

<i>Prítomní:</i>	Bc. Róbert Horváth	<i>Vedúci:</i>	Ing. Dušan Zeleník
	Bc. Peter Jurčík (vedenie)	<i>Miesto:</i>	Softvérové štúdio
	Bc. Peter Macko (zápis)	<i>Dátum a čas:</i>	5.3.2012 14:00
	Bc. Vladimír Ruman	<i>Trvanie:</i>	180 minút
	Bc. Peter Sládeček		
	Bc. Maroš Ubreži		
	Bc. Matúš Vacula		

Prehľad splnenia úloh z predchádzajúceho stretnutia

ID	Popis	Pridelená	Predpokladané ukončenie	stav	Skutočné ukončenie
2848	Zrýchlenie REST služby	Peter Sládeček	5.3.2012	Ukončená	1.3.2012
2943	Konfigurovateľnosť prekladača	Peter Sládeček	5.3.2012	rozpracovaná	
2944	Vyhľadávanie cez N-gramy	Róbert Hrováth Matúš Vacula	5.3.2012	rozpracovaná	
2954	Vypracovanie stratégií v Jave	Peter Jurčík, Vladimír Ruman	5.3.2012	ukončená	5.3.2012

Základné body stretnutia

- Synonymá a ElasticSearch
- Nová verzia ElasticSearch
- Osekanie počtu prekladov slova
- Úpravy REST-ovej služby
 - o Implementácia stratégií
 - o ID posunúť do konštruktora prekladača
 - o predspracovanie viet
- Samostatná stratégia pre N-gramy

Synonymá a ElasticSearch

ElasticSearch podporuje vytváranie indexov v spolupráci so synonymami slov. Pri preklade potom nie je podstatné, či je vo vete dane slovo alebo jeho synonymum. Viac informácií tu: <http://www.elasticsearch.org/guide/reference/index-modules/analysis/synonym-tokenfilter.html>. Potrebujeme preto získať zoznam synonym a vytvoriť dve nové stratégie prekladania.

1. Získanie zoznamu synonym zo SAV
2. Vytvorenie zoznamu synonym z nášho slovníka

V neskoršej fáze je nutné porovnať tieto dve metódy. Pre každú z nich je dôležité vytvoriť **samo-
statný nový** index.

Nová verzia ElasticSearch

V rámci predchádzajúcej úlohy treba upgradovať aj ElasticSearch, keďže Dušan nám u tvorcu vybavil špeciálny build ☺.

Osekanie počtu prekladov slova

Keďže prekladač trpí hlavne počtom slov, ktoré sa vkladajú do ElasticSearch je nutné ich počet znížiť. A to tým spôsobom, že budeme do hľadania vkladat iba pravdepodobnejšie vety a to pomocou dvoch metód:

1. Zoradenie podľa toho, ako sú vložené dané vety do databázy (poradie od PC Translатора dosť dobré)
2. Zoradenie podľa toho aká je TermFrequency daných slov

Treba teda vytvoriť dve nové stratégie translátora a obmedziť v každom počet vyselektovaných slov na 5. Tento parameter musí byť neskôr konfigurovateľný.

Úprava REST-ovej služby

Keďže sme vytvorili stratégie len nutné medzi nimi jednoducho prepínať. To chceme dosiahnuť ďalším parametrom ktorý budeme vkladat do REST-ovej služby. Tento parameter sa prečíta a podľa hodnoty sa inicializuje daná inštancia Translatora a SentenceFindera.

Služba REST dostáva na vstupe jazyky na preklad vo forme skratiek („sk“, „en“, ...). Tie sa v súčasnej dobe presúvajú do Translator-a cez ID-čka, ktoré sú predspracované zo skratiek. Chceli by sme toto spracovanie posunúť do konštruktora Translatora.

Predspracovanie viet je v súčasnej dobe priamo v službe ako samostatná funkcia bolo by dobre spraviť v Translatore pomocnú triedu pre čistenie.

Samostatná stratégia pre N-gramy

Máme vytvorený tvorič N-gramov z viet. V druhom kole potrebujeme spraviť stratégiu sentence findera ktorá by tieto N-gramy hľadala, a tak vytvárala výstupné vety.

Prehľad úloh vyplývajúcich z diskusie

ID	Popis	Pridelená	Predpokladané ukončenie	stav	Skutočné ukončenie
3045	Úprava REST-ovej služby	Peter Macko	11.3.2012	nová	
3044	Oskanie počtu prekladov slov získaných z prekladača	Peter Jurčík	11.3.2012	nová	
2043	vyuzitie možnosti zoznamu synonym v Elastic-Search	Maroš Ubreži	11.3.2012	nová	

Spracoval: Bc. Peter Macko

Overil: Ing. Dušan Zeleník