

<i>Prítomní:</i>	Bc. Róbert Horváth (diskusia)	<i>Vedúci:</i>	Ing. Dušan Zeleník
	Bc. Peter Jurčík	<i>Miesto:</i>	Softvérové štúdio
	Bc. Peter Macko	<i>Dátum a čas:</i>	29.11.2011 11:00
	Bc. Vladimír Ruman	<i>Trvanie:</i>	150 minút
	Bc. Peter Sládeček		
	Bc. Maroš Ubreži		
	Bc. Matúš Vacula (zápis)		

Prehľad splnenia úloh z predchádzajúceho stretnutia

ID	Popis	Pridelená	Predpokladané ukončenie	stav	Skutočné ukončenie
2217	Automatizovať testovanie prekladača	Peter Sládeček Róbert Horváth	06.12.2011	Ukončené	
2232	Integrovať Elasticsearch do prekladača	Peter Macko Maroš Ubreži Matúš Vacula	06.12.2011	Vo vývoji	
2255	Integrovať dokumentáciu 3. a 4. šprintu	Matúš Vacula	29.11.2011	Vo vývoji	
2256	Napísať zápisnicu z 10. Stretnutia a vložiť úlohy do Redmine	Róbert Horváth	29.11.2011	Ukončené	
2259	Vytvoriť plán na letný semester	Róbert Horváth	29.11.2011	Ukončené	
2261	Vytvoriť backlog projektu	Matúš Vacula	29.11.2011	Vo vývoji	
2265	Vytvoriť zhodnotenie plánov projektu	Vladimír Ruman	29.11.2011	Vo vývoji	
2266	Integrovať metodiky	Peter Jurčík	29.11.2011	Ukončené	
2268	Analyzovať možnosti extrakcie faktov z vety	Peter Jurčík Vladimír Ruman	06.12.2011	Vo vývoji	
2286	Pridanie prekladača a restovej služby do Git repozitára	Peter Macko Vladimír Ruman	06.11.2011	Vo vývoji	

Základné body stretnutia

Na stretnutí boli prediskutované nasledujúce body:

- Stav rozpracovaných úloh z minulého stretnutia
- Čistenie slovníka od slov, ktoré nie sú v základnom tvare
- Znižovanie počtu dopytov na databázu
- Spresnenie prekladu

- Nahradenie generátora slovných tvarov
- Vlastné mená a čísla v preklade

Stav rozpracovaných úloh

Automatický test prekladača bol dokončený. V súčasnosti sa nachádza na stránke tímu. Vykonávajú sa automatické ohodnotenia úspešnosti prekladu v porovnaní s konkurenčnými prekladačmi.

Integrácia dokumentácie šprintu nebola realizovaná, nakoľko sa šprint ešte nezačal dokumentovať.

Plán na letný semester bol vyhotovený. Z diskusie vyplynulo, že je žiaduce vypracovať podrobnejšie plány aj pre konkrétne šprinty.

Forma product backlogu bola dohodnutá. Backlog bude upravený do finálnej podoby do konca šprintu.

Na importe z MySQL do Elastic Search sa stále pracuje. Služba je plne funkčná. Potrebné je už len premiestniť dáta.

Znižovanie počtu dopytov na databázu

Preklad po jednotlivých slovách generuje veľké množstvo viet, ktoré vznikli ako kombinácie všetkých preložených slov. Bola diskutovaná možnosť odstránenia slov, ktoré nie sú v základnom tvare, z prekladového slovníka, nakoľko sa slovné tvary budú generovať v nasledujúcom kroku. Problém bol vyriešený pridaním slova DISTINCT do SQL príkazu.

Spresnenie prekladu

Spresnenie prekladu vety sa vykoná priradením hodnoty pravdepodobnosti prekladu k slovu. Tá sa bude odvídať z počtu výskytov slova v korpuse. Táto metóda by mala súčasne znížiť aj čas vykonávania prekladu, nakoľko by sa požadovaná veta z korpusu mala nájsť skôr.

Nahradenie generátora slovných tvarov

Súčasný spôsob generovania slovných tvarov funguje za pomoci slovníka slovných tvarov pre slovenský jazyk. Získať takýto slovník pre iné jazyky, však môže byť problém. Preto sa v diskusii dohodlo, že sa bude využívať tzv. stemmer - analyzátor, ktorý určí slovotvorný základ. Rovnako v korpuse sa bude vyhľadávať podľa slovotvorných základov.

Vlastné slová a čísla v preklade

Boli prediskutované problémy s prekladom viet obsahujúcich vlastné mená. Čísla sa budú identifikovať a nebudú podliehať prekladu. Do prekladu sa dostanú pôvodné čísla. Vlastné mená spôsobujú problémy ak napríklad sa napríklad nachádzajú v slovníku. V prípade ak sa jedná o vlastné meno, je preklad nežiaduci. Identifikovať vlastné meno na základe veľké začiatočného písmena nie je postačujúce. Diskusia ohľadom tohto problému ostala otvorená.

Prehľad úloh vyplývajúcich z diskusie

ID	Popis	Pridelená	Predpokladané ukončenie	stav	Skutočné ukončenie
2443	Vytvoriť plány za jednotlivé šprinty	Róbert Horváth	4.12.2011	Nová	