

# Určovanie kľúčových slov pre odvíjajúce sa príbehy pomocou roja sociálnych agentov

Pavol Návrat a Štefan Sabo

# Motivácia

- Spravodajské príbehy sú často reprezentované stručnými výrazmi, ktoré sú všeobecne známe a jednoznačne identifikujú konkrétny príbeh.
- Spolu s meniacou sa paletou aktuálnych príbehov sa menia výrazy, ktoré sa v rámci spravodajstva často skloňujú.
- Našou snahou je vyvinúť metódu, ktorá by umožňovala tieto zmeny dynamicky sledovať a identifikovať výrazy, ktoré sa dostávajú do popredia.

# Prehľad metódy

- Dynamický prístup, ktorý porovnáva články a hľadá prieniky za účelom identifikácie zaujímavých príbehov.
- Prehľadávanie priestoru článkov inšpirované sociálnym hmyzom, včelami medonosnými.
- Využívame roj jednoduchých nezávislých agentov. Každý agent nesie vlastné kľúčové slovo na základe ktorého vykonáva porovnania článkov.
- Selekcia kľúčových slov – kvalitné kľúčové slová sú preferované a sledované podrobnejšie, nezaujímavé výrazy sú opúšťané.

# Výstupy

- K jednotlivým článkom identifikujeme kľúčové slová spojené s príbehmi a hodnotíme ich relevanciu k článkom.
- Súčasne budujeme sieť článkov na základe prepájania príbuzných článkov.
- Pre kľúčové slová sledujeme vývoj popularity v čase.

# Populárne kľúčové slová

keyword	n (k)	n (k) / N	keyword	n (k)	n (k) / N
Syria	177.30	6.87 %	court	49.90	1.93 %
Egypt	98.10	3.80 %	ECB	49.85	1.93 %
Apple	92.65	3.59 %	attack	49.41	1.91%
Afghan	78.23	3.03 %	Colorado	41.79	1.62 %
Euro	75.50	2.92 %	trial	28.90	1.12 %
shooting	56.32	2.18 %	Libor	27.75	1.07 %
Samsung	55.71	2.16 %	murder	26.38	1.02 %
China	55.30	2.14 %	Aleppo	25.31	0.98 %



# Ďalšie smerovanie

- Hierarchické zoskupovanie kľúčových slov do viacerých úrovní podľa ich všeobecnosti.
- Orientácia na agregáciu výsledkov a vizualizáciu príbehov.