TelmeSubj06.htm

Balíková, Marie (2006) *M-CAST in libraries* [Paper Presented at TEL-ME-MOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] Available at

http://knihovnam.nkp.cz/sekce.php3?page=07_Pro/08_TEL_ME_MOR/TelmeSubj06.htm

Czerniejewski, Borys (2006) *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] Available at

http://knihovnam.nkp.cz/sekce.php3?page=07_Pro/08_TEL_ME_MOR/TelmeSubj06.htm

Lisek, Sebastian (2006) *P2P networks for distributed queries* [Paper Presented at TEL-MEMOR/M-CAST Seminar On Subject Access, Prague, November 24, 2006] Available at

http://knihovnam.nkp.cz/sekce.php3?page=07_Pro/08_TEL_ME_MOR/TelmeSubj06.htm

Strossa, Petr (2006) *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] Available at

http://knihovnam.nkp.cz/sekce.php3?page=07_Pro/08_TEL_ME_MOR/TelmeSubj06.htm

Příloha 1: Příklady

Otázka: Kdy byla Karlova univerzita zpřístupněna ženám?

Odpověď se skládá z přímé odpovědi, úryvku odpovědi a propojení na stránku zdrojového dokumentu
přímá odpověď: "od roku 1897"

úryvek odpovědi: "Od roku 1897 byla ženám zpřístupněna Karlova univerzita"

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Kdy byla Karlova univerzita zpřístupněna ženám? [Pokročilý vyhledávání](#)
 vyhledej ve vzdálených repozitářích [Nastavení](#)
[Hledat](#) [Rozdělit](#) [Nápověda](#)

NKP
Odpověď/Odpovědi:
1897 **Kdy byla Karlova univerzita zpřístupněna ženám?**
pro

1- 10 z 63 nalezených odpovědí (0 min 04 sec) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Další>](#)

1 Český časopis historický

- 10) Od roku 1897 byla ženám zpřístupněna Karlova univerzita.
[Zdroj](#) - 100%
- Prostřednictvím vydávaných archivních pomůcek bylo veřejnosti zpřístupněno archivní bohatství archivu **Univerzity Karlovy** a Král.
[Zdroj](#) - 61%
- Výnos ministra školství z 20. listopadu 1934 ukládal pražské německé **univerzitě**, aby podle § 4-6 zákona z 19. února 1920 (lex Mateš) vydala rektorátu **Karlovy univerzity** staré původní insignie **Karlovy univerzity**, jakmile o to Český rektorát požádá.
[Zdroj](#) - 56%
- Konflikt byl naprosto zbytečný, protože česká **univerzita** měla vlastní, ovšem nové, insignie už od rozdělení **Karlovy univerzity** v německou a českou (1882).
[Zdroj](#) - 56%
- Je uspořádán chronologicky a přináší údaje o osobách, které na lé-kařské fakultě **Karlovy univerzity** půsily jako pedagogové či akademičtí funkcionáři.
[Zdroj](#) - 55%
- Tam k nim promluvil rektor Drachovský, který zdůraznil, že **Karlova univerzita** si přeje předání insignií právní cestou bez de-monstrací.

Ptejte se knihovny Hledání na WWW Copyright

Internet

vizualizace stránky zdrojového dokumentu

http://kramerius.nkp.cz - Kramerius - Dokument stránky periodika - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Výběr titulu
Titul
Ročník
Výsledek
Stránka 74 / 164
PDF Vytvořit
rozsah strani (max: 20)
74 - 164

ROČNÍK 89 **ČESKÝ ČASOPIS HISTORICKÝ** ČÍSLO 2

ČESKÉ ŽENSKÉ HNUTÍ DO ROKU 1948

VĚRA BEDNÁŘOVÁ

České ženské hnutí se stalo součástí moderních dějin českého národa. Jeho počátky sahají do poloviny 19. století, kdy se počaly tříbit názory na poslání a práva žen v souvislosti s činností K. S. Amerlinga a V. Náprstka.¹⁾ Na Moravě se za počátek těchto snah považuje návštěva Boženy Němcové, F. M. Klácela a J. Helcelety v Brně roku 1851, kdy se s B. Němcovou seznámila mladá Veronika Vrbíková-Mikšíčková. Výsledkem tohoto setkání bylo, že Vrbíková roku 1870 založila „Vesnu, dívčí pěvecký spolek v Brně“.²⁾ Ale už roku 1869 v Prostějově byl založen spolek „Vlastimila“³⁾ a pak následovaly další ženské spolky v Cechách i na Moravě. Cílem těchto spolků už nebyly úvahy let předchozích, ale ženy si vytyčily přesný cíl umožnit dívkám vzdělání jako přípravu pro hospodářské osamostatnění. Proto byly velkým přínosem k uskutečnění této myšlenky spolky, které vydržovaly pokračovací školy pro dívky škole odrostlé. V Praze to byl „Ženský výrobní spolek“⁴⁾ založený roku 1871 Karolinou Světlou, v Brně průmyslové školy „Vzdělávacího spolku Vesna“⁵⁾ založené roku 1886 Eliškou Machovou, v Olomouci „Pöfitingeum“⁶⁾ založené roku 1895 a podobný spolek „Světla“⁷⁾ založený roku 1901 A. Pířmanem ve Velkém Meziříčí. R. Gautschovo ministerstvo na žádost německých žen v Rakousku uzákonilo šestileté dívčí lyceum.⁸⁾ Lycea však ženskému hnutí nevyhovovala a splněním jeho požadavků bylo až gymnázium „Minerva“⁹⁾ založené roku 1890 v Praze Eliškou Krásnohorskou. Morava dostala zásluhou Zdenky Wiedermannové dívčí gymnázium roku 1907 ve Valašském Meziříčí a roku 1908 v Brně.¹⁰⁾ Od roku 1897 byla ženám zpřístupněna Karlova univerzita. Po skončení 1. světové války v roce 1918 české ženské hnutí šlo za konkrétními požadavky získat pro ženy rovná práva ve spojení s ženským hnutím mezinárodním, v němž zaujalo významné postavení.¹¹⁾ Zásluhu o to měla senátorka Františka Plamínková, která ženské spolky spojila v „Žen-

©2003-2006 NK ČR, developed by Qbism

Otázka: Co držel Zeus v levé ruce?

Odpověď: žezlo

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Co držel Zeus v levé ruce? [Pokročilé vyhledávání](#)
 vyhledej ve vzdálených repozitářích [Hledat](#) [Nastavení](#) [Nápověda](#)
[Rozdělit](#)

NKP
Odpověď/Odpovědi:
žezlo
 pro **Co držel Zeus v levé ruce?**

1 - 10 z 66 nalezených odpovědí (0 min 11 sec) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Další>](#)

1 Cimrhanzl Tůma - Mythologie čili bájesloví Řekův a Římanův

- V pravé **ruce držel Zeus** bohyni vítězství, jak se právě k němu sklání, chtíc hlavu jeho věncem ozdobit; v **levé držel žezlo**.
[Zdroj](#) - 100%
- Jeikož byla Victoria sym- bolem vítězství, uvádí ji myto loggie v neustálém spojení s Jo- vem a Minervou, a proto také ji spatřujeme vyobrazenou, jak **Zeus** Olympický a Pal las Parte nos na **rukou** svých jí **drží**.
[Zdroj](#) - 73%
- ROK vyobrazuje se 1. jako bohyně, která věncem z kvítí hlavu majíc ověncenou v pravé **ruce drží** obilí a srp, v **levé** roh hojnosti.
[Zdroj](#) - 52%
- Diana na oblacích sedíc **drží** luk v **ruce**, na zádech má toulec se šípy a na hlavě půlměsíc; vede ní leží honící pes.
[Zdroj](#) - 45%
- Hermes **drží** hůl „caduceus“ v **ruce**; za ním spatřujeme zastřené postavy lidské, stíny duší naznačující.
[Zdroj](#) - 45%
- Zeus** vznáší se v oblacích, u nohou jeho sedí orel s rozpatými křídly ma v drápech blesky **drže**.
[Zdroj](#) - 45%

Ptejte se knihovny | Hledání na WWW | Copyright

http://kramerius.nkp.cz - Kramerius - Dokument stránky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda


Abeceda
 Výběr monografie
 Monografie
 Stránka
 31 / 243
 PDF Vytvořit
 rozsah stran: (max. 20)
 31 - 243

LIZARDTECH

— 26 —

Neméně slaven byl *Jupiter* v Římě, maje hlavní chrám na *Kapitolu* — odtud *Capitolinus*. — Jemu na počest konány slavnosti *ludi Romani* (hry římské) v měsíci září, a *ludi Magni* (hry veliké) a. j. V oběť dávali mu byli mladí, jhem ještě nedotknutí býci, barvy bílé. Při zvláštních slavnostech obětovány Jupiterovi *lékatomby*, t. j. 100 kusů obětí.

Obraz. Zeus vyobrazoval se jakožto statečný, vážný muž s korunou na hlavě. V jedné ruce má žezlo, v druhé blesky. Sedí na trůnu. U jeho nohou dli zamilovaný jeho pták, orel, který v zobanu také blesky drží. — Nejdůstojnější podobizna Zeova byla v Olympii od Pheidia zhotovená ze zlata a slonové kosti. Trůn a plášť Zeovův třpytily se zlatem a jinými lesklými kovy, drahokamy a živými barvami. Sedící bůh byl asi 25 metrů vysoký a dotýkal se hlavou skoro až stropu chrámového. Hlava byla ozdobena olivovým věncem, jakožto symbolem odměny za vítězství při hrách olympických. Obličej byl ideál řecké mužné, velenosti božskou však zvýšené krásy. Svrchní část těla byla nahá, ostatní tělo přikrýval říasný plášť, který dosahoval až k nohoun sandály zlatými opatřeným, jež spočívaly na stoličce zlatými lvičky držené. V pravé ruce držel Zeus bohyni vítězství, jak se právě k němu sklání, chtíc hlavu jeho věncem ozdobiti; v levé držel žezlo. Za kreslem vyčínala ona tajemná panna se lvím tělem Sphinx. Socha tato byla tak dokonalá a důstojná, že básník jeden při pohledu na ni se tázal: „Vystoupil Pheidias do Olympu, aby Zeva viděl, anebo sestoupil Zeus s nebe, aby Pheidiovi tvář svou ukázal?“ Každý Řek považoval to za velikou nehodu svého života, nebylo-li mu dopřáno vzácnou sochu tuto vlastnímá očima zříti. Ze stromů posvátný mu byl dub a mezi ptáky orel.



Jupiter Ammon. Zeus Doděanský.

©2003-2006 NK ČR
 developed by Qbizm

Otázka: Kdy byl upálen Mistr Jan Hus

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Kdy byl upálen Mistr Jan Hus? [Přehled](#) [Vzhledování](#)

vyhledej ve vzdálených repozitářích [Hledat](#) [Nastavení](#) [Nápověda](#)

[Rozdělit](#)

NKCP

Odpověď/Odpovědi:
1415 **Kdy byl upálen Mistr Jan Hus?**

1- 10 z 36 nalezených odpovědí (0 min 05 sec) [1](#) [2](#) [3](#) [4](#) [Další>](#)

1 Literární noviny
Týdeník pro kulturně politické a umělecké otázky.

1 • kůra VÝZNAMNÁ DATA V MĚSÍCI ČERVENCI. 6. července 1415 upálen Mistr Jan Hus — 7. čer. 1862 se narodil L. A. Orwell, ruský fyziolog, akademik — 9. července 1947 se narodil František Krážík, český vynálezce — 10. července 1923 zemřel Mikuláš Aleš, národní malíř — 11. července 1932 vláda předmnichovské republiky „rozpuštíla ústředí Rudých odborů“ — 11. července státní svátek lidové republiky Mongolsko — 12. července 1904 se narodil Pablo Neruda, chilský básník — 14. července 1888 se narodil Bohumil Meškus, český básník a překladatel — 15. července 1922 založena Komunistická strana Japonska — 15. července 1937 zahájena plavba na průplavu Moskva—Volha — 10. července 1902 se narodil Stanislav Neumann.

[Zdroj](#) - 100%

2 Července Památný den ČSR — 1415 byl upálen v Kostnici Mistr Jan Hus — 1904 nar.

[Zdroj](#) - 96%

3 Července 1419 byl upálen v Kostnici Mistr Jan Hus — 1872 se narodil v Praze Otto Fater, český novinář, dramatik a divadelní organizátor — 1892 se narodil v Praze PhDr. Václav Vážný, český filolog a vědecký pracovník — 1948 zemřel v Berlíně A. V.

[Zdroj](#) - 96%

4 Července 1415 byl upálen v Kostnici Mistr Jan Hus — 1335 byl popraven Torna Moru, anglický statik-ekonom.

[Zdroj](#) - 96%

5 v Měsí Jan Hu-hpák.

[Zdroj](#) - 19%

2 Žďár
Neodvstěte politický týdeník pro okres Rokycanský, Zbirovský, Hořovický, Blatovický a Kratochvíl.

1 oslavu 489. výročí upálení Mistra Jana Husu s následujícím pořadem: 1. Zahájení slavnosti o 8.

[Zdroj](#) - 97%

2 — Národní divadlo v předvečer výročí upálení Mistra Jana Husa a v den 6. července uspořádá dvě představení.

[Zdroj](#) - 97%

3 Památku upálení Mistra Jana Husa uctily sdružené zdejší spolky v pondělí večer.

[Zdroj](#) - 97%

4 Památku upálení Mistra Jana Husi důstojně uctí spojené místní spolky večer dne 5. července

[Zdroj](#) - 97%

[Přejte se knihovny](#) [Hledání na WWW](#) [Copyright](#)

http://kramerius.nkp.cz - Kramerius - Dokument stránky periodika - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

Výběr titulu
Titul
Ročník
Výtisk
Stránka
23 / 204

PDF [Vytvořit](#)
rozzah strani (max. 20)
23 - 204

ta Kostnice. Jeho (a Husův) žalářník byl Walenrode, biskup z Rigy, který jej dal zavízt ve věži u hřbitova svatého Pavla „a poručil, aby byl přikován. Byl připoután ke sloupu, který byl příliš vysoký, aby na něm mohl sedět, takovým způsobem, že jeho spoutané ruce byly vytaženy nad jeho hlavu.“ V této mučivé poloze byl držen po dva dny a byl živěn pouze chlebem a vodou, až Petr z Mladoňovic (žák a životopisec Husův) zjistil jeho bídu a až se mu podařilo přesvědčit jeho strážce, aby mu ulehčili.“¹³¹⁾

V tomto vězení zůstal Jeronym skoro rok; nejdříve o něm slyšíme v jednom z Husových dopisů, psaných během měsíce června z vězení. Už 6. června 1415 vyjádřil Hus své obavy, že Jeronym bude vydán na smrt,¹³²⁾ a později téhož měsíce se mu zdálo o Jeronymovi, že je ve vězení a připomínal, že Jeronym kdysi řekl: „Jestliže půjdu na koncil, pak se myslím nevrátím.“¹³³⁾ Když bylo dokončeno Husovo vyslýchání, řekl císař Zikmund 8. června otčům koncilu: „A uděláte také konec s jeho tajnými žáky a stoupenci, neboť zakrátko odjedu, a zejména budete jednat o tom muži (opakoval slova „o tom muži“), který je zde ve vězení.“ A oni se ptali „o Jeronymovi?“ A on odpověděl: „Ano, Jeronymovi. Skoncujeme s ním v jednom dni; neboť jeho záležitost bude snazší, neboť Hus je učitel a tento Jeronym je jeho žák.“¹³⁴⁾ Hus měl stále Jeronyma na mysli v posledních dnech svého života. 10. června psal Hus „všemu českému lidu“ český dopis, v němž říkal: „O mistru Jeronymovi, tovariši mém milém neslyším ničeho, než že je v tuhém vězení a čeká na smrt, jako já, a to pro svoji víru, kterou statečně ukázal Čechům — a jsou to Češi, kteří jsou našimi nekrutějšími nepřáteli a vydali nás v ruce jiných nepřátel, aby nás vsadili do vězení. Žádám vás, prostě za něj Pána Boha.“¹³⁵⁾ Poslední Husova zmínka o Jeronymovi, obsažená v českém listu z 27. června, je prorocká: „Právím vám, že Bůh ví proč otálejí s tím, aby zabili mne a mého drahého bratra, Mistra Jeronyma, který, jak věřím, půjde na smrt svatě bez viny a bude si počínati a trpěti statečněji než já, bídný hříšník.“¹³⁶⁾

6. července 1415 byl Hus odsouzen a upálen před branami Kostnice. 11. července napsal Jan „Zelezný“, biskup litomyšlský a předák protivyklifovské strany český králi Václavovi a sděloval mu, že se připravuje soud s Jeronymem.¹³⁷⁾ Komise koncilu, jejímž předním členem byl kardinál Petr z Ailly, sepisovala proti němu články obžaloby. Ale podobně jako v případě Husově, snažil se koncil daleko spíše dosáhnout Jeronymova odvolání než jeho smrti; kající hříšník by byl koncilu daleko prospěšnější než mučedník, tím spíše, že Husova smrt již rozpoutala v Če-

¹³¹⁾ Poggio napsal o Jeronymově věznění: Toto je obdivuhodný příklad dobré paměti: po 340 dnů byl Jeronym v hloubi špinavé, tmavé věže, kde nejen že nic nemohl číst, ale kde ani nic ne-

©2003-2006 NK, ČR, developed by Qbizim

http://kramerius.nkp.cz - Kramerius - Dokument stránky periodika - Microsoft Internet Explorer

Soubor Upravy Zobrazit Oblíbené Nástroje Nápověda

Výběr titulu
Titul
Ročník
Výtisk
Stránka
PDF Vytvořit
rozsah strani (max. 20)

25%

LIZARDTECH

Strana 6. NÁROD. Číslo 13.

do řeči ani odmlouváním, jevenim chvály nebo hany. Na to šteno několikero ze 260 článkův bludných, ježto universita Oxfordská ze spisů Wiklefových byla vytáhla, i vyřčen nad nimi odsudek. Pak přišla řada na při Husovu. Papeživ auditor Berohold z Wildungem četl nejprv třicetero článků ze spisů jeho, potom žaloby svědectvími dotvrzené a celý běh práva proti němu vedeného. Hus chtěl hned z počátku od-

osobu ve svatě trojici, nedal si překaziti, aby neodmlouval „opět“ hlasitě a poněkud i tpoec. Opakoval také časté své slovo, že dověřiv e-králi a ochrannému jeho listu, přišel ke sboru dobrovolně k dokázání neviay své. Sigmundovy lice zřejmě zarděly se, když Hus toto mluvě oči své naň upřel. Potom četl biskup Konkordien-ský odsudek nejprv proti knihám a učení jeho, pak proti němu samému; ony odsouzeny k obni- on pak vyhlášen za zjevného kacíře, an kázal při učení bludné, pohoršlivé i odbojné, zavrel množství lidí, turil čest a moc apoštolské st- lice i cirkve a zarputil se naprosto ve zlosti své: pročž že měl býti szazen s kněžstvi, zba- ven přijatého svěcení a odevzdán rukám svě- ským, ana cirkve nic s ním dále činiti neměla.

Nafizeno jest arcibiskupovi Milánskému i šesti jiným biskupům, aby Husa co kněze od- svátli na místě. Postavivše tedy jej na ležení a davše naň všecka roucha mešní, do rukou pak jeho kalich, žádali opět od něho, aby odv- lal bludy své. I promel se opět s pláčem k ok- lotojícím, že boje se ubližiti svědomí svému, býti lhářem před bohem a dát: pohoršení po- sluchačům i učeníčkům svým, nemohl učení a spíš svých odpřisáhnouti ani odvolati. Biska- pové přišedivše tedy k degradaci jeho, odjali jemu nejprve kalich, potom roucha mešní jedno po druhém se hrozným zlořečením, ve případ- nosti takové předepsaným. Když mu posleze i kněžská plés zrušena býti měla, postavla mezi biskupy hádka o způsobu, kterak to činiti se mělo, on pak uživ příležitosti té, obrátil se ku králi slovy těmito: „aj, ještě tito biskupové ne- umějí ani v trýznění mém se rovnati!“ Po do- konání obřadu toho traplivého postavili mu py- namídní papírovou korunu asi loket zdělí na hlavu, na niž malováná byli tři čertové, sápnají duši hříšnou, s nápisem: „hic est haeresiarcha“ (stotoť je arcikacíř); biskupové pak promeli tato poslední nad ním slova: „cirkve nemajje již nic s tebou činiti, odevzdává tělo tvé moci světské, duši pak tvou ďáblu.“ Hus nezaměšev se při tom, osvědčoval ochotnost svou podnikati všelké muky za příkladem, jak pravil, Kristo- vům i apoštolů jeho. Na rozkaz od krále Sig- mundava dasy odložil tadyž falckrabě Ludvík Heidelberský znaky říšské a vzav vězň v moc svou, odevzdal jej konšelům Konstantčským a t- ní slova: „vezměte Jana Husa, což dle nálezu

Hus na hranici.
Z kancionálu města Litoměřic, „as z r. 1517.“

mlouvat i ospravedlňovat se, což mu ale již nedovoleno; když pak přece mluví se pokou-

©2003-2006 NK ČR
developed by Qbizm

http://www.dbase.cz/manuscriptorium/ft_skoly/mDisplay.asp?Id=H/NKCR_VII_A_18_2RSHARep=http://www.dbase.cz/manuscriptorium/ft_skolylang-CZ

Microsoft Outlook Web Access MS viewer - ver. 1.50.12

O dokumentu Bibliografický popis Prohlázení Záložky Další dokumenty Nápověda

Low 100% 1. Český překlad

Manuscriptorium.com
VII.A.18 - 11r
(241 / 81)
Play text

Zde Jan Hus, osobiví mistr, byv poražen ve veřejné disputaci a odsouzen jako heretik, byl cthodnými otcí zbaven hodnosti a odevzdán světské moci s připojenou prosbou, aby nebyl vydán na smrt, nýbrž aby byl spíše držen v doživotním žaláři. Římský král ho předal bavorskému vévodovi Ludvíkovi, aby ho jako zjevného kacíře potrestal. Ten pravil přistoupivšímu rychtáři říšského města Kostnice, který tam byl přítomen: "Tohoto osobivého mistra Jana Husa, který má být mým rozhodnutím odevzdán smrti, jménem nás obou předej ohni k upálení." Poté ho rychtář města Kostnice vydal do rukou břičů s dodatkem, aby ho upálili bez svrchního oděvu a obuvi. Na hlavu mu byla dána papírová mitra, na niž byli namaalováni dva zli duchové, které my zverne ďáblu. Mezi nimi obrat list

Vícejazyčné vyhledávání: Otázka v angličtině, odpověď v polštině

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST What is the anthem of Erupean Union? [Pokročilé vyhledávání](#)
 vyhledej ve vzdálených repozitářích [Nastavení](#)
 [Nápověda](#)

NKP
Odpověď/Odpovědi:
Ody do radości
 pro **What is the anthem of Erupean Union?**

1- 10 z 68 nalezených odpovědí (0 min 05 sec) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Další>](#)

1 TRAKTAT USTANAWIAJĄCY KONSTYTUCJĘ DLA EUROPY

- 1 Hymn Unii** pochází z „Ody do radości“ z IX Symfonii Ludwiga van Beethovena.
[Zdroj](#) - - 100%
- Państwa Członkowskie uławiają wypełnianie zadań **Unii** i powstrzymują się od podejmowania wszelkich środków, które mogłyby zagrazać urzeczywistnieniu celów **Unii**.
[Zdroj](#) - - 1%
- Konstytucja i prawo przyjęte przez instytucje **Unii** w wykonywaniu przyznaných jej kompetencji mają pierwszeństwo przed prawem Państw Członkowskich.
[Zdroj](#) - - 1%
- Flaga **Unii** przedstawia krąg dwunastu złotych gwiazd na niebieskim tle.
[Zdroj](#) - - 1%
- Dewiza **Unii** brzmi: „Zjednoczona w różnorodności”.
[Zdroj](#) - - 1%
- Walutą **Unii** jest euro.
[Zdroj](#) - - 1%
- Dzień Europy obchodzony jest w całej **Unii** 9 maja.

Ptejte se knihovny | Hledání na WWW | Copyright

Otázka v češtině, odpověď v polštině

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Co je Europ direct? [Pokročilé vyhledávání](#)
 vyhledej ve vzdálených repozitářích [Nastavení](#)
 [Nápověda](#)

NKP
Odpověď/Odpovědi:
serwis, który pomoze Państwu znalezc odpowiedzi na pytania dotyczące Unii Europejskiej
 pro **Co je Europ direct?**

1- 2 z 2 nalezených odpovědí (0 min 02 sec)

1 TRAKTAT USTANAWIAJĄCY KONSTYTUCJĘ DLA EUROPY

- 1 KONSTYTUCJĘ DLA EUROPY**
[Zdroj](#) - - 100%
- Europe **Direct** to serwis, który pomoze Państwu znalezc odpowiedzi na pytania dotyczące Unii Europejskiej.
[Zdroj](#) - - 9%

1- 2 z 2 nalezených odpovědí (0 min 02 sec)

Ptejte se knihovny | Hledání na WWW | Copyright

Otázka v polštině, odpověď v češtině

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Gdzie jest elektrownia jądrowa Bohunice? [Pokročíle vyhledávání](#)
 vyhledej ve vzdálených repozitářích [Hledat](#) [Rozdělit](#)
[Nastavení](#) [Nápověda](#)

NKP

Žádná odpověď pro **Gdzie jest elektrownia jądrowa Bohunice?**

1- 10 z 38 nalezených odpovědí (0 min 02 sec) [1](#) [2](#) [3](#) [4](#) [Další>](#)

1 SMLOUVA O Ústavě pro Evropu

1. USTANOVENÍ O BLOCÍCH 1 A 2 JADERNÉ ELEKTRÁRNY **BOHUNICE** V1 NA SLOVENSKU
[Zdroj](#) - 100%
- Slovensko se zavázalo uzavřít blok 1 Jaderné elektrárny **Bohunice** V1 nejpozději do 31. prosince 2006 a blok 2 této elektrárny nejpozději do 31. prosince 2008 a následně je vyřadit z provozu.
[Zdroj](#) - 100%
1. V období 2004-2006 bude Unie poskytovat Slovensku finanční pomoc na podporu jeho úsilí při vyřazování bloků 1 a 2 Jaderné elektrárny **Bohunice** V1 z provozu a na řešení důsledků jejich uzavření a vyřazení z provozu (dále jen "pomoc").
[Zdroj](#) - 100%
37. Prohlášení o blocích 1 a 2 Jaderné elektrárny **Bohunice** V1 na Slovensku
[Zdroj](#) - 100%
1. V oblastech, v nichž Španělské království zpřístupňuje své informace Evropskému společenství pro **atomovou** energii, udělují příslušné subjekty na žádost a za obchodních podmínek licence členským státům, osobám a podnikům Společenství, pokud jsou držiteli výhradních práv na patenty přihlášené v členských státech Společenství a pokud nemají žádnou povinnost vůči třetím osobám udělovat nebo nabízet výlučné licence nebo částečné výlučné licence na práva vyplývající z těchto patentů.
[Zdroj](#) - 14%
1. Od 1. ledna 1973 jsou Irsku zpřístupněny poznatky sdělované členskými státy, osobám a podnikům v souladu s článkem 13 Smlouvy o založení Evropského společenství pro **atomovou** energii, které je na

[Ptejte se knihovny](#) [Hledání na WWW](#) [Copyright](#)

Otázka ve francouzštině, odpověď v portugalštině

Národní knihovna České republiky - Microsoft Internet Explorer

Soubor Úpravy Zobrazit Oblíbené Nástroje Nápověda

NK Národní knihovna České republiky
National Library of the Czech Republic

Pro novináře

M-CAST Quel est le drapeau de l'Union Européenne ? [Pokročíle vyhledávání](#)
 vyhledej ve vzdálených repozitářích [Hledat](#) [Rozdělit](#)
[Nastavení](#) [Nápověda](#)

NKP

Odpověď/Odpovědi:
círculo de doze estrelas douradas sobre fundo azul 
 pro **Quel est le drapeau de l'Union Européenne ?**

1- 10 z 71 nalezených odpovědí (0 min 02 sec) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Další>](#)

1 TRATADO QUE ESTABELECE UMA CONSTITUIÇÃO PARA A EUROPA

1. A **bandeira da União** é constituída por um círculo de doze estrelas douradas sobre fundo azul.
[Zdroj](#) - 100%
- A moeda da **União** é o euro.
[Zdroj](#) - 98%
- A decisão **européia** é um acto não legislativo obrigatório em todos os seus elementos.
[Zdroj](#) - 98%
- O lema da **União** é: «Unida na diversidade».
[Zdroj](#) - 97%
- A Conferência declara que o Conselho Europeu deverá começar a preparar a decisão **européia** que estabelece as medidas de aplicação da decisão **européia** do Conselho Europeu relativa ao exercício da Presidência do Conselho imediatamente após a **assinatura** do Tratado que estabelece uma Constituição para a Europa, e deverá dar-lhe a sua aprovação política no prazo de seis meses.
[Zdroj](#) - 83%
- As **notas** de banco emitidas pelo Banco Central Europeu e pelos bancos centrais nacionais são as únicas com curso legal na **União**.
[Zdroj](#) - 83%
22. O PROTOCOLO DE BRUXELAS, QUE É O PROTOCOLO DA CONSTITUIÇÃO, RECEPTANTE À APROVAÇÃO

[Ptejte se knihovny](#) [Hledání na WWW](#) [Copyright](#)

Příloha 2: Hodnocení projektu



Final Review Report

Project No.: EDC 22249

Project website: <http://www.m-cast.infovide.pl>

Project acronym: M-CAST

Names of Reviewers: Jörg Schütz
Galia Angelova

Date of meeting: 08-March-2007

INFSO	
Adonis n°	A 408 443
Date	09 MARS 2007
Action	FM
Contrib.	
Info	
Deadline	
File	

A. Work Done

The reviewed project results perfectly match with the Technical Annex of the proposed objectives, and are delivered with very good quality and within the scope of the deadlines. The consortium took very seriously the comments made at the mid-term review; especially the ones concerning the integration of users and the quality of deliverables. The system demonstrator is built with the active involvement of the users-libraries who also invested much effort in an in-depth user requirements study and the testing of the system.

The automated multilingual query-answering features, covering 6 languages, are very well developed. The demonstrator is focused on Polish and Czech library texts. The other languages were added to the consortium linguistic resources from existing previous developments; and thus the search facilities could be farther elaborated.

The project results are in line with the objectives of the eContent programme in the multilinguality area and they prove the potential of a multilingual question-answering system for digital libraries.

B. Product/Service Results:

The delivered portals for searching with free natural language questions in two digital libraries is based on a sound and efficient technical basis, that can/must be customised to cater for the multilingual and diverse requirements of the partners and users. The overall technical solutions are built on a solid methodological ground and current state-of-the-art NLP services, provided by European leaders in the respective fields.

The Consortium manages to demonstrate acceptable speed in processing the really large amounts of linguistic resources. The integrated framework of the M-CAST engine harmonises a variety of library document collections, their metadata in real settings and corresponding linguistic resources supporting the M-CAST language technologies.

C. Project Management

The project management for this project was very good, with an obviously coherent group of partners who have managed to deliver high-quality project results.

It is also clear that much time has been devoted to the actual involvement of users during the second year of the project as well as to dissemination activities in Poland and Czech Republic. Being pioneers in the field in these two new member states, the partners have also organised several workshops to attract the public attention to their results and to establish contacts with the relevant European initiatives.

There was also concern to ensure a solid legal basis for the project with respect to the IPR issues and publishing library materials.

D. View on Project Status

The project can be considered as successfully completed to the highest possible standard. All agreed deliverables have been presented in time and their quality meets the standards and requirements of the eContent programme.

The portal is operational and its functionality matches the expected completeness and quality, according to the Technical Annex. The initial plans for general multilingual search were justified to multilingual question-answering as the latter fits best to the services expected by the library users.

E. Exploitation / Implementation potential

The potential for successful commercialisation of the project results is evident, especially for the software providers from Portugal; France and Italy who are key actors at the European LT market. The lessons learnt in M-CAST will strengthen their position at the market of search engines and the question-answering in particular.

However, plans for potential commercialisation in Poland and Czech Republic are demonstrated as well. The libraries need to search for further funding to keep the M-CAST prototype alive and to turn it to a modern electronic service in a longer term.

F. Recommendations for future work

For future commercial applications, the Consortium may consider extending the linguistic resources and adding new languages.

G. Overall recommendation:

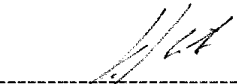
See above.

Successful completion

Modify

Rejected


Reviewer: Jörg Schütz



Signature:

Date: 08-March-2007

Galia Angelova



Signature:

Date: 08-March-2007