

ZÍSKAVANIE ATRIBÚTOV Z POPISU OBJEKTOV

Michaela Linková

Vedúci práce: RNDr. Peter Gurský, PhD.

Ústav informatiky

ÚVOD DO PROBLEMATIKY

- Vytvorenie **sady extrakčných metód**, ktoré by mali automaticky extrahovať atribúty z popisu objektov.

Nerezové čepele nožov skutočne vysokej kvality zabezpečujú konzistentnú textúru a optimálne mixovanie. Protišmykové nožičky zaručujú stabilitu pri práci s týmto skvelým mixérom. **Príkon: 300 W.** Kompaktná smoothie ponúka **2 rýchlosti** + pulz. **Kapacita: 0.5l. Strieborné prevedenie. Materiál džbánov: plast. Materiál tela: plast.** Možnosť piť z nádoby cez otvor vo veku. Nerezové čepele nožov vysokej kvality pre konzistentnú textúru a optimálne mixovanie. Bezpečnostný systém. Protišmykové nožičky pre lepšiu stabilitu. Celorozoberateľný pre jednoduché čistenie. Umývateľný v umývačke. Drvenie ľadu. Rozmer prístroja: 31,3 x 16 x 15,2 cm. **Hmotnosť: 1,75 kg.**



Súhrn

Zaradenie	stolný mixér a šľahač , Smoothie
Výrobca	Kenwood
Farba	strieborná
Hmotnosť	1.75 kg.
Príkon	300 W.
Typ mixéra	stolné
Objem misy	0.5 l.
Nastavenie rýchlosti	áno
Počet rýchlostí	2.
Regulácia rýchlosti	áno.
Materiál nádoby	plastové.
Smoothie	áno.

- Navrhnuté metódy **pre rôzne typy** atribútov
- **Analýza** metód na vstupných príkladoch
- Dôvody **neúspešnosti** metód:
 1. Nájdenie atribútu, ktorý sa v popise reálne nenachádza
 2. Výskyt chyby v popise
 3. Neprítomnosť názvu atribútu v slovníku
 4. Zmena tvaroslovia hodnoty alebo názvu atribútu.
 5. Vynechanie slova alebo slov z viacslovných pomenovaní atribútov
 6. Rozdelenie slov viacslovného pomenovania
 7. Použitie skratky slova názvu

CIELE PRÁCE

- 1. **Návrh metód** na extrakciu rôznych atribútov z voľného textu s využitím aktuálnych nástrojov na prácu s prirodzeným jazykom.
- 2. **Implementácia** navrhnutých metód.
- 3. Vytvorenie dostatočnej **dátovej sady** na overenie vytvorených metód.
- 4. **Testovanie a analýza** korektnosti a úplnosti vytvorených metód.
- 5. **Analýza prípadov**, v ktorých navrhnuté metódy zlyhávajú.

PLÁN PRÁCE

- 1. Návrh metód:
 - Lematizácia
 - Stemming
 - Tvaroslovník
 - Skryté Markove modely
- 2. Webové rozhranie na vytvorenie dátovej sady

ODPORÚČANÁ LITERATÚRA

- 1. J. NothMan at al.: Learning multilingual named entity recognition from Wikipedia. *Artificial Intelligence* 194 (2013) 151–175
- 2. D. Nadeau, S. Sekine, A survey of named entity recognition and classification, *Lingvisticae Investigationes* 30 (2007) 3–26
- 3. D. M. Bikel et al.: Nymble: a High-Performance Learning Name-finder. In *ANLP-97*, Washington, D.C., pp. 194 – 201, 1997.
- 4. J. M. Castillo et al.: Named Entity Recognition Using Support Vector Machine for Filipino Text Documents. *International Journal of Future Computer and Communication*, Vol. 2, No. 5, October 2013
- 5. K. Frantzi, S. Ananiadou, J. Tsujii: The C-value/NC-value Method of Automatic Recognition of Multi-word Terms. In *proceedings of ECDL*, pp. 585-604. ISBN 3-540-65101-2, 1998
- 6. Jana Hiľovská: Syntaktická analýza slovenskej vety pomocou Tvaroslovníka. Diplomová práca PF UPJŠ 2017

ĎAKUJEM ZA POZORNOST