

Obsah

Předmluva	7
Systém M-CAST (Multilingual Content Aggregation System – Vícejazyčný systém agregace informací) M. Balíková	9
Úvod	9
Charakteristika systému M-CAST	11
Konsorcium	11
Systém TRUST	12
Základní schéma systému M-CAST	12
Architektura systému	13
Lingvistické moduly	14
Lingvistický procesor	15
Webové služby systému M-CAST	16
Integrační vrstva Portálu M-CAST	17
Prezentační vrstva systému M-CAST	18
Databáze indexovaných dokumentů	19
Databáze indexovaných dokumentů zpřístupněné pro testování systému M-CAST	19
Kvalita OCR textů	21
Portál M-CAST/M-CAP	21
Typologie dotazů v systému M-CAST	22
Práce s časově omezenými dotazy	22
Proces zodpovězení časově omezené otázky	23
Indexace dokumentů v systému M-CAST	23
Indexace typů dotazů	24
Indexy systému M-CAST	24
Extrakce odpovědi	24
Vyhledávání v systému M-CAST	25
Jednoduché vyhledávání	25
Pokročilé vyhledávání	25
Metoda dotazování v přirozeném jazyce aplikovaná v systému M-CAST	26
Délka dotazů	26
Odpovědi v systému M-CAST	27
Přímá odpověď	27
Úryvek odpovědi	27
Vizualizace zdrojového dokumentu	27
Ukázky vyhledávání v systému M-CAST	28
Vícejazyčné vyhledávání v systému M-CAST	31
Testování v systému M-CAST	35
Hodnocení systému M-CAST	36
Odezva systému	36
Správnost odpovědí	36
Přímá odpověď	36
Úryvek odpovědi	36
Optimální počet odpovědí	36
Výsledná statistika	36
M-CAST v hybridních knihovnách	37
Systém M-CAST a obsahové údaje (projekt TOC – Table Of Content)	37
M-CAST v hybridních veřejných knihovnách	40
Závěr	41
Literatura	42

Zkušenosti získané při tvorbě českého lingvistického modulu systému M-CAST	
<i>P. Strossa</i>	43
Vzorce pro rozpoznávání typů dotazů a potenciálních odpovědí	43
Problém volného slovosledu v češtině	45
Morfologická analýza	45
Morfologické vzory podstatných jmen	45
Morfologické vzory přídavných jmen	49
Morfologické vzory sloves	51
Souhrn vybraných parametrů slovníkových distribucí českých morfologických vzorů ..	52
Otázka „objektivní existence“ a „správnosti“ některých tvarů v české morfologii	53
Problém gramatické homonymie v češtině	54
Literatura	56
M-CAST – postřehy z testování <i>D. Vřešťálová</i>	57
M-CAST (Multilingual Content Aggregation System based on TRUST Search Engine)	
<i>Compiled by M. Balíková</i>	61
System functionality	62
Overall system architecture	63
System modules	63
Request analysis	64
Document indexing	65
Retrieving documents	65
Refining results	66
Displaying results	66
M-CAST system at Czech National Library (NKP)	66
M-CAST system – instances, performance	66
Simple search	67
Advanced search	67
Search options of the M-CAST system: Full text search/Metadata search	67
M-CAST document collection at NKP	67
M-CAST Question Answering method	68
The architecture of the M-CAST QA system	68
The five main phases of M-CAST QA system	69
Questions	69
Types of questions	69
Question length	70
The role of context in M-CAST QA system	70
Answer set generated by M-CAST system	70
Exact Answer	70
Answer-string / snippet	70
Entire document	70
Topic familiarity in QA systems	71
Evaluation process	71
Assessment categories	71
Categories of answer	71
Categories of answer-string/snippet	71
Optimal number of answers	72
Conduct of the testing process	72
Statistics	72
Conclusions	72
Příloha – ukázka otázek	73