

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra informačních technologií

Studijní program: Aplikovaná informatika
Studijní obor: Informační systémy a technologie

VYHLEDÁVÁNÍ JAKO SLUŽBA

DIPLOMOVÁ PRÁCE

Student: Bc. Luděk Veselý
Vedoucí: Prof. Ing. Zdeněk Molnár, CSc.

2017

University of Economics in Prague
Faculty of Informatics and Statistics
Department of Information Technologies

Study programme: Applied Informatics
Study branch: Information Technologies

SEARCH AS A SERVICE

DIPLOMA THESIS

Student: Bc. Luděk Veselý
Supervisor: Prof. Ing. Zdeněk Molnár, CSc.

2017

Tento list nahradte
originálem zadání.

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně a že jsem uvedl všechny použité prameny a literaturu, ze které jsem čerpal.

Datum:

Podpis:

Abstrakt

Tato diplomová práce popisuje návrh a tvorbu fulltextového vyhledávání poskytovaného jako služba.

Klíčová slova

Fulltext, Elasticsearch

Abstract

This diploma thesis describes creation of fulltext search service.

Key words

Fulltext, Elasticsearch

Poděkování

Rád bych poděkoval Ing. Ivanovi Jelínkovi za rady a pomoc při řešení.

Obsah

Seznam zkratek	9
1 Úvod	10
1.1 Cíle práce	10
1.2 Cílová skupina	10
1.3 Použité metody	11
1.4 Struktura práce	11
2 Analýza byznys požadavků na aplikaci	12
2.1 Vysvětlení problému, který je řešen	12
2.2 Specifika prostředí e-commerce	12
2.3 Definice požadavků na vyhledávání	12
2.4 Technické požadavky na aplikaci	12
2.5 Popis oborů, kterých se práce dotýká	12
2.6 Existující služby umožňující implementaci vyhledávání	12
3 Teorie fulltextového vyhledávání	13
3.1 Indexace dokumentů	13
3.2 Analýza dat, v kterých bude vyhledáváno	13
3.3 Zpracování přirozeného jazyka	13
4 Návrh řešení	14
4.1 Nástroje použitelné pro fulltextové vyhledávání	14
4.2 Definice případů užití	14
4.3 Návrh architektury aplikace	14
5 Implementace	15
5.1 Nastavení a nasazení Elasticsearch	15
5.2 Backend	15
5.3 Frontend	15
5.4 Deployment	15
5.5 Testování, ověření	15
6 Závěr	16
6.1 Dosažení vytčených dílů	16
6.2 Diskuze možného budoucího rozšiřování	16

Seznam zkratk

fulltextové vyhledávání	Vyhledávání v indexovaných dokumentech, v české literatuře často označováno jako plnotextové vyhledávání
e-shop	Elektronický obchod
e-commerce	Elektronické obchodování, souhrnné označení obchodní činnosti prováděné na internetu

1 Úvod

V této diplomové práci se zabývám návrhem a implementací nástroje, který umožní provozovatelům e-shopů snadnou a rychlou implementaci fulltextového vyhledávání. Při implementaci fulltextového vyhledávání je totiž třeba mít určitou znalost principů samotného vyhledávání, souvisejících oborů a technologií. Implementace kvalitního vyhledávání tak může být pro provozovatele e-shopu náročná jak časově, tak finančně.

V rámci své diplomové práce navrhuji a vytvářím nástroj, který řeší tyto problémy a usnadňuje tak provozovatelům e-shopů spuštění vyhledávání. Konkrétní cíle diplomové práce popisují v následující kapitole. V další kapitole volím metody, kterými dané cíle naplním.

1.1 Cíle práce

Cílem práce je vytvořit nástroj, který bude nabízen jako služba a umožní rychlou implementaci fulltextového vyhledávání do e-shopu. K dosažení tohoto cíle je třeba naplnit cíle dílčí.

Prvním dílčím cílem je provedení analýzy problému a to jak z pohledu byznysového, tak z pohledu samotné problematiky vyhledávání. Co se týče pohledu byznysu, tak bude třeba analyzovat potřeby zákazníků, zjistit, co je vlastně třeba vytvořit. Z pohledu problematiky vyhledávání je pak třeba provést analýzu této problematiky a utřídit tak znalosti potřebné k poskytnutí kvalitních výsledků vyhledávání.

Dalším dílčím cílem je vytvořit návrh řešení problému na základě znalosti potřeb zákazníků a dále na základě znalosti problematiky vyhledávání. Tento návrh bude v dalším kroku implementovatelný.

Posledním dílčím cílem je implementace dle návrhu. Obnáší to výběr vhodných nástrojů a konečně také ověření samotné implementace. Tu je třeba porovnat vůči požadavkům, aby se ověřilo, že byly splněny.

1.2 Cílová skupina

První cílovou skupinou je vývojář, řešící problém fulltextového vyhledávání produktů při implementaci elektronického obchodu. Takovému čtenáři by měla práce poskytnout dostatečné teoretické znalosti potřebné pro implementaci vyhledávání. Užitečná také může být konkrétní implementace, popisovaná v této práci.

Další možnou cílovou skupinou je provozovatel e-shopu, který přemýšlí, jakým způsobem zlepšit fulltextové vyhledávání. V této práci získá přehled o složitosti samotné implementace, poskytne mu komentovaný soupis možných řešení problémů a případně poskytne funkční službu, kterou může okamžitě na e-shop napojit.

Poslední cílovou skupinou je kdokoli, kdo se zajímá o fulltextové vyhledávání a českém jazyce. V této práci nalezne soupis problémů souvisejících s češtinou, které je třeba řešit. Dále zde nalezne konkrétní implementaci, kterou se může inspirovat při řešení obdobného problému.

1.3 Použité metody

V této kapitole popisují metody použité k naplnění jednotlivých cílů. Pro zkoumání problému oboru provedu analýzu trhu, nabízí se také možnost dotazování potenciálních uživatelů. Pro porozumění problematice vyhledávání provedu rešerši literatury. Vzhledem k množství dostupné literatury bude třeba provést syntézu těchto informací. Pro vytvoření návrhu řešení bude použito ... Samotné vytvoření aplikace bude implementováno(?).

1.4 Struktura práce

V první části práce popisují problematiku vyhledávání v prostředí e-commerce. Definuji zde jednak kontext, v kterém se pohybuji a popisují problémy, které v tomto prostředí řeším. Snažím se identifikovat potencionálního zákazníka a definovat jeho požadavky na vyhledávání. Toto prostředí má svá specifika, která také popisují. V závěru této části porovnávám existující nástroje umožňující implementaci vyhledávání a zjišťuji tak, proč má smysl vytvářet další službu, jaká je její přidaná hodnota.

V druhé části se zabývám samotnou problematikou fulltextového vyhledávání. Popisují zde celý proces od analýzy vstupních dat, přes jejich indexaci, až po samotné vyhledávání. Tyto poznatky budou následně využity k naplnění požadavků na vyhledávání, k zajištění kvalitních výsledků.

V další části práce navrhuji samotnou aplikaci, aby vyhovovala požadavkům a zároveň byla následně implementovatelná. Porovnávám zde dostupné nástroje, definuji přídav užití aplikace a vytvářím model nástroje.

V poslední části popisují konkrétní implementaci v jazyce Go s pomocí databáze Elasticsearch. Výstupem této části je otestovaný spustitelný program, který umožňuje provádět indexaci produktů a jejich následné vyhledávání.

2 Analýza byznys požadavků na aplikaci

2.1 Vysvětlení problému, který je řešen

2.2 Specifika prostředí e-commerce

- Specifika prostředí e-commerce, definice kontextu, ve kterém bude nástroj používán

2.3 Definice požadavků na vyhledávání

- Definice požadavku na samotné vyhledávání, akceptační kritéria

2.4 Technické požadavky na aplikaci

- Technické požadavky (datové formáty, rychlost odezvy...)

2.5 Popis oborů, kterých se práce dotýká

- Získávání informací, NLP, big data...

2.6 Existující služby umožňující implementaci vyhledávání

- Algolia, AWS CloudSearch...

3 Teorie fulltextového vyhledávání

3.1 Indexace dokumentů

3.2 Analýza dat, v kterých bude vyhledáváno

3.3 Zpracování přirozeného jazyka

- NLP
- Sklonování, časování, tvaroslovi obecně
- Vaha slov, stop slova
- Preklepy
- Synonyma, zargon
- Lemmatizace
- Tezaurus
- ...

4 Návrh řešení

4.1 Nástroje použitelné pro fulltextové vyhledávání

- Popis a porovnání vhodných nástrojů
 - MySQL
 - Elasticsearch
 - Sphinx
 - ...

4.2 Definice případů užití

- Uživatel zadá XML
- Uživatel vyhledá - ...

4.3 Návrh architektury aplikace

- Frontend vs backend (API)...

5 Implementace

- Vyber technologii a nastroju (jazyk, framework...)

5.1 Nastavení a nasazení Elasticsearch

- Implementace indexace
- Implementace vyhledávání

5.2 Backend

- Popis tříd, rozhraní, implementační detaily

5.3 Frontend

- Frontend: UI, komunikace s backendem

5.4 Deployment

- Deployment, CI

5.5 Testování, ověření

- Overení funkčnosti a přínosu
- Testování, zda jsou splněna akceptační kritéria
- Porovnání s stávajícím řešením vyhledávání na konkrétním příkladě

6 Závěr

V této diplomové práci jsem...

6.1 Dosažení vytčených dílů

...

6.2 Diskuze možného budoucího rozšiřování

...

Literatura

- [1] STROSSA, Petr. *Počítačové zpracování přirozeného jazyka*.
Vyd. 1. Praha: Oeconomica, 2011 316 s. ISBN 978-80-245-1777-3.
- [2] AYSE Göker a DAVIES John. *Information Retrieval: Searching in the 21st Century*.
Vyd. 1. Library of Congress Cataloging-in-Publication Data, 2009 295 s.
ISBN: 978-0-470-02762-2.
- [3] AGGARWAL Charu C., ZHAI ChengXiang. *Mining Text Data*.
Springer New York Dordrecht Heidelberg London, 2000 522 s. ISBN 978-1-4614-3222-7.
- [4] CLINTON Gormley, ZACHARY Tong. *Elasticsearch: The Definitive Guide*.
O'Reilly Media, 2015 724 s. ISBN 978-1-4493-5854-9.
- [5] KENNEDY William. *Go in Action*.
Manning Publications Co, 2016 264 s. ISBN 978-1-6172-9178-4.
- [6] Elasticsearch. *Elasticsearch Reference* [online].
2017 [cit. 2017-03-03]. Dostupné z:
<https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>

Rejstřík

...

A Obsah přiloženého DVD

- Soubor Diplomova_prace_2017_Ludek_Vesely.pdf
 - Text diplomové práce