

ZDOKONALENÍ VIRTUÁLNÍHO BADATELSKÉHO PROSTŘEDÍ MANUSCRIPTORIA

- OPTIMALIZACE METODIKY VYUŽÍVÁNÍ GRAFÉMŮ PŘI VYHLEDÁVÁNÍ V MANUSCRIPTORIU

Zpráva

verze 1.0

AiP Beroun s.r.o., autoři: Ing. Stanislav Psohlavec, Mgr. Olga Číperová a kol.

Obsah

1	Úvod o dokumentu	3
1.1	Účel	3
1.2	Předpokládaný čtenář	3
1.3	Použité informace	3
1.4	Reference	3
2	Úvod	3
3	Grafémické skupiny	4
4	Testy účinnosti změny metodiky	5
4.1	Testování správnosti metodiky	5
4.2	referenční test nedělené grafémické skupiny	6
5	Test účinnosti rozdělení do gr. skupin	6
5.1	Výsledky prvního setu testů	8
5.2	Závěr prvního testování	8
5.3	Východiska pro druhé testování	9
5.4	Modifikace metody testování	9
6	Modifikace testu účinnosti	10
7	Vyhodnocení přínosu dělení grafémů do skupin	11
8	Závěr	12
	Příloha č.1	14
	Příloha č.2	15
	Příloha č.3	16
	Příloha č.4	17

1 Úvod o dokumentu

Firma AiP Beroun uzavřela s Národní knihovnou České republiky dne 10.11.2011 Smlouvu o spolupráci ve výzkumu a vývoji: Zdokonalení virtuálního badatelského prostředí Manuscriptoria - Optimalizace metodiky využívání Grafémů při vyhledávání v Manuscriptoriu.

1.1 Účel

Tento dokument tvoří zprávu k optimalizování metod využívání grafémů při pokládání dotazu do katalogu Manuscriptoria. Popisuje a zdůvodňuje užití metody a ukazuje výsledné zrychlení a zkvalitnění vyhledávání.

1.2 Předpokládaný čtenář

Tento dokument je určen především pro Zadavatele (NK ČR) a jako podklad pro realizaci opatření zkvalitňujících možnosti rutinního vyhledávání v Manuscriptoriu. Dále je tento dokument určen všem, kteří se podílejí na rozvoji projektu Manuscriptorium jako uživatelé.

1.3 Použité informace

Sbírka grafémů oddělení hudebních a historických fondů NK ČR.

1.4 Reference

V dokumentu se odkazujeme na následující literaturu:

- [1] Řešení problematiky grafémů při vyhledávání v Manuscriptoriu - Zpráva k Dodatku č. 8, AiP Beroun 2010.

2 Úvod

V rámci projektu „Řešení problematiky grafémů“ při vyhledávání v Manuscriptoriu [1] v roce 2010 byla pro uživatele v katalogu Manuscriptoria zpřístupněna možnost při zadávání dotazu v pokročilém formuláři si zvolit rozšířené hledání pomocí grafémických variant. Tuto službu bohužel výrazně omezoval fakt, že prohledávání všech dostupných variant bylo časově náročné. Při využití této volby je každý potenciální grafem ve slově postupně nahrazen všemi grafémickými variantami ze stejné skupiny. Ke každé takovéto záměně se rekurzivně zkouší záměny na dalších pozicích slova. Vzniklých kombinací je neuvěřitelné množství.

V praxi tedy často docházelo k tomu, že výsledný čas vyhledávání překročil povolených 30 sekund i u poměrně krátkých termínů (od 7 znaků výše), což bylo pro uživatele značně nepříjemné.

Na základě těchto zjištění jsme se věnovali v rámci letošního výzkumného úkolu optimalizaci metodiky využívání grafémů. Vycházeli jsme z hypotézy, že není nutné generovat veškeré grafémické variace vyhledávaných slov najednou, protože jednotlivé texty i v nich obsažené citace jsou v podstatě jazykově i historicky homogenní.

Při zkoumání jsme pracovali s původním souborem grafémů, který byl sestaven dle podkladů dodaných NK ČR v roce 2010 a prezentován ve zprávě Řešení problematiky grafémů při vyhledávání v Manuscriptoriu [1].

K ověření dále předkládané hypotézy jsme modifikovali modelové prostředí Manuscriptoria (dále jen „Model“), ve kterém byly vytvořeny prostředky pro řešení této práce a implementovány nástroje pro automatizované provádění testů.

3 Grafémické skupiny

Původní soubor grafémů byl pro testovací účely rozdělen na několik částí primárně podle jazykového hlediska. Oproti zadání preferujícímu jazykové členění se ukázalo jako efektivnější členění podle celých jazykových skupin. Do prvního bloku byly vyčleněny grafémy vyskytující se převážně v latinských textech, další tři bloky obsahovaly grafémy z evropských národních jazyků – grafémické varianty pro jazyky anglosaské, románské a slovanské. V každé z uvedených skupin byly zařazeny znaky a kombinace typické pro daný jazykový okruh. U slovanských jazyků to byly např. znaky s diakritikou (háčky, čárky) či specifické spřežkové kombinace, které v některých slovanských jazycích diakritiku nahrazují. Na základě této rozvahy byly vytvořeny čtyři **grafémické skupiny**:

- A – latina
- B - anglosaské jazyky
- C - románské jazyky
- D - slovanské jazyky

Grafémické skupiny národních jazyků (BCD) ve svém sjednocení významně rozšiřují množství variant, které nejsou obsaženy ve skupině A (latina).

Přehled těchto variant

Těchto odlišností jsme využili existence těchto skupin se zřetelně odlišnými vlastnostmi (Slovník T2) jsme využili tuto charakteristiku a směřovali první blok dotazů nejprve na skupinu A, a potom na sjednocení skupin BCD.

Druhý blok dotazů potom směřoval na sjednocené skupiny AB (latina a anglosaské jazyky) a CD (románské a slovanské jazyky). U tohoto rozdělení jsme vycházeli z předpokladu, že odlišnost počtu a charakteru používaných diakritických znamének a grafémických variant v prvním a druhém bloku bude mít vliv na dobu vyhledávání. Přehled dělení do variant je zřejmý z Přílohy 1, 2, 3, 4.

V první řadě jsme se tedy zaměřili na to, zda a v jaké míře dojde u každé z variant ke zkrácení času pro vyhledávání.

4 Test účinnosti změny metodiky

Testování konkrétních dotazů by zavedlo do procesu významnou subjektivitu, která by silně zatížila výsledky testování. Nelze prakticky vytvořit homogenních testovací soubor smysluplných dotazů, který by byl statisticky relevantní. Smysluplnost dotazu jsme proto nepostavili jako podmínku a předpokládáme, že není pro vyhodnocení rozhodující. Vytvořili jsme nástroje pro náhodné generování dotazu. Při dostatečně velkém množství pokusů o vyhledání se střední doba vyhledání ustálí a je pak obrazem náročnosti generování grafémických variant.

Testování je prováděno na celé databázi Manuscriptoria, v prostředí Modelu.

Základní soubor grafémových skupin je použit dle stavu z počátku r. 2011.

V Modelu je ponecháno přerušení jednotlivého vyhledávání po 25 sekundách.

Dotaz je generován náhodně z celého znakového prostoru EC-kódu a je formálně aplikován na položku 'Titul'. Volba položky nemá na sledované výsledky vliv.

V testu správnosti metodiky bylo určeno, kolik náhodných dotazů je nutno učinit, aby se výsledky přestaly významně měnit a staly se tak signifikantními.

4.1 Testování správnosti metodiky

Hypotéza:

Budou-li průměrné sledované charakteristiky určitého kvanta náhodně generovaných dotazů vykazovat jen malý rozptyl a bude-li se jejich průměr blížit následnému kumulativnímu testu, postačí v dalším rozhodování výsledky v lepším případě kumulativního testu, u časově náročných testů nebude závažnou chybou použít i výsledky kratšího testu na základě menších kvant dotazů.

Test homogenity dotazů při nerozdělených skupinách

5 x 10.000 náhodných dotazů délky 7 znaků

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,353 sec	119 321	96
2	0,366 sec	121 463	98
3	0,345 sec	112 975	114
4	0,311 sec	103 205	91
5	0,310 sec	103 758	112
	průměr 0,337 sec	průměr 112 110	celkem 511

Kumulativní test:

1 x 50.000 náhodných dotazů délky 7 znaků

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
	0,342 sec	115 406	508

Závěr:

Výsledky testu jsou homogenní a poskytují signifikantní informaci o vlastnostech grafémových skupin pro další rozhodování.

Minimální skupina dotazů je stanovena 10.000, optimální 50.000 dotazů.

4.2 referenční test nedělené grafémické skupiny

Výsledky předchozího testu budou použity jako referenční pro závěrečné porovnání účinnosti upravené metody aplikace rozdělení grafémů do logických skupin při vyhledávání v Manuscriptoriu.

5 Test účinnosti rozdělení do gr. skupin

Rozdělení grafémických skupin typu A + BCD

Test: 3 x 20.000 náhodných dotazů délky **6 znaků**
skupina **A**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,003	585	0
2	0,002	426	0
3	0,002	463	0

Test je nenáročný. V dalším bude testováno slovo délky 7 znaků.

Test: 3 x 20.000 náhodných dotazů délky **7 znaků**
skupina **A**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,007	2 017	1
2	0,004	1 209	0
3	0,004	1 157	0

skupina **BCD**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,185	62 318	107
2	0,207	69 951	136
3	0,194	64 412	111

komplet **A + BCD**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,206	66 149	129
2	0,199	63 673	141
3	0,191	63 220	138

Rozdělení grafémických skupin typu AB + CD

Test: 3 x 20.000 náhodných dotazů délky **7 znaků**

skupina **AB**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,116	39 102	68
2	0,121	40 523	51
3	0,117	38 775	79

skupina **CD**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,128	40 142	73
2	0,135	41 632	48
3	0,127	42 849	93

Komplet **AB + CD**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,136	41 561	103
2	0,154	48 276	65 / 4
3	0,149	47 027	78 / 12

Rozdělení grafémických skupin typu A + BC + D

Test: 3 x 20.000 náhodných dotazů délky **7 znaků**

skupina **A**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,006	1 774	1
2	0,005	1 423	1
3	0,004	1 024	0

skupina **BC**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,206	67 224	112
2	0,164	54 029	96
3	0,178	59 306	116

skupina **D**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,135	42 690	70
2	0,168	55 245	91
3	0,139	46 105	74

komplet **A + BC + D**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
1	0,261	86 816	147
2	0,342	108 703	183
3	0,229	67 751	120

5.1 Výsledky prvního setu testů

Pro porovnání jsou použity průměrné hodnoty z předchozích testů.

Typ	střední doba hledávání	střední počet g. variant	Hledání > 5 sec
nedělený	0,342 sec	115 406	508
A + BCD	0,199 sec	64 347	408
AB + CD	0,146 sec	45 621	246
A+BC+D	0,277 sec	87 757	450

5.2 Závěr prvního testování

Výsledky testování jsou optimistické a zdá se, že bylo dosaženo příznivého téměř 50% zrychlení vyhledávání. Zdánlivě je důvod zkoumání uspokojeně ukončit.

Praktické zkušenosti však vedou k opatrnosti.

Nápadný je velmi optimistický výsledek pro testování skupinou A. To pravděpodobně neodpovídá praktické realitě. V praxi se vyskytuje významně častěji příliš dlouhé vyhledávání.

Zvolená metodika potlačuje možný vliv smysluplné tvorby dotazů, kterou bohužel nelze prakticky aplikovat, protože nemůže v reálných časech vést ke statisticky významným závěrům. Na druhou stranu zvolená metodika postihuje rovnoměrně celý prostor možných dotazů, a v kombinaci s vysokou rychlostí vyhledávání poskytuje data pro statisticky významné hodnocení závěry. Optimistické vyhodnocení však vedlo k zamyšlení, zda není zvolená metodika příliš zjednodušená. Dotazy sice pokrývají celý prostor možných dotazů, nicméně právě v tomto celém prostoru je jen malá podmnožina dotazů, majících charakter odpovídající vlastnostem skupin A, B, C, D. Je tedy oprávněná pochybnost, zda výsledky významně neovlivňuje fakt, že dotazy nemají charakter dotazu alespoň zhruba odpovídající vytipovaným skupinám A, B, C, D. Zavedení této vlastnosti dotazů by mohlo zlepšit průkaznost testů.

Samozřejmě by bylo možno vytvořit algoritmus lépe simulující „jazykové varianty“ národních dotazů. Tento směr výzkumu by však nepochybně přesáhl prostor vyčleněný pro řešení tohoto úkolu.

Pro limity dané časem a vyčleněnými prostředky byla vyzkoušena jako druhá varianta testování zaměřená na to, že dotaz bude generován jen ze znaků, které se vyskytují v základu podstatné většiny grafémických skupin.

5.3 Východiska pro druhé testování

To vedlo k úvaze, zda existuje množina znaků, která je pro tyto čtyři skupiny společná. Tato množina existuje. Je to právě množina znaků A.. Z bez diakritiky. Pro ověření nebo spíše interní oponenturu výsledků prvního testování tedy byly testy na Modelu modifikovány.

5.4 Modifikace metody testování

Pro generování dotazů nebyl při druhém testování použit celý rozsah EC kódu (viz. Příloha 1) ale jen znaky A.. Z bez diakritiky. Tyto znaky se vyskytují prakticky ve všech grafémických ekvivalencích.

Dále bylo zavedeno sledování ukončení vyhodnocování variant grafémů, jestliže by hledání překročilo dobu 30 sekund.

Délka generovaného dotazu byla ponechána na sedmi znacích. Testovací časy významně narostly, a proto byly provedeny jen reprezentativní testy a testování bylo redukováno pro velkou časovou náročnost a především proto, že druhé testování zřetelně potvrdilo výsledky prvního testování,

Druhé testování přineslo potvrzení náročnosti hledání grafémických variant ve zjednodušené formě, která se více blíží reálnému vyhledávání.

6 Modifikace testu účinnosti

Dotazy generovány ze znaků
délka dotazu:
sledováno překročení

A... Z,
7 znaků
>5sec / >25sec

znaků originál kumulovaná skupina

střední doba hledávání	střední počet g. variant	Hledání > 5 sec / >25sec
19,376	6 402 501	355 / 111

skupina **A**

střední doba hledávání	střední počet g. variant	Hledání > 5 sec / > 25sec
1,770	386 356	31 / 12

skupina **A + BCD**

střední doba hledávání	střední počet g. variant	Hledání > 5 sec / >25sec
7,495	2 537 643	416 / 95
8,089	2 756 021	423 / 105

skupina **AB + CD**

střední doba hledávání	střední počet g. variant	Hledání > 5 sec / >25sec
8,402	2 746 440	440 / 102
8,453	2 854 935	393 / 88

skupina **A + BC + D**

	střední doba hledávání	střední počet g. variant	Hledání > 5 sec / >25sec
1	7,823	2 615 482	338 / 87
2	7,426	2 553 460	359 / 79

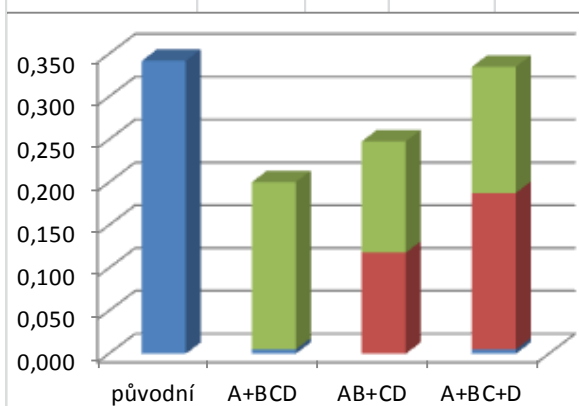
7 Vyhodnocení přínosu dělení grafémů do skupin

Kumulativní test 1 x 50.000 náhodných dotazů délky 7 znaků

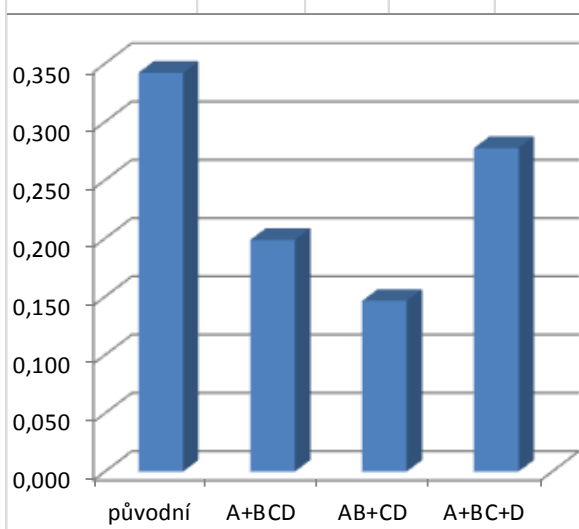
stř.doba	stř. počet	Hledání
hledávání	g.variant	delší
0,342 sec	115 406	508

Složení časů dle měření skupin (EC kód)

	původní	A+BCD	AB+CD	A+BC+D	
1.část	0,342	0,005		0,005	sec
2.část			0,118	0,183	sec
3.část		0,195	0,130	0,147	sec
součet	0,342	0,200	0,248	0,335	sec



dle měření (EC) 0,342 0,199 0,146 0,277



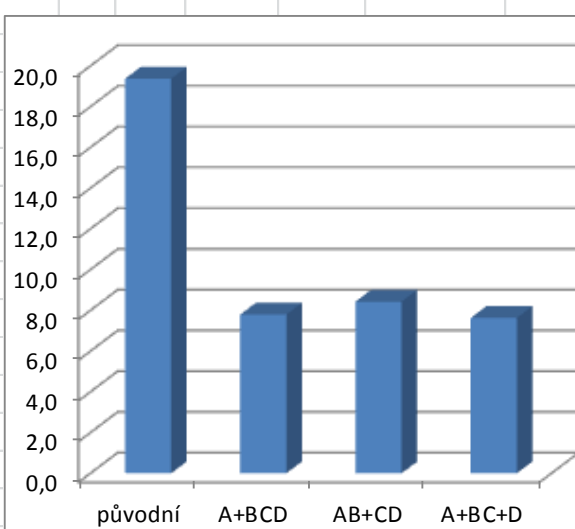
měření s plným EC kódem

A		BCD	A+BCD
0,007		0,185	0,206
0,004		0,207	0,199
0,004		0,194	0,191
AB		CD	AB+CD
0,116		0,128	0,136
0,121		0,135	0,154
0,117		0,127	0,149
A	BC	D	A+BC+D
0,006	0,206	0,135	0,261
0,005	0,164	0,168	0,342
0,004	0,178	0,139	0,229

měření s kódem A... Z

původní	19,367
A+BCD	7,495
AB+CD	8,089
A+BC+D	8,402
A+BC+D	8,453
A+BC+D	7,823
A+BC+D	7,426

dle (A... Z) 19,367 7,792 8,428 7,625



8 Závěr

Optimalizovaná metodika s využitím rozdělení grafémických variant do více skupin přinesla na náhodně generovaných dotazech zkrácení potřebného času na zhruba 50%. Nejprve bude prakticky prověřena verze A+ BC+D.

Metoda byla na základě testů v modelovém prostředí implementována zatím pouze do testovací verze Manuscriptoria. Vzhledem k tomu, že hypotéza předpokládající úsporu času při prověřování grafémických variant v několika po sobě automaticky následujících krocích byla otestována s pozitivním výsledkem, nebylo nutné upravovat formulář vyhledávání v Manuscriptoriu.

Inovovaná metoda vyhledávání grafémických variant je tedy pro uživatele k dispozici beze změny prostředí formuláře pro pokročilé vyhledávání.

Nezávisle na řešení tohoto úkolu budou v testovací verzi MnS vyzkoušeny reálné dotazy, jejichž sady jsou připravovány v době dokončování tohoto úkolu.

Do budoucna by bylo z uživatelského hlediska zajímavé zhodnotit význam podrobnějšího rozdělení grafémických skupin např. podle jazykových a časových rovin a jejich samostatné využití. Pokud bychom vycházeli z předpokladu, že uživatel přijde do katalogu Manuscriptoria s dotazem z určitého jazykového okruhu, pravděpodobně ho nebudou zajímat výsledky, které by ho odkázaly do odlišné jazykové skupiny. Nevýhodou by bylo zesložnění pokročilého formuláře a předpoklad vyšší kvalifikace uživatele. Otázkou je také akceptovatelnost nalezeného balastu a jeho akceptovatelnost. Příkladně, bude-li uživatel pokládat dotaz na záznamy obsahující české slovo „večer“ ve všech grafémických variantách, zřejmě nebude potřebovat záznam německého dokumentu obsahující termín „becher“. Z grafémického hlediska je vyhledání termínu „becher“ k dotazu na „večer“ oprávněné, z hlediska uživatelského dotazu je půvabně nesmyslné. Příklad pochází z testování původní verze grafémických variant. V tomto případě by byla přidanou hodnotou možnost výběru jazykové oblasti.

Podrobnější příklad, odvozený z nové metody tvoření grafémických variant je ilustrován v souboru „Dia_Grafemy_v01_UTF8“ na doprovodném CD-R disku.

Nabízí na dotaz „večer“ metodou A+BC+D nález slov (za grafémické skupiny):

B: bez nálezu,

BC: VEČER, WEČER, VEČER, WEČER, WECŽER, VECHER,

D: VEČER, WEČER, WECŽER, FÄCHER, VECHER, kde slovo FÄCHER může/nemusí být balast. Nález pochází z testování 53992 grafémických variant.

Zde je třeba zvážit přínos očekávání znalostí uživatele. Stejně jsou diskutabilní vlastnosti grafémických zdrojů. Základem je, co vlastně odborník (technický laik) očekává/potřebuje. Určitá míra balastu může být z badatelského hlediska dokonce přijatelnější, než když technika „přemýšlí“ a „předpokládá“ za člověka. Mírný balast dokonce naznačuje, že technicky nebylo nic významného „přehlédnuto“.

Příkladem využití lidské práce při ruční práci s vybranými virtuálními knihami je tvorba kolekcí knih, kde Manuscriptorium má již první praktické výsledky.

Při technické analýze problematiky jsme pravděpodobně narazili na určitou nekonceptnost při shromažďování podkladů pro aplikaci grafémů pro vyhledávání. Zdroje obsahují částečné i úplné duplicity. V praxi může být akceptovatelné velmi podobné grafémické skupiny spojit, což by možná přineslo zrychlení vyhledávání. Nárůst balastu by se musel prokázat zkouškami. Zdá se, že jeho nárůst by mohl být akceptovatelný, případně by k němu dokonce nemuselo dojít.

Cesta, která snadno napadne při pohledu na výše zvýrazněné výsledky – totiž zamezení vyhledání stejných slov – není schůdná, neboť by si vyžádala zásadní matematickou analýzu. Je třeba si všimnout, že nálezy ze skupin BC a D se liší (!), stejný zdroj se prohledává alternativně s různými nálezy.

Podobně by se mohla prověřit cesta úplného technického přepracování či optimalizace struktury zdrojů poskytnutých historiky. V této problematice je podstatné dosažení správného výsledku vyhledávání, nikoliv dodržení formálních pravidel platících v netechnické oblasti.

Existuje alternativní technické řešení problematiky grafémů z hlediska HW. Ukazuje se, že systém, na němž je Manuscriptorium postaveno, v sobě integruje mnohé velmi specializované služby. Tyto stále rostoucí služby vyžadují také rostoucí výkon výpočetní techniky. To lze samozřejmě řešit nárůstem výkonu HW, ale také distribucí úloh mezi více specializovaných zařízení. V projektu ENRICH byla úspěšně vyzkoušena vzdálená spolupráce standardními komunikačními protokoly při poskytování multilingválních služeb. Služby byly využity při překladu segmentů různojazyčného textu. Principiálně podobná problematika je řešena v projektu „Věda a výzkum 2011“, řešícím návaznost na evropský projekt CERL. Modulární řešení dovoluje připojování a odpojování služeb, aniž by byla narušena základní služba. To se potvrdilo při ukončení poskytování překladových služeb firmou SYSTRAN.

Aplikace grafémů při rešeršních službách vyžaduje mimořádný výpočetní výkon, který optimálně může jako službu poskytnout specializovaná stanice HW. Absence této externí systémové služby vede ve svých důsledcích k nižší účinnosti vyhledávání pro širší veřejnost. Správné řešení je služba, která je pro laiky vstřícná. Při prvním kontaktu s touto službou nevadí, když technika „přemýšlí“ a „předpokládá“ za člověka. Pro odborníky však musí být tyto nástroje průhledné, „dávkovatelné“ a případně i vypnutelné.

Rozdělení činností mezi více spolupracujících stanic je cestou dalšího technického vývoje, který může přinést podstatné posílení výkonu systému, zvýšení jeho stability a to při umírněných nákladech.

Řešení tohoto úkolu je typická interdisciplinární záležitost a zasloužila by si pozornost jazykovědnou, matematickou i počítačově systémovou.

Všechny tyto úvahy naznačují další možnosti pokračování vývoje.

Příloha č.1

Původní soubor grafémů

sestavený dle podkladů NK ČR v roce 2010,
dle zprávy Řešení problematiky grafémů při vyhledávání v Manuscriptoriu [1]

AE=OE=E
NN=MN=MPN
L=LL
R=RR
I=Y=J=II=IJ=JI
V=VV=W=U=UU=VU
IE=IHE=HIE
D=DD
Q=QU=CU
S=SS
TI=CI
CA=KA
T=TH=TT
CHA=CA
CHE=CE
SA=ZA
G=C=K
M=MM
IT=ID
ET=ED
SI=SCI
SE=SCE
SU=SUB
X=XS
P=PP
Á=A=AA
É=E=EE=Í
Í=I=Y=Ý=II=IJ=J=IE
Ý=EJ=EY=EI=AY=AI
Ó=O=UO=Ů=Ú=V
Ú=U=UU=AU=OU=V

Ě=IE=YE=E
C=CZ=CH
Č=C=CZ=ČZ=CŽ=CHS
= CS=CH
Ď=D=DI
F=FF=PH
G=J
H=CH
J=G=I=Y=IG=YG
K=C=CH
L=LL
M=MM
N=NN
Ň=N=NI=IN=YN
Q=KV=KW
Ř=RZ=RS=RSS=RZS
S=Š=Z=ZZ
Š=SS=SZ=S
T=TH=TT
Ť=TI=T
U=V
V=W
X=KS
Z=S=ZZ
Ž=Z=ZI=ZZ
SC=SCZ=SČ=ŠČ=SZC
Z= SŽ=CŽ=ŠŤ=SZTI
DL=L
AJ=EJ=AY=AI=EY=IE
ČS=CZS=C
O=VO=WO

A=Â=AA=AH
O=Ô=OO
U=Û=V
E=Ê
I=Î=Y
EI=AI=AY=EY=Î=Y=IJ
OU=AU=OW=AW=Û
EU=EW=IU=IUW
E=Ä=Æ
O=Ö=OE
U=Ü=IU
SS=ZZ=ß
PF=PH=FP=PP
TZ=Z=CZ=ZC=CZC=ZCZ
CH=KCH=HH
F=V
B=P
D=T
G=K=C
K=C=QU=CH=KCH
J=Y=I
CRIST=CHRIST=XPT
M=MB=MM
S=SCH
B=W
E=I
A=O
U=O
GE=G=K

Postup generování skupiny typu A + BCD (viz přiložený CD-R disk)

[illegible]

Postup generování skupiny typu AB + CD (viz přiložený CD-R disk)

13.12.2011

Postup generování skupiny typu A + BC + D (viz přiložený CD-R disk)

[illegible]