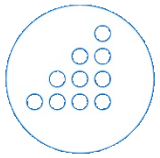


Provozně ekonomická fakulta

Diplomová práce

Autor práce:
Bc. Jiří Švec

Vedoucí práce:
Ing. Jan Přichystal, Ph.D.

Aplikace pro získávání názorů z uživatelských recenzí

Cíl práce

- Webová aplikace
- Automatizované stahování a zpracování recenzí
- Rozšiřitelná o další datové zdroje
- Graficky přehledná a konfigurovatelná
- Uživatel získá přehled o probíraných tématech

Současný stav

- Neexistuje konkurenční aplikace pro hromadné zpracování českých uživatelských recenzí
- Nutnost manuálního vyhledávání a vyhodnocení
- Časově náročný proces

Požadavky

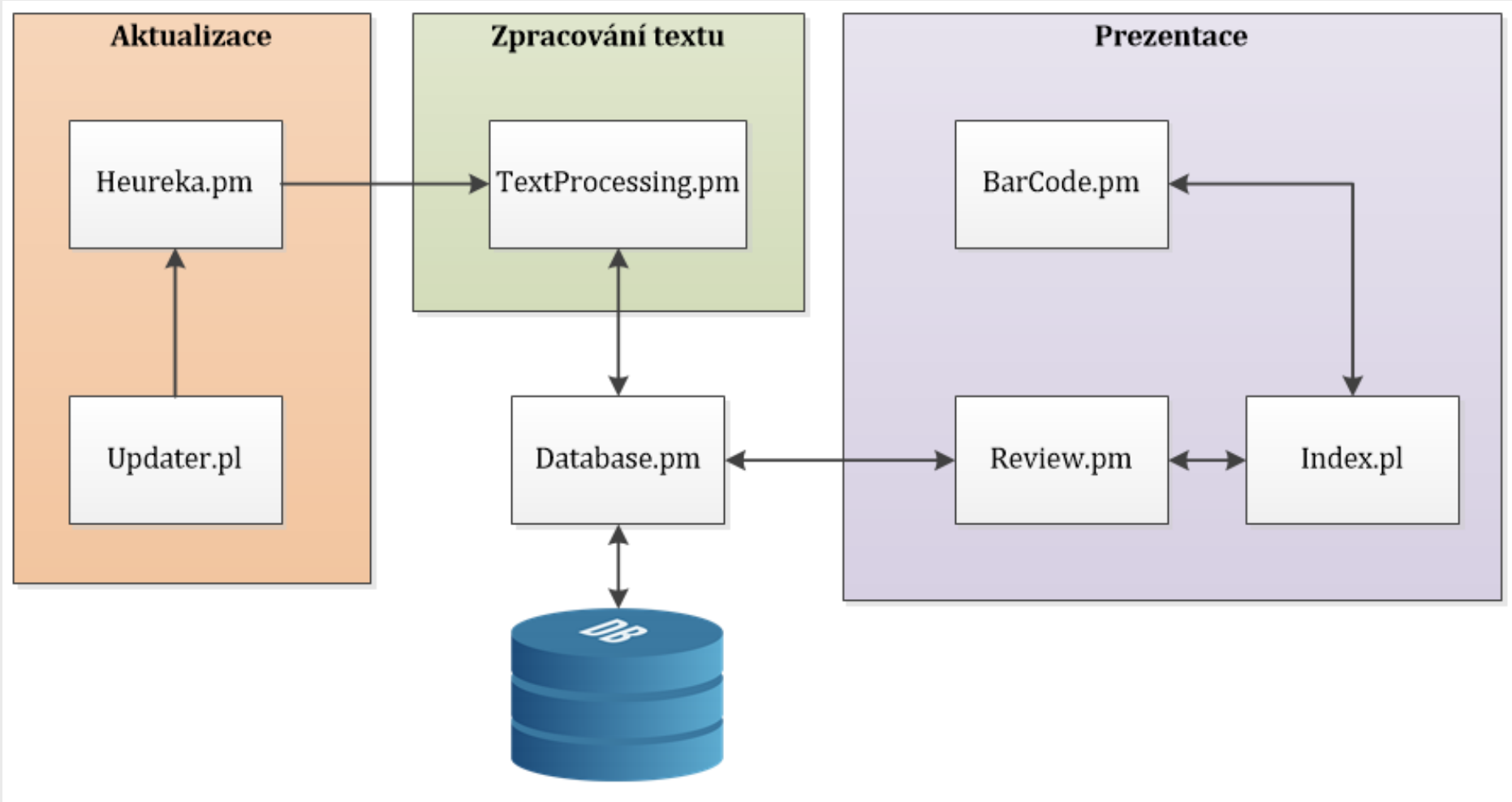
Funkční

- Stahovat recenze
- Zpracovávat stažené recenze
- Ukládat výsledky do databáze
- Nahrávat snímky čárových kódu prostřednictvím kamery mobilního telefonu
- Zobrazovat zpracované recenze

Nefunkční

- Webová aplikace
- Přívětivý vzhled
- Jednoduchá manipulace s aplikací
- Snadná integrace dalších zdrojů recenzí
- GUI v češtině
- Konfigurovatelnost
- Rychlá odezva

Architektura



Konfigurace

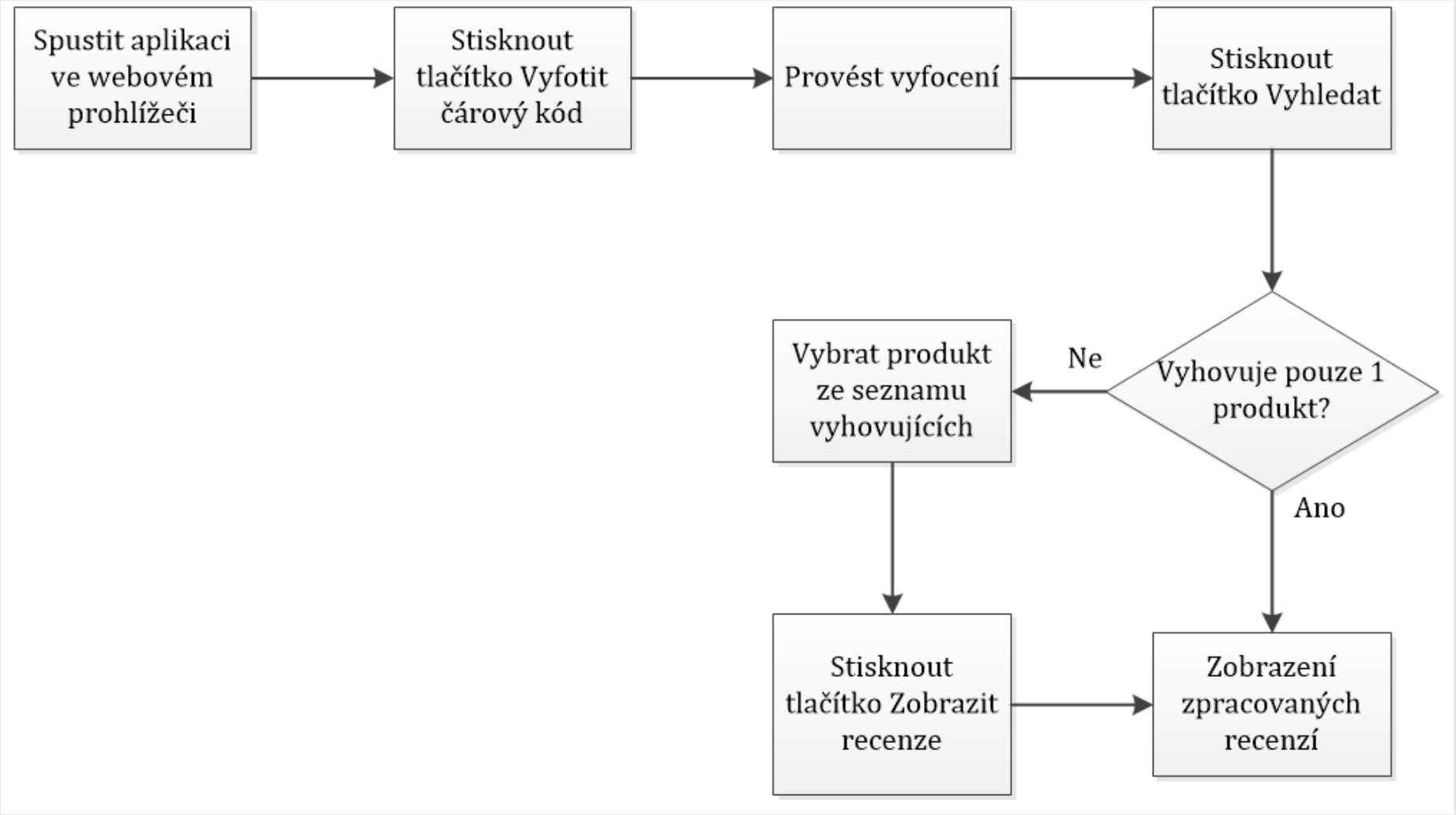
- Nastavení výkonu
- Nastavení aktualizací
- Párování názvů
- Připojení k databázi
- Cesty ke slovníkům
- Nastavení word cloudů
- Nastavení a optimalizace vendora

```
[GLOBAL]
DEBUG=0
THREAD_COUNT=8
THREAD_SLEEP_INT=3
PRODUCTS_PER_FILE_LIMIT=3000
PRODUCT_NAME_PERCENTAGE_MATCH=70
WORD_CLOUD_MIN_WORD_FREQUENCY=2
SHOW_WORD_CLOUD=1
SHOW_WORD_CLOUD_NEIGHBORS=1
UPDATE_EVERY=30
DELETE_ALL_ON_UPDATE=1

# downloaded from https://sites.google.com/site/kevinl
STOPWORDS_PATH=conf/stopwords_cz.txt

# downloaded from http://www.lexiconista.com/datasets,
LEMMATIZATION_PATH=conf/lemmatization_cs.txt

# path to zbarimg tool, empty if installed on linux
ZBARIMG_HOME=


#-----------------------------------------------
[DATABASE]
HOST=localhost
PORT=3306
USER=root
PASSWORD=root
SCHEMA=dp

#-----------------------------------------------
[HEUREKA]
URL=http://www.heureka.cz
POPULATE_PRODUCTS_LIST=0
POPULATE_PRODUCTS=0
POPULATE_REVIEWS=0
GET_ERRORS_LIMIT=5
GET_ERRORS_SLEEP=5
MAX_CATEGORY_PAGE=15000
```

Použité metody a nástroje

- Web scraping
- Synchronizační vzor fronta
- Normalizace
- Stopwords
- Tokenizace
- Stemizace
- Lemmatizace
- Shlukování
- Sousední dvojice
- Asociační pravidla
- Levenshteinova vzdálenost
- Responzivní design

Prezentace názorů 1 / 2



Prezentace názorů 2 / 2

View

Pro zobrazení relevantích recenzí stisktněte tlačítko **Vyfotit čárový kód** a následně **Zobrazit recenze**.

Vyfotit čárový kód

Vyhledat

Apple iPhone 5S 16GB

Frekventovaná slova

Klady a zápory

Klady

- Autonomnost
- Škálovatelnost
- Konfigurovatelnost
- Jedinečnost
- Rozšiřitelnost
- Inkrementální stahování produktů
- Paralelní zpracování
- Responzivní GUI

Zápory

- Zdlouhavý proces pořízení a zpracování dat
- Závislost na slovnících
- Nelze zajistit párování názvů produktu

DĚKUJI ZA POZORNOST