

Zápisnica 6

Dátum	24.10.2018
Miesto	FIIT STU, 4.26
Zapisovateľ	Dávid Csomor

Úvod

- Poslať najdlhší článok vedúcemu, nech môže nastaviť limit (asi netreba)
- Dokážeme identifikovať rozlišovacia stránku?
 - o také spoznáme ak má na začiatku „môže byť“ (iba ak je to na začiatku)
 - o zrušiť rozlišovacie stránky
- Zachovať odseky ak sa nám to podarí
- Dávať bodku:
 - o za nadpisy vo wiki
 - o za nadpisy v článkoch
- Tlačidlá typu „tokenizovať“, by sme mali vedieť niekedy v budúcnosti schovať
- Odstraňovať z web-novín „Bratislava...“

Prehľad šprintu

- Crawler je hotový ako samostatný skript
- Features pozeral Peťo:
 - o Skúšal niekoľko indexov na sledovanie náročnosti textu
- Metrika od vedúceho:
 - o koľko z našich slov je v TOP 500
- Pridať TAB: „Analyzovaný text“
- Pridať filter na korpus (SELECT)
- V TABe „Zoznam článkov“ pridať počet znakov/slov článku
- 300 pridaných osobností postačí
- Invertovaný index má spravený prototyp rozhrania a základnú schému

Invertovaný index

- Vytvoriť IE podľa predpisu v TFS – Features - „9321 Vytvoriť invertovaný index pre tokeny z článkov“

Ostatné

- Keď už budeme mať ten invertovaný index, tak až potom by sme spravili slovník
- Pouvažovať nad aktivitou, ktorú by sme vložili do nášho inteligentného analyzátora
 - o zoberieme článok, nejaké typické slova a kto prvý uhádne kategóriu
 - o napr. slovo „futbalista“ a do akej kategórie patrí
 - o prípadne „Akinátor“
- Treba poupratovať User Stories