

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

1	Úvod	9
1.1	Cíle práce	9
1.2	Cílová skupina	9
1.3	Použité metody	10
1.4	Struktura práce	10
2	Analýza byznys požadavků na aplikaci	11
2.1	Vysvětlení základních pojmů	11
2.2	Vysvětlení problému, který je řešen	13
2.3	Specifika prostředí elektronického obchodování	14
2.4	Definice požadavků na vyhledávání	14
2.5	Technické požadavky na aplikaci	15
2.6	Popis oborů, kterých se práce dotýká	15
2.7	Stávající možnosti implementace vyhledávání	16
3	Teorie fulltextového vyhledávání	18
3.1	Indexace dokumentů	18
3.2	Analýza dat, v kterých bude vyhledáváno	18
3.3	Zpracování přirozeného jazyka	18
4	Návrh řešení	19
4.1	Nástroje použitelné pro fulltextové vyhledávání	19
4.2	Definice případů užití	19
4.3	Návrh architektury aplikace	19
5	Implementace	20
5.1	Nastavení a nasazení Elasticsearch	20
5.2	Backend	20
5.3	Frontend	20
5.4	Deployment	20
5.5	Testování, ověření	20
6	Závěr	21
6.1	Dosažení vytčených dílů	21
6.2	Diskuze možného budoucího rozšiřování	21

1 Úvod

V této diplomové práci se zabývám návrhem a implementací nástroje, který umožňuje provozovatelům elektronických obchodů snadnou a rychlou implementaci plnotextového vyhledávání. Při jeho implementaci je totiž třeba mít určitou úroveň znalosti principů samotného vyhledávání včetně souvisejících oborů a technologií. Implementace kvalitního vyhledávání tak může být pro provozovatele elektronického obchodu náročná jak časově, tak finančně. Tento nástroj je poskytován jako služba se všemi výhodami i omezeními s tímto principem spojenými. Konkrétní cíle diplomové práce popisuji v následujících odstavcích.

1.1 Cíle práce

Cílem práce je vytvořit nástroj, který je nabízen jako služba a umožňuje snadnou implementaci plnotextového vyhledávání do elektronického obchodu. 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 byznysového, tak je třeba analyzovat potřeby zákazníků, zjistit, jaká jsou specifika oblasti elektronického obchodování a definovat kritéria, jejichž naplnění je pro provozovatele elektronických obchodů klíčové. Z pohledu problematiky vyhledávání je pak třeba provést analýzu této disciplíny a utřídit tak znalosti potřebné k poskytnutí kvalitních výsledků plnotextového vyhledávání.

Dalším dílčím cílem je vytvořit návrh řešení problému jednak na základě znalosti potřeb zákazníků a také na základě znalosti problematiky plnotextového vyhledávání. Tento návrh musí být v dalším kroku implementovatelný.

Posledním dílčím cílem je samotná implementace aplikace dle jejího návrhu. Obnáší to výběr vhodných nástrojů, jejich nasazení do produkčního prostředí, naprogramování jednotlivých služeb a konečně také ověření samotné implementace. Tu je třeba porovnat vůči požadavkům a ověřit tak jejich naplnění.

1.2 Cílová skupina

První cílovou skupinou je vývojář, řešící problém plnotextového vyhledávání produktů při implementaci elektronického obchodu. Takovému čtenáři by měla práce poskytnout dostatečné teoretické znalosti potřebné pro implementaci vyhledávání.

Užitečná také může být konkrétní podoba implementace, která je v této práci popisována.

Další možnou cílovou skupinou je provozovatel elektronického obchodu, který přemýšlí, jakým způsobem zlepšit (případně zavést) plnotextové vyhledávání. V této práci získá přehled o složitosti samotné implementace, poskytne mu komentovaný soupis možných řešení a konečně také poskytne funkční službu, kterou může okamžitě na webový portál napojit.

Poslední cílovou skupinou budiž kdokoli, kdo se zajímá o problematiku plnotextového vyhledávání v 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 části popisuji metody použité k naplnění jednotlivých cílů. Pro zkoumání problému oboru provedu analýzu trhu, nabízí se také možnost dotazování potenciálních uživatelů. Pro porozumnění problematice vyhledávání provedu rešerši literatury. Vzhledem k množství dostupné literatury bude třeba provést syntézu těchto informací. Při vytváření návrhu řešení bude použito modelování, výstupem by tedy měl být model řešení.

1.4 Struktura práce

V první části práce popisuji problematiku vyhledávání v prostředí elektronického obchodování. Definuji zde jednak kontext, v kterém se pohybuji a dále také popisuji problémy, které v tomto prostředí existují. Snažím se identifikovat potenciálního uživatele a definovat jeho požadavky na plnotextové vyhledávání. Toto prostředí má svá specifika, která také popisuji a vysvětluji, proč jsem se na tuto oblast zaměřil. 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 teorií plnotextového vyhledávání. Popisuji 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 vyhledávání.

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

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

2 Analýza byznys požadavků na aplikaci

V první části diplomové práce popisují problematiku plnotextového vyhledávání v prostředí elektronických obchodů. Nejprve je čtenář uveden do problematiky elektronického obchodování a obeznámen s základními pojmy a principy. Následně je popsán problém, který je řešen a jsou definovány konkrétní podmínky, za kterých bude výsledný nástroj přínosný a použitelný. Jsou zde také diskutovány stávající možnosti implementace plnotextového vyhledávání a porovnány alternativní služby.

2.1 Vysvětlení základních pojmů

Pro usnadnění orientace v této práci nejprve vysvětlují dále používané pojmy. Jedná se vesměs o známé termíny, přesto považují za důležité uvést jejich význam na pravou míru.

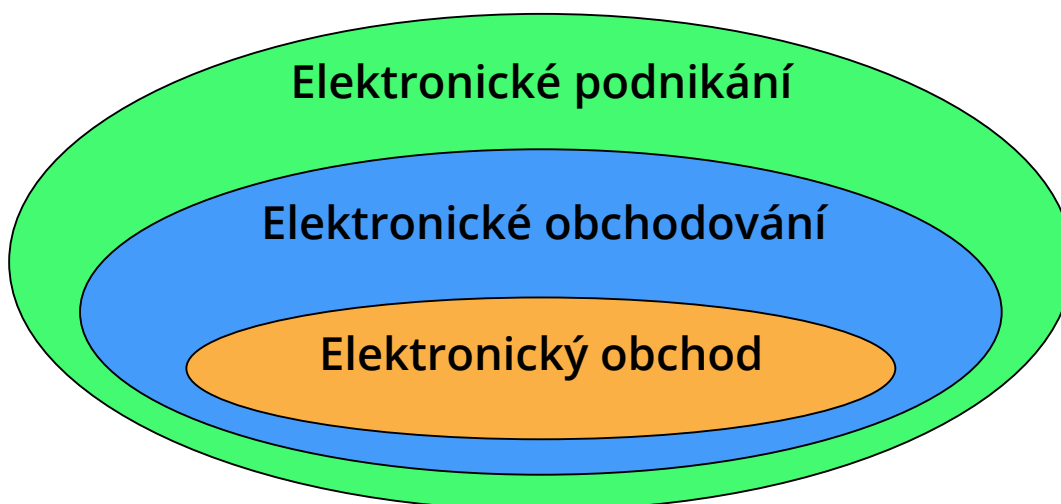
Elektronické obchodování

Termín elektronické obchodování (e-commerce) označuje veškerou činnost spojenou s obchodováním realizovaným prostřednictvím sítě internet [1, strana 11]. Spadá sem distribuce, nákup, prodej, marketing, servis produktů. Elektronické obchodování se dále podle dělí zaměřením na cílové skupiny, z nichž nejpodstatnější jsou B2C (zaměřeno na koncové zákazníky) a B2B (zaměřeno na obchodníky) [1, strana 17]. Zároveň je však elektronické obchodování podmnožinou elektronického podnikání (e-business), které označuje veškeré obchodní a výrobní aktivity, zatímco elektronické obchodování se týká samotného prodeje zboží a služeb.

Elektronický obchod

Elektronický či internetový obchod (e-shop) je webová aplikace, jejímž prostřednictvím provozovatel obchodu prodává zboží nebo služby zákazníkům [1, strana 16]. Jedná se o podmožinu elektronického obchodování, jde tedy o jeden z možných prodejních kanálů v prostředí internetu. Elektronický obchod je specifickou webovou aplikací, v níž se opakují určité vzorce. Konkrétně je to katalog produktů, kategorizace zboží, vkládání položek do košíku, platba objednávky, možnost kontaktu s zákaznickou podporou prostřednictvím on-line chatu přímo v aplikaci.

V elektronickém obchodě můžeme shledávat řadu podobností s obchodem kamenným, který zákazníci navštěvují osobně. Některé vlastnosti elektronického obchodu jsou zřejmě inspirovány obchody kamennými (procházení zboží podle kategorií, přidávání položek do košíku, uplatnění slevových poukázek při platbě). Často však elektronický obchod využívá možností, které webové prostředí nabízí (filtrace a řazení položek podle parametrů, vrácení se k naplněnému košíku, porovnávání zboží napříč obchody).



Obrázek 2.1: Vztah elektronického podnikání, obchodování a podnikání

Zcela zásadní výhodou elektronického obchodu oproti návštěvě obchodu kamenného je možnost vytvoření objednávky z pohodlí domova a její následné doručení zásilkovou službou. Tento přístup především šetří čas strávený cestou do obchodu a čas strávený v obchodě. Například v segmentu nákupu potravin může být výhodou to, že je takový nákup hygieničtější a zboží bude doručeno v lepším stavu díky uzpůsobeným vozům dopravců. Výhody ale vznikají i pro provozovatele - nemusí zřizovat prodejnu včetně jejího vybavení a zaměstnance, mohou efektivněji navrhnout sklad, nebo dokonce využít automatizace při kompletaci objednaného zboží.

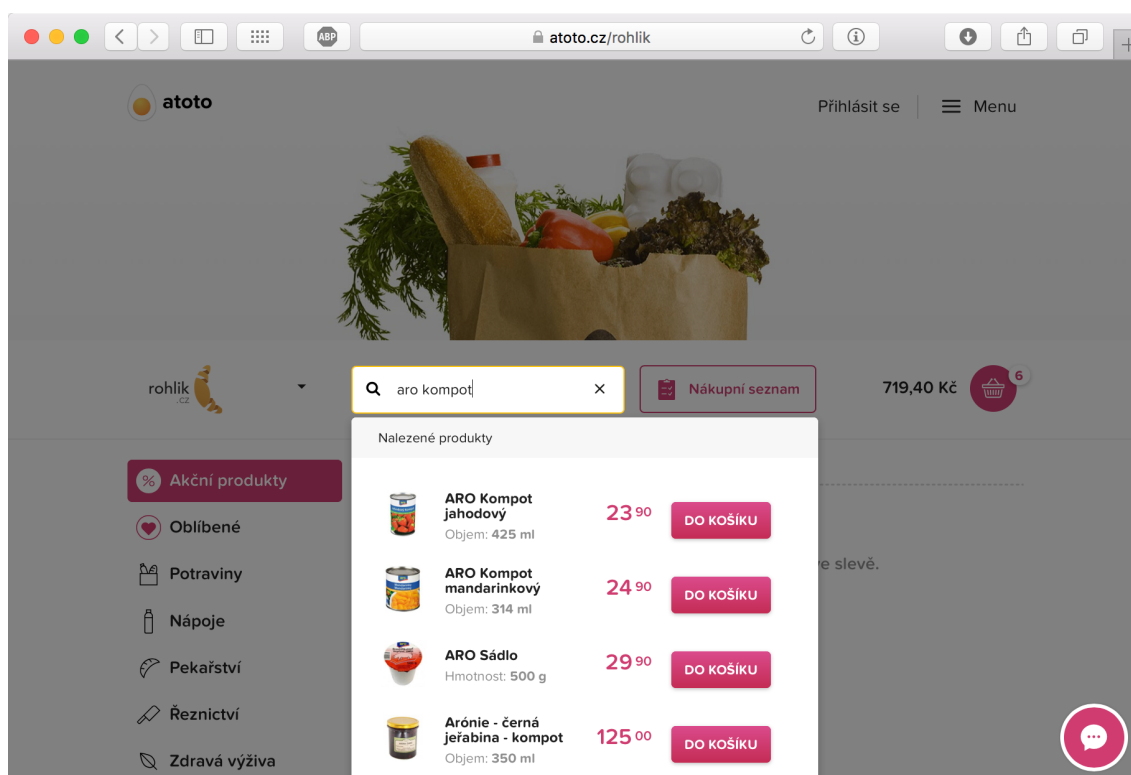
Plnotextové vyhledávání

Plnotextové vyhledávání (full-text search), je způsob vyhledávání v textových datech, kdy se vyhledává uživatelem formulovaný dotaz v invertovaném souboru (někdy také indexovém souboru nebo invertovaném indexu), tedy v souboru obsahující výrazy, podle nichž je daný záznam vyhledatelný [2, strana 15]. Proces, kterým vzniká indexový soubor se nazývá indexace. Tato problematika je podrobněji popisována v druhé kapitole.

2.2 Vysvětlení problému, který je řešen

Plnotextové vyhledávání je rychlým prostředkem nalezení konkrétního produktu v elektronickém obchodě. Typicky je využito zákazníky, kteří jsou na webu s konkrétní potřebou a jsou schopni formulovat výraz, podle kterého produkt hledají. Může jít o název typu produktu, jeho značku, variantu nebo kategorii. Pro takového uživatele je procházení webu procházením kategorií zdlouhavé a právě plnotextové vyhledávání mu může zásadně zrychlit cestu k nalezení hledaného produktu.

Důležitost plnotextového vyhledávání navíc roste s rostoucím počtem produktů a kategorií, v kterých je složité se orientovat. Uživatel tak může vyhledávání využít už pro nalezení kategorie, která se nachází v dlouhém a nepřehledném menu.



Obrázek 2.2: Ukázka plnotextového vyhledávání

Implementace plnotextového vyhledávání však není triviální záležitostí. Vstup uživatele je v přirozeném jazyce, takže je třeba jej odpovídajícím způsobem zpracovat, aby bylo vůbec vyhledávání na webu použitelné. Studium této problematiky může být zdlouhavé a v konečném důsledku drahé. Pokud bude k dispozici hotové řešení, které bude snadno nasaditelné na elektronický obchod a bude umožňovat jednoduché uzpůsobení potřebám konkrétního webu, bude to výrazná úspora jak času, tak peněz provozovatele obchodu.

2.3 Specifika prostředí elektronického obchodování

V prostředí elektronického obchodování existují jisté vzorce které se opakují a odlišují toto prostředí od ostatních. Vyhledávají se zpravidla produkty (případně služby), které jsou zařazeny do určitých kategorií a disponují atributy jako cena, název, kód, popis, url, obrázek, dostupnost a dalšími parametry, které bývají číselné (hmotnost, objem) nebo výčtové (barva, značka).

Z toho vyplývá co a podle čeho se bude vyhledávat. Produkty mívají jednoznačné identifikátory, které jsou navíc standardizovány (EAN, ISBN). Pomocí těchto kódů je možné vyhledat produkty napříč elektronickými obchody nebo je párovat v cenových srovnávacích. Pro vyhledávání je dále důležitý název produktu, případně jeho varianty, která upřesňuje konkrétní verzi produktu. Méně často je nutné vyhledávat v popisu produktu, kde bývá uveden jak slovní popis, tak výpis parametrů.

U produktů dále bývá evidována cena, která může být interně uložena v kombinaci s marží. Obchody operující se zlevněným zbožím pak mívají uvedenou cenu jak před slevou, tak po této slevě, aby zákazník viděl, kolikaprocentní sleva je na produktu. I samotná cena produktu (případně marže prodejce) může hrát svou roli ve vyhledávání.

2.4 Definice požadavků na vyhledávání

Požadavky na samotné vyhledávání jsou ovlivněny prostředím, ve kterém vyhledávání probíhá a uživateli, kteří vyhledávání využívají. Primárně by mělo být možné nalézt daný produkt podle jeho názvu, nebo části názvu. To samé platí pro kód produktu. Produkt by měl být ale vyhledatelný i podle dalších atributů, jako je název kategorie nebo název značky výrobce. Někdy dokonce uživatel nemusí vyhledávat konkrétní produkt, ale jen produkty dané značky nebo kategorie, což by měl systém také umět rozpoznat.

Podoba vyhledávaných výrazů bude zadávána zákazníky v přirozeném jazyce, s čímž by si mělo vyhledávání také poradit. Konkrétně jde o tvarosloví, kdy může uživatel zadat výraz například v jiném pádu, než je uvedeno v názvu produktu. Dále jde o vyrovnání se s chybami, které mohou vzniknout při formulaci vyhledávaného výrazu - překlady nebo pravopisné chyby. V neposlední řadě jde také o vztah slova k jeho významu, kdy může jedno slovo mít více významů (homonymum) nebo naopak více slov může odpovídat jednomu výrazu (synonymum).

Další problematikou při zadávání hledaného výrazu uživatelem, je poskytování relevantních výsledků už ve chvíli, kdy uživatel formuluje dotaz, tedy jej teprve píše. V takovém případě je třeba odhadnout, který výraz chce napsat a tuto informaci využít při zobrazení odpovídajících výsledků.

Ve chvíli, kdy systém zná produkty, které odpovídají hledanému výrazu by měl výsledky poskytovat ve vhodném pořadí, měl by tedy pracovat s relevancí výsledků vzhledem k zadanému výrazu. Toto může být velmi obtížné, protože každý uživatel má zájem o jiné produkty a tak pro něj mohou být ro jeden výraz relevantní jiné produkty, než pro někoho jiného. Dále do tohoto pořadí můžou vstupovat požadavky

provozovatele elektronického obchodu v případě, kdy má specifické požadavky na zboží, které chce nabízet přednostně. Důvodů pro takové chování může být více, může souviset s marží konkrétních produktů, nebo s filozofií samotného obchodu.

2.5 Technické požadavky na aplikaci

Na aplikaci je kladeno několik technických požadavků. Nejedná se o funkční požadavky, nýbrž o požadavky související s použitelností, výkonností, spolehlivostí a podporou. Naplnění těchto požadavků zajistí následný bezproblémový chod aplikace a uspokojení potřeb zákazníka.

Prvním technickým požadavkem na aplikaci je její rychlost. Ta je důležitá pro spokojenost koncového zákazníka, který provádí vyhledávání. Všechny části aplikace nejsou z hlediska rychlosti tak kritické, jedná se primárně o samotné vyhledávání. Akceptovatelná rychlost API v tomto případě je 100 ms, přičemž čím bude rychlost vyšší, tím lépe [13]. Rychlost by neměla výrazně ovlivňovat množství současných požadavků. Aplikace by měla být navržena tak, aby ji bylo možné snadno škálovat při zvyšující se zátěži.

Dalším požadavkem je formát samotného API. Bude třeba použít takový formát, který bude co nejsnadněji implementovaný v používaných webových technologiích pokud možno bez instalace jakýchkoli rozšíření. To je důležité pro rychlé nasazení aplikace do provozu a odbourání možných vstupních bariér při rozhodování o využití nástroje. Pro načtení produktů z elektronických obchodů by bylo ideální využít stávajících exportů pro jiné systémy, kterými by obchody mohly implementovat. Pokud by nic takového neexistovalo, musí být vytvoření export produktů pro provozovatele elektronického obchodu co nejsnazší.

Posledním technickým požadavkem je dokumentace samotného API. Dokumentace musí být dobře pochopitelná, tedy psaná přehledně, ideálně s konkrétními ukázkami použití. Výhodou je také možnost vyzkoušet si práci s API vůči testovacímu prostředí, kde nebude riziko vzniku chyb na produkčních datech.

2.6 Popis oborů, kterých se práce dotýká

Samotné řešení problematiky vyhledávání je výrazně interdisciplinární obor. Je nutné mít značné povědomí o získávání informací, počítačovém zpracování přirozeného jazyka, což souvisí jak s počítačovou lingvistikou, tak úzce s teorií formálních jazyků a překladačů. Existuje také vztah mezi lingvistikou a logikou, jakožto vědě o myšlení, která se odehrává v kategoriích lidského jazyka. Podobný vztah lze nalézt také s umělou inteligencí, vzhledem k tomu, že používání přirozeného jazyka je inteligentní činností. Díky tomu, že probíhá ukládání dat, je důležitá znalost teorie datových modelů. Vzhledem k množství ukládaných dat se také můžeme dotýkat oboru Big Data. V neposlední řadě je třeba porozumět potřebám potenciálních uživatelů, tedy mít určitou znalost podnikání a obchodování.

2.7 Stávající možnosti implementace vyhledávání

Na trhu existuje několik služeb, které umožňují implementaci plnotextového vyhledávání. Ty se však liší funkcí, obtížností implementace i cenou. Níže popisují nejvýznamější z nich zejména vzhledem k požadavkům na aplikaci.

Využití relační databáze

První možností, jak vyhledávání implementovat je využití stávající databáze, kterou elektronický obchod disponuje. Ať už se jedná o open source (MySQL, Postgres) nebo komerční (SQL Server, Oracle) databázi, možnosti plnotextového vyhledávání jsou zde omezené. Výhodou je to, že jsou data stále v jednom úložišti, není třeba řešit správu (instalaci, konfiguraci nebo zálohování) dalšího nástroje. Ani vývojáři se nemusí učit nic nového, pouze využijí stávající databázi. Vyhledávání zde probíhá v nejjednodušší formě pomocí operátoru `LIKE`. Některé databázové systémy mají disponují pokročilejšími funkcemi, kterými lze vyhledávání zlepšit [14].

Elasticsearch, Solr, Sphinx

Využitím nástrojů přímo určených pro implementaci plnotextového vyhledávání lze dosáhnout nejlepších výsledků, ať už se jedná o Elasticsearch či Apache Solr využívající Apache Lucene [15], nebo Sphinx. Jde o typický krok v okamžiku, kdy přestává funkčnost relační databáze pro potřeby vyhledávání dostávat. Tyto nástroje umožňují pokročilé nastavení indexace a zároveň jsou dostatečně rychlé, aby byly schopny provádět vyhledávání v řádu milisekund. Zřejmou nevýhodou tohoto řešení je nutnost znalosti dalšího nástroje, jeho správa a řešení synchronizace se stávající databází.

Algolia

Algolia je pokročilý nástroj umožňující implementaci plnotextového vyhledávání [9]. Zaměřuje se na poskytování kvalitních výsledků v co nejkratším možném čase (v řádu milisekund). Snaží se usnadnit práci programátorům tím, že nabízí řadu připravených integrací pro konkrétní programovací jazyky a frameworky. Kromě výsledků vyhledávání umožňuje faceting, tedy dodání dat pro tvorbu filtrů, podporuje řadu jazyků nebo vyhledávání podle geografické lokality. V neposlední řadě je k dispozici dashboard zobrazující stav systému a statistiky vyhledávání.

Pokud si zákazník vystačí s 10 000 vyhledatelnými produkty, může nástroj Algolia používat zdarma, pouze je povinné uvést ve výsledcích vyhledávání logo firmy. Placené verze začínají na 59 dolarech za měsíc za 100 000 produktů.

Swifttype Site Search

Swifttype Site Search je nástroj podobný nástroji Algolia, zaměřuje se však více na uživatelské rozhraní a celkově na snadnost použití [10]. Mezi jeho hlavní funkce pat-

ří vyhledávání v okamžiku formulace hledaného výrazu, možnost filtrace výsledků, dodatečná úprava pořadí výsledků ručním zásahem na úrovni konkrétních výsledků, nebo na základě některého z atributů produktu. V neposlední řadě je také k dispozici přehledná analytika proběhlého vyhledávání. Cena za provoz služby začíná na 299 USD měsíčně.

AWS CloudSearch

Firma Amazon poskytuje v rámci svého cloudu službu CloudSearch [11]. Jejími přednostmi jsou především vysoká výkonnost a škálovatelnost. Velkou vstupní bariérou je však její počáteční složitost. Uživatel musí nejprve proniknout do některých základních konceptů Amazon Web Services, až poté může začít pracovat se samotným vyhledáváním. Co se funkčnosti týče, nabízí služba podporu 34 jazyků, možnost nastavení váhy jednotlivých atributů, nebo doplňování textu během psaní. Cena záleží na využití výpočetní kapacity, její určení je tedy složitější.

Google Custom Search Engine

Google Custom Search Engine [12] se nejvíc odlišuje od ostatních služeb. Vyhledávání je plně řízeno algoritmem společnosti Google a také výsledky vyhledávání vypadají obdobně, jako výsledky vyhledávání na www.google.com. Odlišné je i samotné napojení na službu - výsledky vyhledávání jsou na web vloženy jako samostatná stránka, na které je vidět logo společnosti Google. Základní varianta je však zdarma, je to tedy levná cesta, jak rychle zprovoznit vyhledávání na webu. Pokročilejší varianta umožňující konfiguraci stojí 100 USD ročně.

Porovnání

tabulka...

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] SUCHÁNEK Petr. *E-commerce*.
Vyd. 1. Praha: Ekopress, s.r.o., 2012 144 s. ISBN 978-80-86929-84-2.
- [2] STROSSA, Petr. *Počítačové zpracování přirozeného jazyka*.
Vyd. 1. Praha: Oeconomica, 2011 316 s. ISBN 978-80-245-1777-3.
- [3] 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.
- [4] AGGARWAL Charu C., ZHAI ChengXiang. *Mining Text Data*.
Springer New York Dordrecht Heidelberg London, 2000 522 s. ISBN 978-1-4614-3222-7.
- [5] STURGEON Phil. *Build APIs You Won't Hate*.
Vyd. 1. Philip J. Sturgeon, 2015 188 s. ISBN 978-0692232699.
- [6] KENNEDY William. *Go in Action*.
Manning Publications Co, 2016 264 s. ISBN 978-1-6172-9178-4.
- [7] CLINTON Gormley, ZACHARY Tong. *Elasticsearch: The Definitive Guide*.
O'Reilly Media, 2015 724 s. ISBN 978-1-4493-5854-9.
- [8] Elasticsearch. *Elasticsearch Reference* [online].
2017 [cit. 2017-03-03]. Dostupné z:
<https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>
- [9] Algolia. *Hosted cloud search as a service* [online].
2017 [cit. 2017-03-03]. Dostupné z:
<https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>
- [10] Swiftype. *Site Search by Swiftype* [online].
2017 [cit. 2017-03-30]. Dostupné z: <https://swiftype.com/site-search>
- [11] AWS. *Amazon CloudSearch* [online].
2017 [cit. 2017-03-30]. Dostupné z: <https://aws.amazon.com/cloudsearch/>

- [12] Google. *Custom Search Engine* [online].
2017 [cit. 2017-03-30]. Dostupné z: <https://cse.google.com/cse/>
- [13] GigaSpaces Blog. *Amazon found every 100ms of latency cost them 1% in sales* [online].
2017 [cit. 2017-03-26]. Dostupné z:
<https://blog.gigaspaces.com/amazon-found-every-100ms-of-latency-cost-them>
- [14] PostgreSQL: Documentation. *Full Text Search* [online].
2017 [cit. 2017-03-26]. Dostupné z: <https://www.postgresql.org/docs/9.5/static/textsearch.html>
- [15] The Apache Software Foundation. *Apache Lucene* [online].
2017 [cit. 2017-03-26]. Dostupné z: <https://lucene.apache.org>

Rejstřík

...

A Obsah přiloženého DVD

- Soubor Diplomova_prace_2017_Ludek_Vesely.pdf
 - Text diplomové práce