

<i>Prítomní:</i>	Bc. Róbert Horváth	<i>Vedúci:</i>	Ing. Dušan Zeleník
	Bc. Peter Jurčík	<i>Miesto:</i>	Softvérové štúdio
	Bc. Peter Macko	<i>Dátum a čas:</i>	11.4.2012 17:00
	Bc. Vladimír Ruman (zápis)	<i>Trvanie:</i>	120 minút
	Bc. Peter Sládeček (diskusia)		
	Bc. Maroš Ubreži		
	Bc. Matúš Vacula		

### Prehľad splnenia úloh z predchádzajúceho stretnutia

ID	Popis	Pridelená	Predpokladané ukončenie	stav	Skutočné ukončenie
3318	Odstránenie duplicity N-Gramov	Maroš Ubreži	11.4.2012	ukončená	10.4.2012
3319	Vytvorenie jednotlivých častí dokumentácie šprintov	Členovia tímu	11.4.2012	ukončená	11.4.2012
3320	Pospájanie dokumentácie do finálnej verzie	Peter Jurčík	11.4.2012	ukončená	11.4.2012
3495	Vytvorenie inštalačnej príručky	Peter Macko	11.4.2012	ukončená	11.4.2012
3496	Inštalačná príručka k mysql	Peter Sládeček	11.4.2012	ukončená	11.4.2012
3497	Vytvorenie návodu ako pracovať so službou	Matúš Vacula	11.4.2012	ukončená	11.4.2012
3497	Vytvorenie návodu ako používať knižnicu translatora	Róbert Horváth	11.4.2012	ukončená	11.4.2012
3499	Vytvorenie návodu ku tomcatu a nasadeniu služby	Peter Macko	11.4.2012	ukončená	11.4.2012
3500	Vytvorenie navodu ako nahodit Elasticsearch a veci k tomu	Vladimír Ruman	11.4.2012	ukončená	11.4.2012
3501	Vytvorenie	Peter Jurčík	11.4.2012	ukončená	11.4.2012

---

používateľskej  
a konfigurač-  
nej príručky  
k webovému  
rozhraniu

---

## Základné body stretnutia

Na stretnutí boli prerokované nasledujúce body:

- Odovzdanie dokumentácie a produktu
- Diskusia k IIT.SRC
- Prerokovanie funkcionálnych úloh z minulého stretnutia
- Určenie nových úloh

## Odovzdanie dokumentácie a produktu

Na začiatku stretnutia sme odovzdali nášmu vedúcemu dokumentáciu k doterajším šprintom za letný semester. Taktiež sme odovzdali inštalačné a používateľské príručky a tiež celý vytvorený produkt na CD.

Vedúci prijal dokumenty bez pripomienok.

## Diskusia k IIT.SRC

Vzhľadom na blížiacu sa súťaž IIT.SRC sme diskutovali o tvorbe posteru. Vedúci uviedol na základe vlastných skúseností niekoľko návrhov. Nakoniec sme určili dvoch členov zodpovedných za návrh plagátu a to Bc. Róberta Horvátha a Bc. Petra Macka.

Ďalej sme analyzovali potrebné vybavenie pre našu prezentáciu na tejto súťaži. Zhodli sme sa, že nám stačí obyčajný stôl.

## Prerokovanie funkcionálnych úloh z minulého stretnutia

### N-gramy

Na minulom stretnutí sme sa rozhodli dať ešte šancu riešeniu pomocou n-gramov. Túto úlohu mal na starosti Bc. Vladimír Ruman. Na stretnutí uviedol možné zlepšenia tohto riešenia a teda aj odstránenie hlavných nedostatkov, ktoré toto riešenie má. Riešenie spočíva v použití Term frequency na slová v n-gramoch, ešte pred ich poslaním do elasticsearchu. Takto sa veľmi nevhodné preklady slov pomocou slovníka (napríklad preklad slova „is“ ako „informačný systém“) dostatočne penalizujú a naopak môžu vďaka tomu dostať šancu pravdepodobnejšie preklady.

Po diskusii sme zhodnotili n-gramy ako možnú cestu a rozhodli sme o ďalšom pokračovaní v tejto oblasti. Tiež sme identifikovali potrebu upraviť spôsob spájania n-gramov.

## Synonymá

Okrem riešenia cez n-gramy rozvíjame aj využitie synonymým pri vyhľadávaní. Na stretnutí sme overovali fungovanie implementovaného počítania term frequency na strane elasticsearchu.

Hlavným problémom je nepochopiteľné správanie prekladača, ktorý pracuje so synonymami. Spočíva v neschopnosti nájsť hľadané vety aj keď vieme, že sa v korpuse nachádzajú. Preto sme riešenie tohto problému identifikovali ako prioritu pre ďalšiu prácu so synonymami.

## Určenie nových úloh

Úlohy do ďalšieho týždňa sú zamerané hlavne na to aby prekladač dosahoval lepšie výsledky pri prekladaní a aby sa odstránili problémy, s ktorými sa teraz stretávame. Ďalšou dôležitou časťou je príprava na súťaž IIT.SRC.

## Prehľad úloh vyplývajúcich z diskusie

ID	Popis	Pridelená	Predpokladané ukončenie	stav	Skutočné ukončenie
3542	Vytvorenie posteru pre IIT.SRC	Róbert Horváth, Peter Macko	23.4.2012	nová	
3543	Dopracovať úlohu s tagovaním čísloviek	Peter Sládeček	23.4.2012	nová	
3544	Nájdenie vhodných viet v korpuse pre preklad	Peter Jurčík	23.4.2012	nová	
3545	Zlepšenie metódy spájania n-gramov	Matúš Vacula	23.4.2012	nová	
3546	Vytvoriť testovací index s vhodnou vzorkou údajov	Maroš Ubreži	23.4.2012	nová	
3547	Aplikovať term frequency na n-gramy	Vladimír Ruman	23.4.2012	nová	

**Spracoval:** Bc. Vladimír Ruman

**Overil:** Ing. Dušan Zeleník