NAKI-II-UJC-UKONCENE - Task #3707

Task # 3637 (Closed): Etapa 01 - Nová strukturace poradenských hovorů – návrh automatické segmentace

Zpracování dat z emailů - témata

20.01.2016 18:18 - Zajíc Zbyněk

Status:	Closed	**Start date:**	20.01.2016
Priority:	Normal	**Due date:**	01.08.2017
Assignee:	Skorkovská Lucie	**% Done:**	50%
Category:		**Estimated time:**	0.00 hour
Target version:			

Description

Zkusit na emailech (až je ÚJČ předá na FTP #3696)
-detekci klíčových slov
-klustrovací metody

History

#1 - 09.03.2016 13:54 - Zajíc Zbyněk

data od ÚJČ budou na poli v korpusech v adresáři NAKI-II-UJC, heslo napsána na redmine wikki projektu naki-ii-ujc-privat

#2 - 29.04.2016 12:30 - Skorkovská Lucie

- Due date changed from 01.05.2016 to 23.05.2016

- % Done changed from 0 to 20

Přehled výsledků připravit před schůzí s UJČ.

#3 - 24.05.2016 09:33 - Skorkovská Lucie

- File kmeans.log added

- % Done changed from 20 to 50

Testování shlukovacích algoritmů na dopisech ukázalo, že dokážeme najít často se vyskytující jevy v dotazech:

- psaní cikán / rom
- přechylování příjmení
- psaní ulic s předložkou
- jaký pád se používá při oslovení
- psaní počátečních velkých písmen
- pravidla velkých písmen u psaní názvů měst
- psaní přídavných jmen "řídící" "měřící" "kropící" .....
- skloňování příjmení
- že jazyková poradna není právní poradna
- norma úpravy písemností strojem

Další postup

- v přípravě zapojení lemmatizace, normalizace - může trochu zlepšit výsledky
- chtělo by to anotované dopisy

#4 - 25.05.2016 11:14 - Zajíc Zbyněk

Aleš říká, že online přepis není problém (kontinuálně i např. po pauze předávat slova), proto zkoumej i **klasifikaci témat online**.

#5 - 31.05.2016 14:01 - Skorkovská Lucie

- Due date changed from 23.05.2016 to 30.06.2016

- lematizace dost vylepšuje shlukování témat
- na jakékoli další experimenty to chce anotovana data

#6 - 31.05.2016 14:32 - Skorkovská Lucie

- File kmeans_50clusters_lemma.log added

#7 - 10.04.2017 08:10 - Zajíc Zbyněk

- Due date changed from 30.06.2016 to 01.08.2017

#8 - 18.11.2019 09:03 - Zajíc Zbyněk

- Status changed from Assigned to Closed

Files

kmeans.log	30.7 KB	24.05.2016	Skorkovská Lucie
kmeans_50clusters_lemma.log	162 KB	31.05.2016	Skorkovská Lucie