

ZÍSKAVANIE ATRIBÚTOV Z POPISU OBJEKTOV

Michaela Linková

Vedúci práce: RNDr. Peter Gurský, PhD.

Ústav informatiky

PROBLEMATIKA

- 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. **Príkion: 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.**



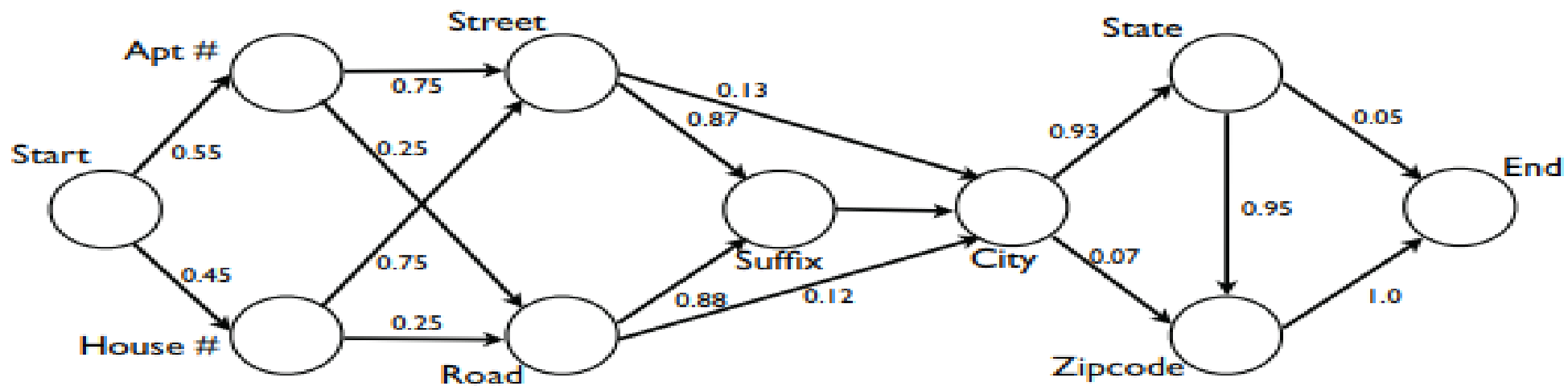
Súhrn

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

PREHĽAD SÚČASNÉHO STAVU

- Named Entity Recognition
 - Support Vector Machine
 - Decision tree
 - Conditional Random Fields
 - Hidden Markov Model
- Extrakcia informácií

HIDDEN MARKOV MODEL



Obrázok 1 HMM pre parsovanie adries

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.

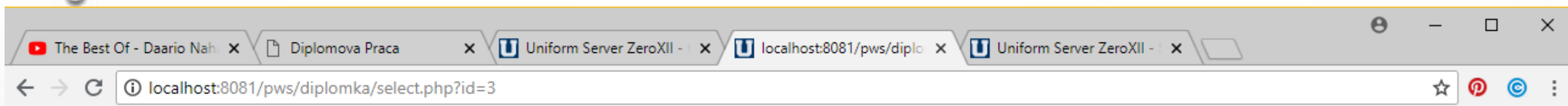
NÁVRH METÓD

- Lematizácia
- Stemming
- Tvaroslovník
- Hidden Markov Model

WEBOVÉ ROZHRAŇIE

Dataset

Description	Action
Alcor Zest	<input type="button" value="Select"/>
Manta MID710	<input type="button" value="Select"/>
Blow Kidstab	<input type="button" value="Select"/>
Prestigio PMP5080C	<input type="button" value="Select"/>
GoGEN TA 8600	<input type="button" value="Select"/>
Trekstor SurfTab Breeze 7.0	<input type="button" value="Select"/>
Sencor Element 7Q104	<input type="button" value="Select"/>
Prestigio MultiPad 7.0 Ultra + PMP3670B	<input type="button" value="Select"/>



Sencor Element 7Q104

Jednou z hlavných výhod tohto modelu je jeho skvelý pomer medzi cenou a výkonom. Výkon je zaistený 4jadrovým procesorom s celkovou frekvenciou 1 GHz. Tento procesor zabezpečí plynulý chod aj tých najnáročnejších aplikácií, bezproblémové prehrávanie HD videí, či surfovanie po internete. Tablet disponuje operačnou pamäťou, ktorá činí 1 GB a stačí aj na pokročilý multi-tasking aplikácií. Tablet je vybavený veľmi sympatickým 7palcovým displejom s celkovým rozlíšením 1280×800px. Takéto rozlíšenie vám zaručuje skvelé predstavenie a sýtosť obrazu, aké ste ešte nezažili. Sledovanie filmov a seriálov sa stane nezabudnuteľným zážitkom. Svoje dôležité okamihy môžete zvečniť pomocou zadného 2Mpx fotoaparátu a vďaka prednému 0,3Mpx fotoaparátu si budete môcť urobiť pekné selfie s vašimi priateľmi. Pre uloženie vašich fotiek a iných multimédií tú je 8 GB internej pamäte, ktorú môžete rozšíriť pomocou pamäťových microSD kariet až na 32 GB.

Výrobca:

Operacny system:

Velkost pamate: GB

Hmotnost: g

Velkost RAM pamate: GB

Velkost displeja: inch

Bluetooth: Ano

Wi-Fi: Ano

GPS: Ano

PLÁN PRÁCE

- 1. Dokončenie webového rozhrania
- 2. Návrh metód
- 3. Implementácia metód
- 4. Analýza metód

TECHNOLÓGIE



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