

Fulltextové vyhľadávanie so štruktúrovaným výsledkom

Vedúci práce: RNDr. Peter Gurský, PhD.

Autor: Patrik Sedlák



VYHĽADÁVANIE: SAMSUNG 7"

Nájdene v produktoch (208 výsledkov)

Nájdene v stránkach (24 stránok)

Nájdene v článkoch (3 články)

i Pre hľadaný výraz Samsung 7" existujú aj aktuálne vypredané produkty. [Zobraziť aj tieto produkty](#) alebo [zobraziť len tieto produkty](#)



Samsung Galaxy S7 (biely)

- Uhlopriečka displeja: 5.1"
- Poskytovaný OS: Android
- Vnútrotná kapacita pre ukladanie: 32 GB

5.1"



Samsung Galaxy S7 edge (čierny)

- Uhlopriečka displeja: 5.5"
- Poskytovaný OS: Android
- Vnútrotná kapacita pre ukladanie: 32 GB

5.5"



Samsung Galaxy S7 edge (biely)

- Uhlopriečka displeja: 5.5"
- Poskytovaný OS: Android
- Vnútrotná kapacita pre ukladanie: 32 GB

5.5"



Samsung Galaxy S7 (zltý)

- Uhlopriečka displeja: 5.1"
- Poskytovaný OS: Android
- Vnútrotná kapacita pre ukladanie: 32 GB

5.1"

VYHĽADÁVANIE: SAMSUNG 7"

Nájdene v produktoch (208 výsledkov)

Nájdene v stránkach (24 stránok)

Nájdene v článkoch (3 články)

i Pre hľadaný výraz Samsung 7" existujú aj aktuálne vypredané produkty. [Zobraziť aj tieto produkty](#) alebo [zobraziť len tieto produkty](#)



Samsung Galaxy S7 (biely)

- Uhlopriečka displeja: 5.1"
- Poskytovaný OS: Android
- Vnútrorná kapacita pre ukladanie: 32 GB

5.1"



Samsung Galaxy S7 edge (čierny)

- Uhlopriečka displeja: 5.5"
- Poskytovaný OS: Android
- Vnútrorná kapacita pre ukladanie: 32 GB

5.5"



Samsung Galaxy S7 edge (biely)

- Uhlopriečka displeja: 5.5"
- Poskytovaný OS: Android
- Vnútrorná kapacita pre ukladanie: 32 GB

5.5"



Samsung Galaxy S7 (zltý)

- Uhlopriečka displeja: 5.1"
- Poskytovaný OS: Android
- Vnútrorná kapacita pre ukladanie: 32 GB

5.1"

Kategória?

Heureka.sk » **Vyhľadávanie Samsung 7"**

Nájdene kategórie (20)

[Ochranné fólie pre tablety \(116\)](#), [Puzdrá na tablet PC a čítačky ekníh \(111\)](#), [Foto - Video batérie - neoriginálne \(95\)](#), [LCD monitory \(56\)](#), [Tablety \(46\)](#), [Tvrdené skla pre mobilné telefóny \(44\)](#), [Stylusy \(15\)](#), [Náplne a tonery - originálne \(7\)](#), [Gamepady \(7\)](#), [Náplne a tonery - kompatibilné \(6\)](#), [Viac kategórií »](#)

Cena (€) Od: Do:

Tovar **Nové** bazárové

Zoradiť podľa ▼

Zobrazujem 1 - 20 zo 590 produktov [Ďalšie »](#)



Samsung Galaxy S7 Edge G935F 32GB

(Mobilné telefóny) ★★★★★ (24 recenzií)

Vyvážený design kráčajúci ruka v ruke s dokonalou funkčnosťou? To je nový model Galaxy Edge, ktorého špičková výbava posúva hranice inteligentných telefónov opäť o kúsok ďalej.

[Uložiť na porovnanie](#)

679,00 - 953,78 €
v [67 obchodoch](#)

[Porovnať ceny](#)



Samsung Galaxy A3 2016 A310F

(Mobilné telefóny) ★★★★★ (11 recenzií)

Tento vdarený smartphone spoločnosti Samsung ponúka dokonalý vzhľad, vysoký výkon a výborný fotoaparát, s ktorým vyhotovíte perfektné zábery aj v zlých svetelných podmienkach.

[Uložiť na porovnanie](#)

219,99 - 355,06 €
v [70 obchodoch](#)

[Porovnať ceny](#)

Kategórie 😊


7 palcov ? 😞

Motivácia

- Žiadne alebo nepostačujúce **vyhľadávanie nad viacerými atribútmi** a viacerými **kategóriami** produktov
- Neutriedenie výsledkov (napríklad podľa toho, na koľko percent spĺňajú kritéria vyhľadávania, teda podľa ratingu)
- **Výsledky nie sú štruktúrované**



Ciele práce

- **Navrhnuť metódu** fulltextového vyhľadávania nad produktmi internetových obchodov, ktorej výsledkom je okrem zoznamu produktov, ktoré najviac zodpovedajú vstupnej otázke, aj štruktúrovaná informácia o úspešnosti vyhľadávania podľa domén a atribútov.
 - **Implementovať** navrhnutú metódu ako modul v projekte Kapsa.
 - **Analyzovať** rýchlosť, presnosť a úplnosť navrhnutej metódy a jej porovnanie bežným fulltextovým vyhľadávaním nad reálnymi dátami.
- 

Ako fulltextovo vyhľadávať ?

Využijeme Elasticsearch, ktorý:

- vyhľadáva nad JSON štruktúrovanými dokumentmi pričom výsledok je opäť štruktúrovaný JSON dokument
- poskytuje **štandardné fulltextové vyhľadávanie**
- dáta v reálnom čase
- vysoká dostupnosť
- je zdarma



elasticsearch.



Naša reprezentácia produktov v ES

```

 "_index": "kapsa",
 "_type": "documents",
 "_id": "AVfmU0024_a04GFQcuXu",
 "_score": 0.4958561,
 "_source": {
   "id_object": 68685,
   "id_domain": 4,
   "id_source": 17,
   "domain": [
     : "Tablety",
     : null
   ],
   "attribute_4": "99 eur",
   "attribute_24": "26320",
   "attribute_32": "193.7 mm",
   "attribute_36": "344 g",
   "attribute_40": "0.512 GB",
   "attribute_33": "4 mm",
   "attribute_37": "4000 mAh",
   "attribute_41": "nie",
   "attribute_34": "5 mm",
   "attribute_38": "7 \\"",
   "attribute_42": "áno",
   "attribute_35": "3874 mm3",
   "attribute_39": "8 GB",
   "attribute_44": "áno",
   "attribute_26": "Samsung Galaxy 2.0 P3110 8GB - posledné vystavené kusy",
   "content": "cena 99 eur Hmotnosť Weight 344 g Battery capacity Kapacita batérie 4000 mAh Uhlopriečka displeja Display size 7 \\" Storage Kapacita úložného priestoru HDD 8 GB Local identifier Lokálny identifikátor 26320 Memory RAM Pamäť RAM 0.512 GB GPS nie Bluetooth áno WiFi áno Výška 193.7 mm Šírka 4 mm Hĺbka 5 mm Objem 3874 mm3 Názov Name Samsung Galaxy 2.0 P3110 8GB - posledné vystavené kusy"

```


Dopyt pre požadovanú veľkosť

POST `_search`

```
{  
  "query": {  
    "match": {  
      "content": {  
        "query": "samsung size 7",  
        "fuzziness": 2  
      }  
    }  
  },  
  "highlight" : {  
    "fields" : {  
      "content" : {  
        "fragment_size": 20,  
        "term_vector": "with_positions_offsets",  
        "highlight_query" : {  
          "match": {  
            "content": {  
              "query": "samsung size 7",  
              "fuzziness": 2  
            }  
          }  
        }  
      }  
    }  
  }  
}
```

<- dopyt

<- vyznačenie výskytov (highlight)



Highlightované zhody

```
"_index": "kapsa",
"_type": "documents",
"_id": "AVfmU0024_a04GFQcuXu",
"_score": 0.4958561,
"_source": {
  "id_object": 68685,
  "id_domain": 4,
  "id_source": 17,
  "domain": [
    : "Tablety",
    : null
  ],
},
"attribute_4": "99 eur",      "attribute_24": "26320",    "attribute_32": "193.7 mm",
"attribute_36": "344 g",     "attribute_40": "0.512 GB", "attribute_33": "4 mm",
"attribute_37": "4000 mAh",  "attribute_41": "nie",     "attribute_34": "5 mm",
"attribute_38": "7 \\",     "attribute_42": "áno",    "attribute_35": "3874 mm3",
"attribute_39": "8 GB",     "attribute_44": "áno",    "attribute_26": "Samsung Galaxy 2.0 P3110 8GB - posledné vystavené kusy",
"content": "cena 99 eur Hmotnosť Weight 344 g Battery capacity Kapacita batérie 4000
mAh Uhlopriečka displeja Display size 7 \ Storage Kapacita úložného priestoru HDD 8
GB Local identifier Lokálny identifikátor 26320 Memory RAM Pamäť RAM 0.512 GB GPS
nie Bluetooth áno WiFi áno Výška 193.7 mm Šírka 4 mm Hĺbka 5 mm Objem 3874 mm3 Názov
Name Samsung Galaxy 2.0 P3110 8GB - posledné vystavené kusy"
```

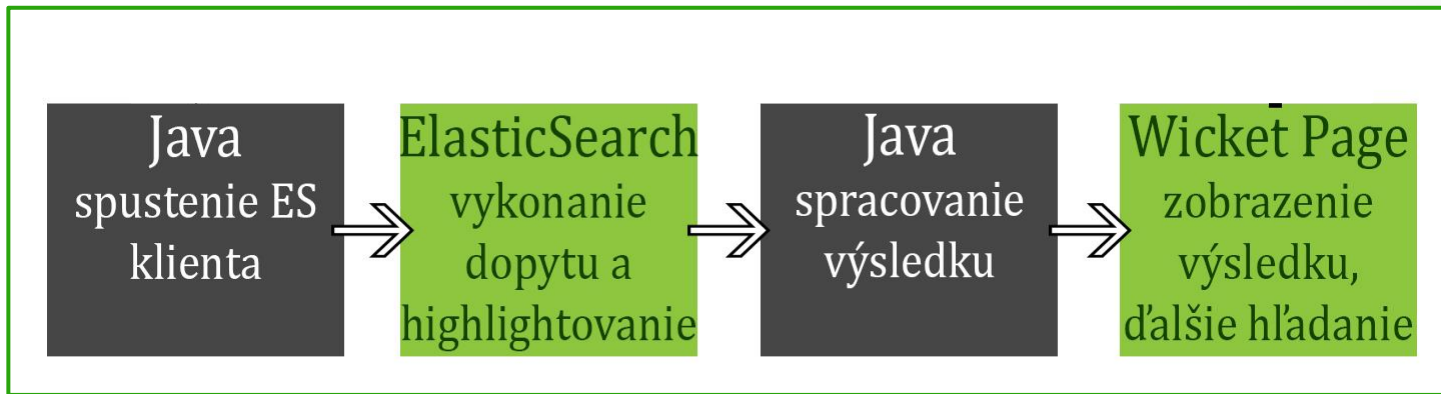
Štruktúrovaný výsledok

- V ES sa vykoná **dopyt nad poľom „content“** podľa vstupnej otázky a vytvorí sa **highlight**
- V Jave **spracujeme** výsledok (vyextrahujeme z highlightu slová vyhodnotené ako zhody)
- Prejdeme celým výsledkom a zisťujeme, ktoré **atribúty** obsahujú slová vyhodnotené ako zhody a počítame pre ne **početnosť**
- Zistíme **početnosť** aj **pre jednotlivé domény**



Štruktúrovaný výsledok

- Na základe získaných informácií vytvoríme **strom domén a atribútov** popisujúci výsledky vyhľadávania



Štruktúrovaná informácia o výsledku

Hľadaj

Tablety (58)

Názov (31)

Uhlop. displeja (24)

Hĺbka (15)

Notebooky (23)

Uhlop. displeja (15)

Názov (12)

Asus MeMO Pad 7 ME70C-1B010A, biely

kategória: tablet

cena: 79 eur

kapacita batérie: 3220 mAh

uhlop. displeja: 7"

-
-
-

Analýza navrhnutej metódy

- **Presnosť** – popisuje kvalitu vrátených výsledkov

$$\textit{presnosť} = \frac{\textit{relevantné vrátené}}{\textit{všetky vrátené}}$$

- **Úplnosť** – popisuje kvantitu vrátených výsledkov

$$\textit{úplnosť} = \frac{\textit{relevantné vrátené}}{\textit{všetky relevantné v DB}}$$

- **Rýchlosť** – stanovenie limitu pre počet výsledkov, ak počet výsledkov je rovný limitu, dopyt je príliš všeobecný



Krok za krokom

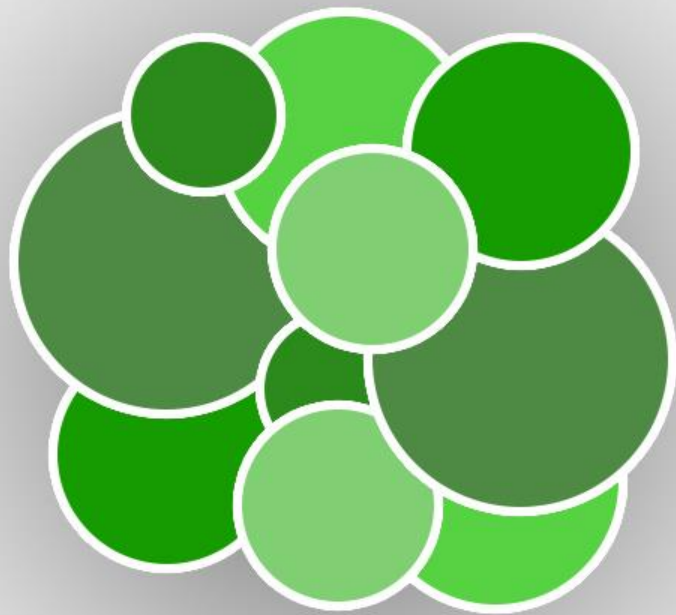
- ✓ Naštudovať Wicket framework
- ✓ Nainštalovať Elasticsearch a naučiť sa s ním pracovať
- ✓ Naštudovať literatúru k danej tematike
- ✓ Navrhnuť, respektíve modifikovať už existujúcu metódu fulltextového vyhľadávania
 - Otestovať a analyzovať navrhnutú metódu
 - Skoncipovať a formulovať text k práci



Odporúčaná literatúra

- [1] Carstens, Carola ; Womser-Hacker, Christ: Ontology Based Query Expansion - Retrieval Support for the Domain of Educational Research. 2012 [online] Dostupné na <<http://d-nb.info/1023809400>>.
- [2] Baeza-Yates, Ricardo A. and Ribeiro-Neto, Berthier: Modern Information Retrieval, 1999 ISBN 020139829
- [3] Tvarožek, M., Bieliková, M. Bridging Semantic and Legacy Web Exploration: Orientation, Revisitation and Result Exploration Support. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. [online] Dostupné na <<http://acadmedia.wku.edu/Zhuhadar/WI-IAT%202010/Data/4191c326.pdf>>
- [4] Jozef Kuper: Unifikácia metadát o produktoch internetových obchodov. Diplomová práca, PF UPJŠ, 2015





Ďakujem za pozornosť