Výsledky z vlastního programu pro trigramový model

Český text s diakritikou

Název souboru	Logaritmus pravděpodobnosti, že text byl vygenerován českým korpusem	Logaritmus pravděpodobnosti, že text byl vygenerován slovenským korpusem
text1-CZ	-1837.549	-2614.960
text2-CZ	-2175.443	-3031.171
text3-CZ	-1022.450	-1287.579
text4-CZ	-318.269	-431.142
text5-CZ	-509.942	-713.538
text6-CZ	-1003.920	-1169.694
text7-CZ	-101161.099	-133117.655

Český text bez diakritiky

Název souboru	Logaritmus pravděpodobnosti, že text byl vygenerován českým korpusem	Logaritmus pravděpodobnosti, že text byl vygenerován slovenským korpusem
text1-CZ-ND	-1798.650	-2084.475
text2-CZ-ND	-2138.313	-2429.390
text3-CZ-ND	-991.603	-1029.491
text4-CZ-ND	-315.752	-362.478
text5-CZ-ND	-495.752	-568.461
text6-CZ-ND	-960.864	-1017.583
text7-CZ-ND	-99070.702	-108947.449

Slovenský text s diakritikou

Název souboru	Logaritmus pravděpodobnosti, že text byl vygenerován českým korpusem	Logaritmus pravděpodobnosti, že text byl vygenerován slovenským korpusem
text1-SK	-2081.423	-1654.489
text2-SK	-1854.502	-1457.083
text3-SK	-1290.181	-1043.241
text4-SK	-444.295	-390.791
text5-SK	-461.821	-370.153
text6-SK	-644.197	-558.331
text7-SK	-47187.889	-41481.329

Slovenský text bez diakritiky

Název souboru	Logaritmus pravděpodobnosti, že text byl vygenerován českým korpusem	Logaritmus pravděpodobnosti, že text byl vygenerován slovenským korpusem
text1-SK-ND	-1792.675	-1628.622
text2-SK-ND	-1562.111	-1423.913
text3-SK-ND	-1137.401	-1013.746
text4-SK-ND	-426.529	-382.276
text5-SK-ND	-360.699	-364.353
text6-SK-ND	-535.435	-537.137
text7-SK-ND	-99070.702	-39720.031

Výsledky z fastText

Český text s diakritikou

Název souboru	Jazyk s nejvyšší pravděpodobností
text1-CZ	CZ (0.999)
text2-CZ	CZ (0.999)
text3-CZ	CZ (0.997)
text4-CZ	CZ (0.993)
text5-CZ	CZ (0.999)
text6-CZ	CZ (0.978)
text7-CZ	CZ (0.997)

Český text bez diakritiky

Název souboru	Jazyk s nejvyšší pravděpodobností
text1-CZ-ND	CZ (0.545)
text2-CZ-ND	CZ (0.355)
text3-CZ-ND	CZ (0.385)
text4-CZ-ND	CZ (0.377)
text5-CZ-ND	CZ (0.457)
text6-CZ-ND	CZ (0.279)
text7-CZ-ND	CZ (0.539)

Slovenský text s diakritikou

Název souboru	Jazyk s nejvyšší pravděpodobností
text1-SK	SK (0.929)
text2-SK	SK (0.995)
text3-SK	SK (0.949)
text4-SK	SK (0.657)
text5-SK	SK (0.994)
text6-SK	SK (0.997)
text7-SK	SK (0.966)

Slovenský text bez diakritiky

Název souboru	Jazyk s nejvyšší pravděpodobností
text1-SK-ND	SK (0.215)
text2-SK-ND	SK (0.455)
text3-SK-ND	SK (0.246)
text4-SK-ND	PL (0.507)
text5-SK-ND	SK (0.525)
text6-SK-ND	SK (0.563)
text7-SK-ND	SK (0.439)

Závěr

Vytvořený program úspěšně vyhodnotil všechny texty s diakritikou. V případě textů bez diakritiky špatně vyhodnotil, že text v souborech text5-SK-ND a text6-SK-ND je český, byť se jedná o text slovenský. Rozdíl logaritmických hodnot pravděpodobností byl v případech těchto textů u obou relativně malý. Důvodem chybného vyhodnocení je nejspíš malé množství různých druhů textů ve slovenském korpusu. V případě, kdy byl místo trigramového modelu použit kvadrigramový model, byly jazyky všech textů vyhodnoceny správně.
Knihovna fastText dokázala správně vyhodnotit všechny texty s diakritikou. U textů bez diakritiky byla jediná chyba v případě textu v souboru text4-SK-ND, který vyhodnotila jako polský.