

Mendelova univerzita v Brně
Provozně ekonomická fakulta

Vylepšenie fulltextového vyhľadávacieho algoritmu pre elektronické obchody

Diplomová práca

Vedúci práce:
Doc. Ing. František Dařena, Ph.D.

Bc. Dominika Jančovičová

Brno 2016

Rada by som poďakovala svojej rodine za zázemie a svojmu priateľovi za podporu a trpezlivosť počas celej doby štúdia. Tiež chcem poďakovať vedúcemu mojej diplomovej práce, pánovi Doc. Ing. Františkovi Dařenovi, Ph.D., za vedenie, cenné pripomienky a trpezlivosť pri písaní tejto práce. Ďalej ďakujem pánovi Ing. Romanovi Malovi, Ph.D. za poskytnuté dáta a odborné konzultácie.

Čestné prehlásenie

Prehlasujem, že som túto prácu: **Vylepšenie fulltextového vyhľadávacieho algoritmu pre elektronické obchody**

vypracovala samostatne a všetky použité pramene a informácie sú uvedené v zozname použitej literatúry. Súhlasím, aby moja práca bola zverejnená v súlade s § 47b zákona č. 111/1998 Sb., o vysokých školách v znení neskorších predpisov, a v súlade s platnou *Smernicou o zverejňovaní vysokoškolských záverečných prác.*

Som si vedomá, že sa na moju prácu vzťahuje zákon č. 121/2000 Sb., autorský zákon, a že Mendelova univerzita v Brně má právo na uzavretie licenčnej zmluvy a použitie tejto práce ako školského diela podľa § 60 odst. 1 Autorského zákona.

Ďalej sa zaväzujem, že pred spísaním licenčnej zmluvy o využití diela inou osobou (subjektom) si vyžiadam písomné stanovisko univerzity o tom, že predmetná licenčná zmluva nie je v rozpore s oprávnenými záujmami univerzity, a zaväzujem sa uhradiť prípadný príspevok na úhradu nákladov spojených s vznikom diela, a to až do ich skutočnej výšky.

V Brne dňa 4.1.2016

.....

Abstract

Jančovičová, D. The improvement of the fulltext searching algorithm for the electronic shops. Diploma thesis. Brno, 2016.

The diploma thesis is dedicated to research, suggestion of solution and implementation of innovating algorithm of the fulltext search in the web pages of an electronic shop. The theoretical part deals with the searching in the e-shops. Customer's search behavior is described, further information about the field of web analysis is given and some of the characteristics of fulltext search engines are offered. The practical part is concerned with the analysis of the visits of the web shop, the assessment of current situation of searching, the determination of customer's preferences. This is the base for formulating business rules and improving search algorithm. Finally, the thesis deals with the implementation in scripting language PHP.

Keywords: Fulltext, searching, questionnaire, web analytics, PHP.

Abstrakt

Jančovičová, D. Vylepšenie fulltextového vyhľadávacieho algoritmu pre elektronické obchody. Diplomová práca. Brno, 2016.

Diplomová práca sa venuje výskumu, návrhu riešenia a implementácií algoritmu vylepšujúceho vyhľadávanie na webovej stránke elektronického obchodu. Práca obsahuje všeobecnú časť o vyhľadávaní v e-shopoch a popisuje správanie užívateľov pri vyhľadávaní. Ďalej približuje problematiku webovej analytiky a charakteristiku niektorých nástrojov fulltextového vyhľadávania. Praktická časť sa zaoberá analýzou prístupov zákazníkov do webového obchodu, zhodnotením súčasného stavu vyhľadávania a zisťovaníu preferencií zákazníkov pri vyhľadávaní. Nakoniec sa zaoberá formuláciou pravidiel pre vylepšenie vyhľadávania a ich následnou implementáciou v jazyku PHP.

Kľúčové slová: Fulltext, vyhľadávanie, dotazník, webová analytika, PHP.

Obsah

1	Úvod a cieľ práce	14
1.1	Úvod	14
1.2	Cieľ práce	15
2	Teoretická časť riešenej problematiky	16
2.1	Vysvetlenie dôležitých pojmov	16
2.2	Problematika vyhľadávania v e-shopoch	18
2.3	Správanie užívateľov	20
2.4	Webová analytika	22
2.5	Analýza nástrojov fulltextového vyhľadávania	25
2.5.1	Elasticsearch	27
2.5.2	MySQL	27
2.5.3	Sphinx	27
3	Metodika riešenia	30
4	Analýza súčasného stavu	32
4.1	Správanie zákazníkov	32
4.2	Aktuálne využívané vyhľadávacie nástroje	36
4.3	Zostavenie dotazníka a analýza výsledkov	39
5	Návrh riešenia a implementácia	43
5.1	Špecifikácia obchodných pravidiel pre vyhľadávacie nástroje	43
5.2	Implementácia	45
5.2.1	Prostredie, príprava a kódovanie	45
5.2.2	Základný princíp optimalizovaného algoritmu	47
5.2.3	Vyhľadávacie módy a možnosti usporiadania	49
5.2.4	Frázy obsahujúce preklepy	52
6	Testovanie	55
7	Diskusia	57
7.1	Zhodnotenie dosiahnutých výsledkov a cieľov	57
7.2	Ďalšie možnosti rozšírenia práce	58
8	Záver	59
9	Literatúra	60
	Prílohy	63
A	Vyhľadávacie termíny Fantazie	64
B	Dátové CD	65

Zoznam obrázkov

Obr. 1: Zhlukovanie, zdroj: web zajo.sk	18
Obr. 2: Systém kategórií, zdroj: web alza.sk	19
Obr. 3: Základné KPI vyjadrujúce vyhľadávanie	32
Obr. 4: Výsledok neúspešného vyhľadávania	37
Obr. 5: Vyhľadávanie termínu Truhlík	40
Obr. 6: Predvolené poradie produktov podľa respondentov	41
Obr. 7: Predvolené usporiadanie podľa respondentov	41
Obr. 8: Vyhľadávanie termínu Fantazie fuchsie	42
Obr. 9: Vyhľadávanie termínu Truhlík 40	42
Obr. 10: Vývojový diagram popisujúci základné fungovanie algoritmu	48
Obr. 11: Vyhľadávanie termínu Fantazie	64

Zoznam tabuliek

Tabuľka 1: Rozdiel medzi zákazníkmi podľa využitia možností interného vyhľadávania	33
Tabuľka 2: Konverzie nákupu podľa využitia vyhľadávania	33
Tabuľka 3: Prehľad vyhľadávacích fráz v internom vyhľadávaní TOP 10	35
Tabuľka 4: Ukazovateľ hodnota za jedno vyhľadávanie TOP 10	36
Tabuľka 5: Testované výrazy a počty ich výsledkov	56

1 Úvod a cieľ práce

1.1 Úvod

V dobe nezastaviteľného pokroku informačných technológií, vzrastajúcej populácie a s tým súvisiacej veľkej konkurencii v podnikaní záleží stále na viac faktoroch a musí sa klásť dôraz aj na maličkosti. Nie je tomu inak ani na webe. Nezáleží na tom, či je reč o stránkach s filmovými databázami, podnikovými stránkami či rôznorodými obchodmi, ktoré majú svoj základ v online podnikaní. Na to, aby prilákali návštevníkov, prípadne aby si ich udržali a vzbudili v nich záujem sa vracieť, je nutné isté prispôsobovanie sa trhu.

Niekoľko rokov dozadu webové stránky slúžili na poskytovanie základných informácií o určitom subjekte. Mali jednoduchý vzhľad, niekoľko riadkov či odstavcov textu, vizuálne odlíšený hlavný panel s ponukou. Samozrejme doba pokročila a dnes takéto stránky považujeme maximálne za stredoškolský projekt študentov. To si však mnoho spoločností neuvedomuje. Nevidia, že už to nie je len pre informačné účely, ale podľa kvality stránky mnoho ľudí posudzuje aj kvalitu služieb či výrobkov, je to ich zrkadlo, vizitka a mocný marketingový nástroj.

Väčšina ľudí presedí viac ako tretinu dňa v práci, má svoje záľuby a rodinu a konštatuje, že nemá čas. Nakupovanie cez internet sa stalo veľmi obľúbeným, šetrí čas aj finančné prostriedky a keby sme zašli do extrémov aj životné prostredie. Najšť však obchod, ktorý spĺňa všetky predstavy všetkých zákazníkov je nereálne. Nie nadarmo sa hovorí, koľko ľudí toľko chutí. Jedným z mnohých spoločných prvkov bude určite snaha o čo najväčšie zjednodušenie pre návštevníka, ktorého chceme premeniť na zákazníka. Na to potrebujeme, aby všetko do seba zapadalo. Kvalitné produkty za adekvátnu cenu sú pre zákazníka základ. Dôležité je ukázať mu konkurenčnú výhodu práve toho nášho obchodu, čo môže nákupom získať a cestu k tomuto cieľu mu čo najviac spríjemniť.

Od výberu tejto témy som si začala viac všimáť prevedenie webových stránok a ich možnosti vyhľadávania, ak to vôbec umožňujú. Web obsahuje veľa textu, množstvo kariet, súbory na stiahnutie. Hlavná ponuka je tak využívaná v kombinácii s vyhľadávaním, nie vždy je však táto možnosť dostatočne funkčná. Pri dobre prepracovanej hlavnej ponuke som schopná nájsť čo potrebujem aj bez vyhľadávacieho poľa. Ak však niektorý produkt zadám do vyhľadávania, nie vždy sú výsledky nápomocné. Občas je množina výsledkov prázdna, inokedy sú neprehľadné a skôr vo výnimočných prípadoch sú uspokojivé. Ako je možné, že sama nájdem čo hľadám, ale výsledky vyhľadávania nie? Ba čo viac, ako je možné, že najznámejší vyhľadávač mi naprieč celému webu nájde čo hľadám na stránke, ktorá mi to isté nájsť nevie? Jedným z problémov by mohol byť zvyk užívateľov zadávať kľúčové slová bez diakritiky, pričom vyhľadávanie s tým nepočíta. Taktiež užívatelia môžu produkt hľadať pomocou viacerých slov, ktoré ho popisujú, ale nepatria bezprostredne k sebe. Týmto a ďalším dôvodom sa bude zaoberať táto práca. Bude zameraná na odstránenie problémov, ktoré spôsobujú zobrazenie prázdnej množiny výsledkov, aby zákazník dostal čo najlepšiu spätnú väzbu na svoj dotaz.

Takáto úloha sa nedá uskutočniť iba na základe myšlienok a nápadov zopár zainteresovaných osôb, ale vyžaduje dlhodobý prístup založený na sledovaní správania ľudí za rozličných okolností. Táto diplomová práca sa zameriava na získanie poznatkov o nákupnom správaní zákazníkov na internetovom obchode v súvislosti s interným vyhľadávaním medzi ponúkanými produktami a ich detailnom popise. Dotazníkovým prieskumom a sle-

dovaním nasadenej analytickej služby budú získané informácie, ktoré sa stanú podkladom pre zmenu algoritmu vyhľadávania.

1.2 Cieľ práce

Cieľom práce je vylepšenie vyhľadávacieho algoritmu na stránke elektronického obchodu, ktorý využíva vlastné vyhľadávanie pomocou fulltextového nástroja. Vylepšenie spočíva v úprave algoritmu tak, aby nájdené výsledky lepšie uspokojovali potreby zákazníkov, boli relevantné voči zadávaným kľúčovým slovám a ich výsledné usporiadanie podporovalo predajnosť produktov. Na zistenie užívateľských preferencií bude využitý prostriedok marketingového výskumu – dotazníkový prieskum v kombinácii s analýzou súčasného správania existujúcich zákazníkov podľa nástroja webovej analytiky, ktorý tieto aktivity sleduje a zaznamenáva. Po preskúmaní možností súčasného vyhľadávania bude pristúpené k definovaniu pravidiel, ktorými sa toto vyhľadávanie bude riadiť. Pravidlá zistené z prevádzky konkrétneho elektronického obchodu budú stanovené vo všeobecnej rovine z dôvodu ich aplikácie na rôzne e-shopy. Každý e-shop umožňuje prezeranie produktov pomocou navigačnej lišty, ktoré sú z pravidla rozdelené na kategórie. Pre rýchlejšiu orientáciu slúži fulltextové vyhľadávanie. Na plné využitie jeho funkčnosti však treba, aby poskytovalo prístupné výsledky. To bude dosiahnuté vďaka eliminácií prázdnej množiny výsledkov. Ďalej bude vyhľadávanie optimalizované tak, aby bolo čo najprívetivejšie pre užívateľa. To sú výhody ako napr. korekcia chýb v zadanom slove či nezohľadňovanie diakritiky. Po definícii pravidiel vyhľadávania bude pristúpené k implementácii a testovaniu navrhovaného riešenia.

2 Teoretická časť riešenej problematiky

Pod pojmom elektronický obchod, inak aj e-shop, internetový obchod alebo online obchod sa podľa portálu BusinessInfo.cz vo všeobecnosti rozumie podnikanie prostredníctvom elektronických prostriedkov. To zahŕňa obchodovanie s tovarom alebo službami, rovnako dôležité sú však aj súvisiace kroky spojené s reklamou, uzatváraním zmlúv a plnením ich obsahu vrátane popredajnej podpory a služieb, sprostredkovanie platieb či poskytovaním dodatočných informácií o výrobkoch či riešenie reklamácií. (BusinessInfo.cz, 2015)

E-shop je webová aplikácia špeciálne vytvorená k sprostredkovaniu obchodných transakcií na internete ľubovoľného charakteru – B2B (Business To Business), teda podnikateľ podnikateľovi alebo B2C (Business To Customer), podnikateľ zákazníkovi ako konečnému spotrebiteľovi. Základom je podľa Ecksteinovej katalóg tovaru obvykle radený do kategórií, viac ale tieto možnosti budú popísané v kapitole 2.2. Ich súčasťou je tiež možnosť vyhľadať tovar na základe rôznych parametrov, čo umožní zrýchlenie procesu nákupu a zlepšenie orientácie. (Ecksteinová, 2010)

V dobe, kedy je dostupné nespočetné množstvo elektronických obchodov je potrebné sa zamyslieť nad ich optimalizáciou. Napríklad optimalizované vyhľadávanie produktov môže pomôcť k jeho zvýhodneniu oproti konkurencií. Aby bolo možné vykonať vylepšenie, je vhodné zistiť úspešnosť aktuálneho vyhľadávania. Na stanovenie tejto úspešnosti sa používajú pojmy uvedené v kapitole 2.1. Podrobnostiam o vyhľadávaní a prostriedkom pre jeho optimalizáciu sa venuje táto záverečná práca v celom svojom obsahu.

2.1 Vysvetlenie dôležitých pojmov

Táto práca sa zaoberá tematikou z viacerých oblastí, preto je vhodné na začiatok uviesť zopár pojmov, ktoré sa v nej budú opakovať. Čitateľ bude oboznámený s niektorými možno neznámymi pojmami a štúdium ďalšieho textu tak bude menej komplikované.

Vyhľadávanie

Všeobecný pojem vyhľadávania netreba nijak zvlášť vysvetľovať, pre potreby webovej problematiky sa však tento termín špecifikuje podľa miesta a obsahu výsledkov vyhľadávania na:

- **Organické**
Organické výsledky sú vyhľadávačom vrátené po indexovaní stránok na základe ich obsahu a relevancie kľúčových slov. Pre daný web je oveľa cennejšie byť uvedený v zozname prirodzených výsledkov ako v inzerátoch PPC (sponzorovaných odkazoch, viď. ďalší text), pretože návštevníci webu majú sklony viac dôverovať týmto výsledkom ako plateným. Väčšina spotrebiteľov má pocit, že tieto výsledky skôr odrážajú potreby toho, čo by mohli chcieť. (Barcelona Creative Group, 2015)
- **Sponzorované**
Platené vyhľadávanie je spoplatnená služba, ktorú poskytujú vyhľadávače pre zviditeľnenie organizácií alebo ich webu. Platiaci zákazník si určí, ktoré kľúčové slová zadane do vyhľadávača majú zobraziť reklamu na jeho web. Výsledky plateného vyhľadávania sa objavujú nad alebo vpravo od organických výsledkov. Inzerenti súťažajú prostredníctvom cenových ponúk ponúkaných za určité kľúčové slovo, a v závislosti

na výške ponuky a celkovej hodnote slova sa vyhľadávač rozhodne o umiestnení a poradí uvažovaného inzerátu (webu). Poradie je zväčša určené sumou, ktorú je inzerent ochotný zaplatiť za klik (tzv. PPC – pay per click). Cena sa mení s konkurencieschopnosťou zvoleného slova. (Barcelona Creative Group, 2015)

- **Interné**

Interné vyhľadávanie je funkcionalita ponúkaná na webových stránkach, ktorá návštevníkom umožní jednoduchšie prehliadanie obsahu v prípade, že majú záujem túto službu využiť. Dalo by sa povedať, že opakom je externé vyhľadávanie. Externé teda znamená vyhľadávanie cez vyhľadávač v celej sieti internetu, a interné je integrované v rámci daného webu. (Yahoo! Help, 2010)

Metrika

Pri použití termínu metrika sa odkazujeme na priamu číselnú hodnotu, ktorá reprezentuje časť dát vo vzťahu k jednej alebo viacerým dimenziám. Príkladom môže byť hrubý predaj za týždeň. Mierou je predaj v peňažných jednotkách, a dimenziou je čas, teda týždeň. Ten môže byť určený na rôznych hierarchických úrovniach (deň, týždeň, mesiac...). Vytvorenie asociácie medzi mierou a špecifickou hierarchickou úrovňou tvorí základ metriky. (Gonzalez, 2015)

Kľúčový ukazovateľ výkonnosti

KPI (Key Performance Indicator, kľúčový ukazovateľ výkonnosti) je metrika, ktorá je viazaná k cieľu. Najčastejšie táto hodnota ukazuje, aká veľká je odchýlka od preddefinovaného cieľa.

Organizácie po celom svete využívajú kľúčové ukazovatele pre zhodnotenie svojej výkonnosti. Ukazovateľ KPI, ktorým sa tiež niekedy hovorí kľúčový ukazovateľ úspechu (anglicky Key Success Indicator, skrátene KSI) alebo vyvážená výsledková listina (anglicky Balanced Score Card, BSC) sa používajú v strategickom riadení na oznámenie o stave podniku. Akonáhle organizácia stanoví svoje činitele cieľov a kľúčových výsledkov, potrebuje najšť spôsob pre meranie svojho postupu a týmito mierami sú práve kľúčové ukazovatele výkonnosti. Zvyčajne sú zobrazované ako pomer aktuálneho stavu k stavu cieľovému a sú dizajnované tak, aby neustále vyjadrovali progres v stanovenom pláne. (Gonzalez, 2015)

Podobne je to aj v oblasti webovej analytiky, kde tieto kľúčové ukazovatele výkonnosti predstavujú webovú metriku, ktorá je podstatná pre úspech organizácie na Internete. Dôraz je kladený na slovo podstatná, pretože každá zmena ukazovateľa KPI by mala vyvolať záujem zo strany organizácie, keďže dobre určené ukazovatele KPI vytvárajú očakávania a spôsobujú akcie. (Clifton, 2009) Podľa Petersona akékoľvek KPI, ktoré keď sa náhle a nečakane zmení a neinšpiruje nikoho, aby poslal email, zodvihol telefón alebo rýchlo nevyhľadal pomoc, nie je hodné reportovania. (Peterson, 2006) Pri tvorbe KPI je dôležité, že:

- vo väčšine prípadov nie je ukazovateľom KPI samotné číslo, ale nejaký pomer, percentuálny podiel alebo priemer, vďaka čomu je možné dáta prezentovať v určitom kontexte;
- ukazovateľ KPI musí byť časovo zviazaný, to zdôrazňuje zmenu a jej rýchlosť;

- ukazovateľ KPI je hlavným indikátorom akcií kritických z hľadiska obchodu. KPI sa vždy prezentujú v určitom kontexte, aby hodnota týchto ukazovateľov mala vypovedaciu schopnosť.

Kľúčové slová a frázy

Návštevníci využívajú tieto slová pri akomkoľvek vyhľadávaní, či už ide o produkty, služby alebo informácie. Kľúčovými sa nazývajú z dôvodu svojho významu pre zvýšenie potenciálnej prevádzky na webe. (Sostre, Leclair, c2007)

2.2 Problematika vyhľadávania v e-shopoch

Užívatelia používajúci vyhľadávanie často vyjadrujú záujem o prostredie, ktoré usporadúva výsledky do významových skupín. To napomáha porozumieť výsledkom a rozhodnúť sa, akú akciu vykonať ako nasledujúcu. Dlhodobá štúdia, v ktorej bola užívateľom poskytnutá možnosť zobrazovania výsledkov podľa skupín odhalila, že ľudia zmenili svoje vyhľadávacie návyky v nadväznosti na dostupný mechanizmus tvorby zoskupení.

Vyhľadávanie: bundy

Brand	Model	Original Price (€)	Sale Price (€)	Discount (%)
POLARTEC	Ferrara Lady JKT Egret	89,90	39,90	55%
POLARTEC	Ferrara Lady JKT Orchid	89,90	39,90	55%
POLARTEC	Ferrara Lady JKT Tarragon	89,90	39,90	55%
POLARTEC	Ferrara Lady JKT Grenadine	89,90	39,90	55%
POLARTEC	Vitoria Lady Jkt Deep Orchid	129,90	84,90	34%
POLARTEC	Vitoria Lady Jkt Beet	129,90	84,90	34%
POLARTEC	Vitoria Lady Jkt Grenadine	129,90	84,90	34%
ZERQ therm	Molveno Lady JKT Scarlet	89,90	34,90	61%
ZERQ therm	Molveno Lady JKT Birch	89,90	34,90	61%
ZERQ therm	Molveno Lady JKT Violet	89,90	34,90	61%
ZERQ therm	Molveno Lady JKT Lime	89,90	34,90	61%
ZERQ therm	Molveno Lady JKT Egret	89,90	34,90	61%

Obr. 1: Zhlukovanie, zdroj: web zajo.sk

Zjednodušene povedané, produkty môžu byť rozdelené hierarchicky do jednej kategórie alebo na viac úrovní. Podľa detailnosti zobrazovaných výsledkov užívateľa prispôbili

svoje návyky pri vyhľadávaní a podľa prostredia, ktoré im zobrazuje výsledky sú v niektorých prípadoch nútení bližšie špecifikovať požadovanú položku.

Existujú dve najznámejšie metódy: *systém kategórií*, konkrétne možnosť využívania špecifických filtrov, ktorý umožní priamo zobrazenie iba tých položiek, ktoré sú vyžadované a *zhlukovanie*, ktoré všetky zobrazené položky zoskupí podľa istých pravidiel. (Herst, 1999)

Na obrázku 1 vidno, že tento obchod po zadaní vyhľadávaného textu jednoducho zobrazí všetky vyhovujúce produkty, *zhlukuje* výsledky na jednej úrovni podľa spoločného pravidla – hľadané boli bundy.

Notebooky ASUS

Herné | Profesionálne | Štýlové | Základné
 Dotykové | Pre náročných | Tablet PC | 2 v 1
 Podľa parametrov | Batohy na notebooky | Tašky na notebooky | Puzdrá na notebooky
 Príslušenstvo k notebookom | Spríevodca notebookmi

Značky a parametre | 0,- Od | Do 2 408,-
 Stav tovaru: Všetko | Iba nové | Iba rozbalené, zánovné | Iba na sklade | Kdekoľvek

TOP | Najpredávanejšie | Najdrahšie | Najlacnejšie | 302 položiek

Ľudová voľba

ASUS X553MA-BINGO-SX537B biely
 Notebook - Intel Celeron Dual Core N2840 Bay Trail-M, 15.6" LED 1366x768 lesklý, RAM 4GB, Intel HD Graphics, HDD 500GB 5400 otáčok, DVD, WiFi, Bluetooth 4.0, webkamera, HDMI, USB 3.0, Windows 8.1 64-bit with Bing - zadarmo upgrade na Windows 10

Zlatá stredná cesta

ASUS ZENBOOK UX303LA-R4389H kovový
 Ultrabook - Intel Core i5 5200U Broadwell, 13.3" LED 1920x1080 WV antireflexná, RAM 4GB, Intel HD Graphics 5500, SSD 128GB, WiFi ac, Bluetooth 4.0, HD webkamera, HDMI, mini DisplayPort, Wida, USB 3.0, podsvietená klávesnica, 3 článková Li-pol batéria, puzdro, Windows 8.1 64-bit - zadarmo upgrade na Windows 10

Pre náročných

ASUS ZENBOOK UX303LB-DQ005H kovový
 Ultrabook - Intel Core i7 5500 Broadwell, dotykový 13.3" LED 3200x1800 QHD + WV lesklý, RAM 12GB, NVIDIA GeForce 940mm, SSD 256GB, WiFi ac, Bluetooth 4.0, HD webkamera, HDMI, mini DisplayPort, USB 3.0, podsvietená klávesnica, 3 článková Li-pol batéria, puzdro, Windows 8.1 64-bit - zadarmo upgrade na Windows 10

€274 bez DPH €228,33 | Kúpiť | Na sklade > 5 ks

€789 bez DPH €657,50 | Kúpiť | Na sklade > 5 ks

€1 186 bez DPH €988,33 | Kúpiť | Na sklade > 5 ks

Obr. 2: Systém kategórií, zdroj: web alza.sk

Systém kategórií tvorí množina označení organizovaných spôsobom, ktorý odrzkadľuje podstatu danej oblasti. Na obrázku 2 sú to napríklad kategórie Profesionálne, Štýlové, Herné... Sú vytvárané prevažne manuálne, hoci priradenie dokumentov do kategórií môže byť do istej miery vykonané aj automatizovane. Pri rozdelení na herné a profesionálne by sa podľa určitých parametrov dalo rozhodnúť aj automatizovane, do ktorej kategórie ktorý počí-

tač patrí. Ale napr. pri štýlových je to skôr o subjektívnom rozhodnutí, pričom do tejto kategórie môže spadať počítač z oboch spomínaných kategórií.

Dobré systémy kategórií sú logicky relatívne kompletne, a ich štruktúra je predvídateľná v súlade s výsledkami vyhľadávania pre daný súbor informácií. Najbežnejšie využívané štruktúry kategórií vo všeobecnosti sú: *horizontálne*, *hierarchické* a *zobrazenie pomocou filtrov*. Horizontálne kategórie sú jednoducho zoznamy tém alebo produktov. Tie môžu byť zobrazené pomocou filtrov, v istých skupinách alebo usporiadané podľa rôznych kritérií. Väčšina webových stránok zaraďuje informácie do všeobecných kategórií, pričom výber jednej konkrétnej obmedzí množinu výsledkov, ktorá sa okamžite aktualizuje iba na zvolené informácie či produkty spadajúce do danej kategórie. Niektoré weby zaraďujú výsledky horizontálne a štúdie ukazujú pozitívne výsledky užívateľských reakcií. Niekedy môže byť zložitá zostaviť správnu podmnožinu vhodnú zaradenia do danej kategórie, napr. z dôvodu rozsiahleho webu, v ktorom by tak vzniklo množstvo kategórií, pod kategórií, podmnožín a skupín, pričom by v každej skupine bolo zopár produktov, a web by sa tak stal neprehľadným. Táto možnosť sa tak zdá byť výhodnejšia pre užšie ciele kolekcie dát. (Hearst, 1999)

Vo webovom obchode na obrázku 2 po zadaní kľúčových slov boli vyhľadané všetky notebooky Asus, no poskytuje nám navyše možnosť upresniť svoje požiadavky a jednoduchšie sa preklikať k informáciám či produktom, ktoré hľadám. Hlavná kategória sa týka notebookov, podmnožina ich delí do určitých skupín. Je na nás, či toto delenie chceme využiť alebo sa výsledkami radšej preklikáme sami a využijeme k tomu filtre prípadne ďalšie skupiny, ktoré produkt viac konkretizujú. v tomto prípade je využitá kombinácia vyhľadávania s doplnkovými možnosťami.

2.3 Správanie užívateľov

Všetky hľadania nemusia začínať napísaním kľúčového slova do vyhľadávacieho poľa. Veľa webových stránok poskytuje užívateľom možnosť preštudovať štruktúru informácií ktoré poskytujú, aby mal užívateľ možnosť dostať sa k počiatočnému bodu pre vyhľadávanie. Pojmy ako navigácia a browsing (prezeranie) sa používajú viac menej porovnateľne. Odkazujú na stratégiu, akou užívateľ získava informácie, pohybuje sa po jednotlivých stránkach a vyberá možnosti ďalšieho postupu. V prípade, že navigačné menu uspokojuje jeho potreby, uprednostňuje prezeranie pred vyhľadávaním. Táto stratégia je často preferovaná, pretože je mentálne menej náročná. Rozpoznanie istej informácie je jednoduchšie ako jej udržanie či vyvolanie z pamäte. Ak však hľadanie predmetu záujmu trvá príliš dlho, je časté využitie druhej spomínanej varianty. Prezeranie funguje len tak dlho, ak každé nové zobrazenie posunie užívateľa bližšie k jeho cieľu. Ak sa však niekde cestou nedostane k novej informácii, stáva sa táto skúsenosť frustrujúcou z hľadiska využiteľnosti a siahne po vyhľadávaní.

Výskumníci vyvinuli mnoho teoretických modelov o procese vyhľadávania uskutočňovaného ľuďmi. Klasická predstava o modeli procesu vyhľadávania informácií podľa popisu Sutcliffe and Ennis (Hearst, 1999) je formulovaný ako cyklus skladajúci sa zo štyroch hlavných aktivít:

- identifikácia problému;
- zrozumiteľné vyjadrenie potrieb;

- formulácia požiadavku;
- vyhodnotenie variant.

Tento štandardný model obsahuje na prvý pohľad skrytý predpoklad, že potreby užívateľa sú v danom čase nemenné a proces vyhľadávania je tvorený postupným zdokonaľovaním vyhľadávanej frázy až do vtedy, kým nie sú nadobudnuté relevantné informácie naviazané na pôvodný cieľ. Modernjšie prístupy zdôrazňujú dynamickú podstatu vyhľadávacieho procesu a upozorňujú, že užívatelia sa rýchlo učia a svoje potreby upravujú podľa nájdených výsledkov. Tento spôsob sa niekedy nazýva tiež „model zbierania bobúl“ (berry picking). V dnešnej dobe pri okamžitej odpovedi na dotaz je jednoduché začať vyhľadávanie jedným všeobecným slovom, ktoré sa podľa výsledkov preformuluje a zabezpečí tak postup bližšie k cieľu. Je vysoko pravdepodobné, že užívatelia preformulujú svoj dotaz, pretože je to jednoduchší spôsob ako ho presne špecifikovať hneď na prvý pokus. (Hearst, 1999)

Pri hľadaní pomocou webového vyhľadávača akokoľvek zadané a upresnené kľúčové slová prinášajú výsledky, ktoré sú rozdelené na sponzorované a organické. Užívatelia webových vyhľadávačov v obrovskej miere využívajú výsledky organického vyhľadávania, ktorého hodnoty dosahujú až 94 % z celkového počtu. Vyplýva to z výskumu GroupM UK and Nielsen na základe vzorky 1,4 miliarda vyhľadávaní vykonaných 28 miliónmi obyvateľov Veľkej Británie v júni roku 2011. Tento výskum sa na strane organického vyhľadávania zamerával aj na rozdiel vo vyhľadávaní podľa obchodnej značky, či jej názov bol použitý alebo nie. Vo všeobecnosti sa 68 % užívateľov uspokojilo s poskytnutými výsledkami na prvých troch pozíciách. Po rozdelení na vyhľadávanie so zadanou značkou, názvom firmy a pod. návštevníci využili prvé tri výsledky, ktoré predstavujú 90 %. Na druhej strane užívatelia, ktorí pri vyhľadávaní nepoužili značku, majú vyššie tendencie prehliadať aj výsledky na nižších pozíciách. Prvé tri pozície tvoria iba 61 %. (Goodwin, 2012)

Na strane platených výsledkov výskum odhalil isté demografické špecifikácie o užívateľoch, ktorí majú väčšie sklony k prehliadaniu platených výsledkov. Takými sú ženy, ktoré tvoria 53 %, muži tvoria iba 47 %. Ďalším faktorom je vek – 35 % tvoria užívatelia mladší ako 34 rokov vrátane, pričom platí, že s rastúcim vekom sa zvyšuje aj množstvo užívateľov, ktorí prehliadajú platené výsledky. Takže 65 % kliknutí na platené výsledky pochádza od ľudí starších ako 34 rokov. (Goodwin, 2012)

Iné výskumy boli zamerané na organické výsledky TOP 10, z ktorých vyšli rozdielne percentuálne hodnoty, avšak v princípe boli neustále rovnaké: čím vyššie sa výsledok nachádzal, tým viac návštevníkov na daný odkaz kliklo a naopak. Z toho prirodzene vyplýva, že väčšina užívateľov si nevšimla ani zmenu, že z TOP 10 sa stalo TOP 7, pretože na nižšie priečky vyhľadávania sa dostalo len malé percento z nich. (Goodwin, 2012) Návštevníci sú zaujatí v prospech prvých výsledkov, pretože z dôvodu svojej popularity tieto výsledky prinášajú vyššiu pravdepodobnosť, že splnia ich očakávania. Štúdie tiež ukazujú, že ľudia sú neochotní odhadovať, koľko relevantných informácií našli. Čím majú slabšie vedomosti o danej problematike, tým skôr sa uspokojia s dosiahnutými výsledkami s presvedčením, že našli všetko potrebné. Navyše často končia po niekoľkých pokusoch aj napriek tomu, že by mohli dosiahnuť presnejšie alebo lepšie informácie. (Hearst, 1999)

Vo výskume od Kinshuk Jerath, Liye Ma, and Young-Hoon Park (2014) sa autori zameriavajú na rovnakú problematiku. Využívajú dáta zozbierané za obdobie jedného mesiaca, sústredia sa na 120 kľúčových slov náhodne vybraných spomedzi 1200 slov, ktoré

boli použité počas 1,63 milióna vyhľadávaní. Pridanou hodnotou ich výskumu proti ostatným je zameranie sa na popularitu kľúčových slov, ktoré tvoria rozhodujúci činiteľ v správaní užívateľov po uskutočnení vyhľadávania. Konkrétne platí, že priemerný počet klikov na jedno vyhľadávanie (vrátane platených výsledkov) je väčší pre menej populárne kľúčové slová ako pre tie viac vyhľadávané. Ďalším poznatkom bolo, že užívatelia môžu byť rozdelení do segmentov podľa podstaty ich klikov. To sa dá interpretovať ako rozdielne štádium záujmu spotrebiteľa podľa témy, o ktorej hľadá informácie či podľa produktu, o ktorého kúpu má záujem. Bolo zistené, že menej zainteresovaní spotrebiteľia hľadajú viac populárne a frekventované kľúčové slová a zameriavajú sa viac na organické výsledky, zatiaľ čo tí s vyšším záujmom hľadali skôr špecifické a menej známe kľúčové slová. Pri vyhľadávaní boli vytrvalejší a uskutočnili viac klikov na jednu návštevu a relatívne viac sa zameriavali na sponzorované výsledky. Podľa tohto výskumu pri zisťovaní vzájomných vzťahov kľúčových slov s prítomnosťou obchodnej značky, s názvom predajcu či vzťahom medzi objemom vyhľadávania a dĺžkou kľúčového slova, sa tieto faktory ukazujú ako nepodstatné. Výsledky výskumu teda indikujú, že popularita kľúčových slov je dôležitou charakteristikou pri vyhľadávaní ako značka alebo predajca.

2.4 Webová analytika

Webová analytika je podnikový proces využívaný k neustálemu zlepšovaniu online podnikania. Jedným z najpopulárnejších analytických služieb je Google Analytics, ktorý kvantitatívne meria, čo sa na webovej stránke deje. (Cutroni, 2010) Stručné, ale dôkladné vymedzenie pojmu webová analytika podľa Avinasha Kaushika (Rethink Web Analytics, 2015) je to analýza kvalitatívnych a kvantitatívnych dát z webovej stránky a neustála snaha o zlepšovanie online skúseností potenciálnych a súčasných zákazníkov, ktorá sa premieťa do požadovaných výsledkov.

Táto definícia zahŕňa podľa Cutroniho tri hlavné úlohy, s ktorými sa každé podnikanie musí vyrovnávať, pokiaľ chce robiť webovú analytiku.

- Merať dáta kvantitatívne a kvalitatívne
Webová analytika nie je bez dát možná. Veľa organizácií si neuvedomuje, že je potrebné mať rôzne typy dát na to, aby sa dal pochopiť výkon webových stránok. Nástroje ako Google Analytics, Omniture, WebTrends and Yahoo! Web Analytics generujú kvantitatívne dáta, sledujú mieru preklikov návštevníkov. Tieto dáta poskytujú informácie, odkiaľ návštevníci prichádzajú, čo na stránke vykonávajú – vyjadrujú dianie na webovej stránke. Pre dôkladnú analýzu je však potrebné zbierať aj kvalitatívne dáta, ktoré hovoria o dôvodoch, prečo sa návštevníci správajú tak, ako to vidíme vďaka kvantitatívnym dátam. Tieto kvalitatívne dáta pochádzajú z rôznych zdrojov ako napríklad z rozhovorov alebo prieskumov. Ak návštevníci odpovedia na jednoduché otázky, má to veľký význam pre pochopenie ich potrieb a poskytnú nám informácie, ako im návštevu spríjemniť.
- Neprestajne zlepšovať web
Všetky dáta a analýzy musia byť v procese neustáleho zlepšovania. To je hlavná podstata webovej analytiky – vykonávať určité akcie nad získanými dátami s istým cieľom. Veľmi zjednodušenou reprezentáciou procesu webovej analýzy je nekonečný cyklus merania, analýzy a následnej zmeny. Najhoršia časť je zistiť, čo a ako zmeniť.

Dáta nám väčšinou poskytnú informácie o tom, kde sa nachádza problém, ale nie je z nich jasné, ako to napraviť. Najčastejšie sa vytvoria rôzne riešenia daného problému a testujú sa. Testovanie znamená, že sa potenciálne riešenia postupne zobrazia návštevníkom a následne sa meria, ktoré z nich dosahuje najlepšie výsledky.

- Zlúčiť stratégiu meraní so stratégiou obchodovania
Hlavným cieľom webovej analytiky je zvyšovanie výnosov. Nie je preto podstatné iba meranie návštevnosti, ale aj to, ako sa dianie na webe premietne do výsledkov obchodovania. Každá webová stránka existuje z nejakého dôvodu, a ten sa musí zohľadňovať aj pri sledovaní kľúčových faktorov.

So spomínanými úlohami pomáhajú nástroje webovej analytiky, a asi najznámejším z týchto zástupcov je Google Analytics.

Technológia Google Analytics

Google Analytics (GA) je webová analytická služba od spoločnosti Google zameraná na maloobchodné stránky malých a stredných spoločností. Vďaka tomuto nástroju môžu vlastníci webových stránok získavať štatistické dáta o svojich užívateľoch. GA umožňuje sledovať aktuálnu aj historickú návštevnosť a poskytuje rôzne prehľady dát, z ktorých medzi najdôležitejšie patria:

- Cieľové publikum
Podáva základné informácie o návštevníkoch, z akých miest a krajín pochádzajú, aké zariadenia využívajú, či sa jedná o nových alebo vracajúcich sa zákazníkov, tok návštevníkov – prechod stránkami na základe rôznych filtrovacích kritérií.
- Zdroje návštevnosti
Tiež nazývané ako akvizícia, teda proces získavania alebo nadobudnutia istých aktív alebo cieľov tohto procesu. V prípade GA ide o nadobúdanie návštevníkov z rôznych zdrojov. Tými môžu byť napríklad priamy odkaz na webové stránky, organické vyhľadávanie, platené vyhľadávanie, vďaka sociálnym sieťam alebo prostredníctvom referencií. Je možné zistiť, vďaka akým vyhľadávacím frázam sa návštevníci dostali na web, ktorá bola vstupná stránka, prípadne ktorá sociálna sieť prispieva najväčším počtom návštevníkov.
- Obsah
V aktuálnej verzii GA je táto sekcia nazývaná Správanie. Vyjadruje, ako dobre fungujú jednotlivé stránky webu. Je tu vidieť, cez ktoré stránky sa návštevníci dostávajú na web a cez ktoré stránky ho najčastejšie opúšťajú. Obsahuje aj prehľad o tom, čo je hľadané v internom vyhľadávaní. Tieto informácie tvoria cenný podklad pre vylepšenie webu. Získané poznatky môžu slúžiť ako tipy na to, čo je potrebné zvýrazniť alebo aký sortiment je vhodné pridať.
- Konverzie
V tejto časti sa sledujú dáta spojené so situáciami, kedy návštevník vykoná užitočnú akciu. Okrem typickej objednávky to môže byť registrácia, prihlásenie k odoberaniu noviniek, stiahnutie produktového katalógu a pod. Nachádzajú sa tu napr. aj informácie o tom, ktorý spôsob prilákania zákazníkov prináša najvyššie konverzie.

Táto služba je na trhu od roku 2005, sprístupnená všetkým záujemcom s Google účtom je zdarma od polovice roku 2006. Využíva asynchrónny kód, ktorý je možné umiestniť do záhlavia stránky. Beží na pozadí paralelne spolu s načítaním stránky, takže nespôsobuje problémy s dobou načítania. (Sharif, 2012) V aktuálnej verzii je možnosť vytvárania vlastných reportov, je možná analýza v reálnom čase, grafy zobrazujúce tok užívateľov a prechody medzi stránkami. GA často prechádza zmenami designu, prípadne modifikáciou názvov základných sekcií. Nanešťastie, pretože je služba závislá na cookies, nemôže zhromažďovať dáta od užívateľov, ktorí majú cookies blokované. (Rouse, 2011)

Predmet merania a vyhodnocovania

V každom prehľade sa pre jednotlivé typy dát zobrazujú rôzne typy hodnôt. Medzi najdôležitejšie podľa Ungra patria (Ungr, 2015):

- **Návštevy**
Znamenajú, koľko ľudí priviedol nejaký zdroj návštevnosti, kľúčové slovo vo vyhľadávачi alebo referenčná stránka na web. Návštevnosť je dôležitá, ale treba ju správne interpretovať, preto je potrebné ju sledovať vždy v súvislosti s tržbami, konverzným pomerom alebo s mierou opustenia stránok.
- **Miera opustenia**
Tento atribút hovorí o tom, koľko návštevníkov si zobrazilo iba jednu stránku a potom web opustili. Vo väčšine prípadov je miera opustenia na e-shope problém, pretože znamená, že nebola vykonaná objednávka. K tomu treba aspoň dve stránky – detail produktu a nákupný košík. V prípade určitého kľúčového slova, ktoré generuje vysokú mieru opustenia je pravdepodobné, že je toto slovo irelevantné pre svoju vstupnú stránku. Vysoké percento opustenia pri konkrétnych stránkach môže znamenať ich neprehľadnosť alebo to, že ponuka nie je dostatočne zaujímavá. Obvyklá priemerná miera opustenia elektronického obchodu sa pohybuje v rozmedzí medzi 25 – 50 %.
- **Tržby**
Orientačne hovoria, koľko financií priniesol konkrétny zdroj. Pri porovnaní s nákladmi sa dá zistiť výhodnosť daného zdroja. Takýmto príkladom môžu byť kampane v AdWords alebo Skliku. Porovnaním tržieb a nákladov na kredit a prácu sa dá zistiť, či sú kampane výhodné. Obdobne sa dajú vyhodnocovať aj iné marketingové kampane.
- **Priemerná hodnota objednávky**
Hovorí o tom, aká je priemerná hodnota nákupu z daného zdroja. Najviac sa z marketingového hľadiska oplatí investovať do kampaní, ktoré majú najvyšší prísun zákazníkov s vysokými hodnotami nákupov. Naopak pri zdrojoch s prísunom nízkych hodnôt nákupov je potrebné sledovať celý objem a vyhodnotiť, či sa investovať do týchto zdrojov oplatí.
- **Konverzný pomer**
Je to percentuálny ukazovateľ, ktorý hovorí o pomere medzi celkovým počtom návštevníkov a tými, ktorí dokončia konverziu, teda najčastejšie objednávku alebo inú činnosť, ktorá je definovaná ako uspokojivý cieľový stav. Najviac sa oplatí investovať

do zdrojov s vysokým konverzným pomerom, pri opačných hodnotách je potrebné hľadať príčiny problémov.

Funkcionalita

GA je schopný zbierať dáta z niekoľkých rôznych platforiem vďaka využívaniu technológií trackovania. Je to agregatívny informačný systém, ktorý zbiera dáta zo štandardných webových stránok, mobilných webových stránok, aplikácií Adobe Air, iPhone a Android. S postupným vývojom technológií Google progresívne pridal viac metód pre zber dát a zvolil rôzne spôsoby distribúcie obsahu k ľuďom.

2.5 Analýza nástrojov fulltextového vyhľadávania

Fulltextové vyhľadávanie je jedna z techník slúžiaca na vyhľadávanie v elektronických dokumentoch alebo v databázach. Pri hľadaní sa prehľadajú všetky slová v texte a toto komplexné prehľadanie sa nazýva fulltextové vyhľadávanie. Je vhodné pre veľký objem neštruktúrovaného textu, ktorý dokáže prehľadať rýchlo a efektívne. Využívanie tejto technológie umožňuje niekoľko programov, o ktorých bude reč v tejto kapitole. Jedná sa napr. o Sphinx a Elasticsearch.

Pre pochopenie rozdielu medzi obyčajným hľadaním a fulltextovým je vhodný príklad na MySQL databáze. Nasledujúci príkaz nájde všetky záznamy, ktoré v stĺpci *popis* nájdu výraz *beautiful* alebo *programming*.

```
SELECT * FROM posts WHERE description LIKE '%beautiful%' OR
description LIKE '%programming%';
```

Problém sa ale komplikuje, keď chceme vyhľadať viac slov, ktoré sa nenachádzajú v jednom poli alebo bezprostredne za sebou. V ukážkovom príkaze sú to iba dve slová, keď ich bude viac, jeho vykonanie bude pomalšie, pretože každé slovo sa vyhodnocuje individuálne. Tento spôsob taktiež neposkytuje informáciu o relevantnosti hľadaného a nájdeného. Relevancia môže byť definovaná ako miera toho, ako blízko sú vrátené výsledky voči zadávanému výrazu. Je to dôležitá hodnota, pretože užívateľ chce vidieť najpresnejšie výsledky vo vrchných pozíciách.

Po pridaní fulltextového indexu do určitého stĺpca máme k dispozícii MySQL fulltext vyhľadávanie. Príkaz vyzerá nasledovne:

```
SELECT * FROM posts WHERE MATCH (description) AGAINST ('beautiful_
programming');
```

Výhodou fulltextu je rýchlejšie spracovanie, výsledky môžu byť usporiadané podľa relevancie, veľmi dobre pracuje aj na veľkých databázach s miliónmi záznamov. Fulltext sa využíva hlavne pre veľké objemy neštruktúrovaných dát, ak vznikla potreba optimalizovať výsledky alebo ak sú špecifické požiadavky na vyhľadávanie. (Abbas, 2011)

Vyhľadávacie nástroje využívajú dva druhy kritérií pre vyhodnotenie zhody medzi dokumentom a užívateľovým zadaním vyhľadávania.

1. Logické podmienky

Tento typ navracia hodnotu Boolean ako výsledok operácie zadanej užívateľom. Logické výrazy môžu byť komplexnejšie, zahŕňať viac stĺpcov, matematické kombinácie, funkcie a pod. Napríklad:

```
price < 100
LENGTH(title) <= 20
(author_id=123 AND YEAROF(date_added) >= 2000)
```

Text rovnako ako aj metadáta (date_added) môžu byť vyhodnocované ako logické výrazy. Dodatočné podmienky môžu byť uložené na základe existencie alebo neexistencie kľúčového slova v danom rade (výraz: cat AND dog BUT NOT mouse), alebo na základe zhody kľúčových slov v rámci skúmaného radu (fráza „To be or not to be“). Pretože výsledkom logického výrazu je hodnota TRUE alebo FALSE, môžeme takto ohodnotiť každého spracovávaného kandidáta a potom rozhodnúť o jeho zaradení do výsledkovej listiny alebo vylúčení z nej.

2. Fulltextové dotazy

Tieto sa rozpadajú na mnoho subtypov. Nasledujúce sa zaraďujú do všeobecnej kategórie vyhľadávania podľa kľúčových slov.

a) Boolean hľadanie

Je podobné vyššie spomínanému logickému, ale používa užší rozsah podmienok, ktoré jednoducho skontrolujú, či sa dané slovo v dokumente nachádza alebo nie. Výraz cat AND dog tak vráti každý dokument, ktorý obsahuje obe slová pričom nezáleží na mieste ich výskytu. Cat NOT dog tak vráti každý dokument, ktorý prvé slovo obsahuje a druhé sa nikde nevyskytuje.

b) Frázové vyhľadávanie

Slúži na vyhľadanie presnej zhody so zadaním, ktoré tvorí viac slovné zadanie frázy, takmer vždy označené úvodzovkami namiesto vyhľadávania každého slova samostatne. V tomto prípade nestačí informácia o výskyte slova, ale je potrebné poznať aj miesto výskytu. Na to, aby fungovalo vyhľadávanie tohto typu je potrebné vytvoriť indexy, ktoré nebudú ukladať iba informácie o tom, či je kľúčové slovo obsiahnuté v danom dokumente, ale aj jeho pozíciu.

c) Vyhľadávanie na základe podobnosti

Tento spôsob je viac flexibilný, využíva pozíciu na vyhľadanie zhody v dokumentoch tak, že skúma, či sa kľúčové slová nachádzajú v rámci určitej vzdialenosti medzi sebou. Syntax sa líši v závislosti od systému, ale pre Sphinx môže vyzeráť takto:

```
„cat dog“~5
```

To znamená nájsť všetky dokumenty, kde sa slovo cat a dog vyskytuje vo vzdialenosti 5 slov.

d) Vyhľadávanie na základe polí

Dokumenty takmer vždy mávajú viac ako iba jedno pole a programátori často potrebujú obmedziť časť vyhľadávania na určitú oblasť. Napríklad nájsť všetky správy od určitej osoby – od Petra, ktoré obsahujú v predmete MySQL:

```
@from Peter @subject MySQL
```


2.5.1 Elasticsearch

Elasticsearch je vyhľadávací a analytický software vyvinutý v jave. Všetky funkcie sú dostupné cez JSON a java API. Môže byť využitý na rôzne druhy dokumentov a poskytuje prispôsobiteľné fulltextové vyhľadávanie, štrukturované vyhľadávanie, analýzy alebo kombinácie. Je využívaný veľkými korporáciami. Principiálne funguje rovnako ako Sphinx, musí mať zaindexované záznamy, v ktorých potrebuje vyhľadávať. Dokáže usporadúvať výsledky podľa rôznych kritérií a má rôzne módy, ktoré určujú, o akú zhodu pri vyhľadávaní má ísť. Poskytuje navyše vytvorenie jazykového analyzátora, ktorému sa dá nadefinovať množina predložiek, spojok, synonym. Umožňuje využívať geolokáciu, takže výsledky vyhľadávania môžu byť presnejšie podľa typu dotazu a miesta výskytu užívateľa. To sú niektoré z dôvodov, prečo je využívaný hlavne veľkými korporáciami. Má množstvo rozličných nastavení, ktoré sa dajú kombinovať a prispôbovať okolnostiam. (Elasticsearch, © 2015)

2.5.2 MySQL

Databáza MySQL podporuje fulltextové indexovanie a vyhľadávanie, má však pár obmedzení. Indexy môžu byť vytvorené s tabuľkami InnoDB alebo MyISAM, a iba pre stĺpce typu char, varchar alebo text. Existujú tri typy vyhľadávania. Prvá možnosť je vyhľadávanie frázy v prirodzenom jazyku, ďalšia je pomocou boolean výrazov. Posledným spôsobom je rozšírenie vyhľadávania ako modifikácia prirodzeného jazyka. Hľadaný reťazec je použitý na vyhľadanie v prirodzenom jazyku, najviac relevantné výsledky sú pridané do vyhľadávaného reťazca a vyhľadávanie sa zopakuje. Dotaz zobrazí výsledky druhého hľadania. (MySQL, 2015)

2.5.3 Sphinx

Sphinx je samostatný open-source vyhľadávací software vydávaný pod licenciou GPL (General Public License), ktorý poskytuje rýchlu a tomu zodpovedajúcu funkcionálnosť vyhľadávania. (Aksyonoff, 2015) Bol navrhnutý na prácu s SQL databázou, ktorá ukladá dáta a na prístup k dátam sa využíva skriptovací jazyk. Ku svojmu behu však nevyžaduje žiadnu špecifickú databázu. Natívne podporuje MySQL, PostgreSQL, databázy kompatibilné s ODBC (MS SQL, Oracle), podporuje stop slová, hodnoty typu boolean, frázy, podobnosť slov a mnoho ďalších typov dotazov. Je možná škálovateľnosť distributívneho vyhľadávania, jednoduchá integrácia s dátovými zdrojmi v SQL a XML a mnoho ďalších funkcií. (Sphinx Technologies Inc, 2015)

Štruktúra sa principiálne dá porovnať s databázou. Dáta, v ktorých chceme vyhľadávať, tvoria obyčajný text. Ten však obsahuje veľa rovnakých slov, ktoré sa v celku opakujú a uchovávať text ako celok by bolo pamäťovo náročne a neprehľadné. Tento text sa teda musí rozdeliť na slová, ktoré je následne možné uchovávať ako indexy. Každé takéto slovo je uložené s užitočnými charakteristikami, ako je napr. počet výskytov, alebo jeho pozícia. V databáze sú dáta uložené v tabuľkách s preddefinovanými stĺpcami a je s nimi možné vykonávať operácie ako vkladanie, úprava či výmaz. Tomuto je ekvivalentný pojem Sphinx index. Riadok databázy je ekvivalentný Sphinx dokumentu – povedzme že to je akákoľvek forma textu, ktorá je predmetom nášho záujmu. No a vyhľadateľné polia dokumentu a pripojené atribúty (metadáta, napr. informácie o autorovi, dátum pridania)

predstavujú stĺpce databázy. Taktiež sa využíva dotazovací jazyk (query language), ktorý umožňuje pracovať so vzťahmi medzi vyhľadávanými frázami, najčastejšie sa využívajú binárne operátory – and, not, or.

Pri spracovávaní požiadavky od užívateľa Sphinx využíva fulltextový index pre rýchly pohľad na každú fulltextovú zhodu. Tým sa myslí dokument, ktorý sa zhoduje so všetkými kľúčovými slovami. Potom skúša doplnkové podmienky vyhľadávania, ktorého základom nie je kľúčové slovo. Podľa týchto dodatkových kritérií je jasné, či daný dokument má byť vrátený alebo nie. Dokument, ktorý je aktuálne prehľadávaný, sa nazýva kandidát (candidate). Kandidát, ktorý spĺňa každé kritérium sa nazýva zhoda (match). Zhody sú potom hodnotené (ranked) a to tak, že Sphinx vypočíta a priradí určitú hodnotu relevancie, zoradí zhody podľa tejto hodnoty a vráti top N najlepších zhôd aplikácii, ktorá si to vyžiadala. Týchto top N sa nazýva množina výsledkov (result set). (Aksyonoff, 2015)

Z dôvodu využívania jazyka PHP a s ohľadom na zložitosť a komplexnosť spomínaných fulltextových vyhľadávačov bol pre implementáciu zvolený Sphinx. Po výbere môže byť prístupné k podrobnejšiemu popisu funkcií, ktoré budú užitočné pre implementáciu.

Sphinx API poskytuje funkcie pre vyhľadávanie v závislosti na pr esnosti vyhľadávania. Sú to:

- SPH_MATCH_ALL, nájde zhodu vo všetkých hľadaných slovách;
- SPH_MATCH_ANY, nájde zhodu v ktoromkoľvek slove z vyhľadávanej frázy;
- SPH_MATCH_PHRASE, vyhľadá zhodu frázy, vyžaduje dokonalú zhodu medzi hľadaným a zdrojovým textom;
- SPH_MATCH_BOOLEAN, nájde zhodu ako výsledok boolean výrazu využívajúceho operátory OR a NOT, napr. výraz *fantazie !truhlík* – hľadaj fantazie, ale nemá obsahovať truhlík;
- SPH_MATCH_EXTENDED2, nájde zhodu ak je hľadaná fráza vyjadrená ako výraz v internom jazyku Sphinxu, môže využívať operátory ako boolean výrazy, k tomu operátor MAYBE, operátor vyjadrujúci atribút/atribúty, v ktorých má byť slovo hľadané, operátor presnosti, ktorý definuje, koľko slov sa môže maximálne nachádzať medzi dvoma hľadanými slovami frázy a mnoho ďalších. (Sphinx, 2015)

Ďalej poskytuje módy pre usporiadanie výsledkov.

- SPH_SORT_RELEVANCE, usporiadava výsledky podľa relevancie zostupne, najlepšie prvé;
- SPH_SORT_ATTR_DESC, usporiadava podľa zvoleného atribútu zostupne, najvyššie prvé;
- SPH_SORT_ATTR_ASC, usporiadava podľa zvoleného atribútu vzostupne, najnižšie prvé;
- SPH_SORT_TIME_SEGMENTS, usporiadava výsledky podľa časového údaju;
- SPH_SORT_EXTENDED, využíva kombinácie podobne ako SQL, rôzne stĺpce s kombináciami vzostupne/zostupne;

- SPH_SORT_EXPR, usporiadava aritmetické výrazy

Prvá z možností ignoruje akýkoľvek ďalší parameter a vždy sa riadi podľa relevancie. Relevancia v tomto prípade znamená, že váha sa slovu pridáva podľa množstva jeho výskytov vo vyhľadanom zázname. Relevantnejšie výsledky sú tie, ktoré zadaný výraz obsahujú viackrát ako zvyšné. Ďalšie spomínané módy k svojmu fungovaniu potrebujú pridať názov atribútu, s ktorým majú pracovať.

3 Metodika riešenia

Vypracovanie odbornej práce si vyžaduje prípravu z hľadiska definície problému a problematiky ako celku. Správne pochopený problém s jasnými požiadavkami na jeho výslednú funkcionálnosť je vhodným začiatkom pred samotným riešením. Tieto požiadavky sú popísané v kapitole 1.2. Na jeho splnenie budú nutné nasledovné kroky:

1. analýza súčasného stavu;
2. dotazníkový prieskum;
3. definovanie obchodných pravidiel vyhľadávania;
4. simulácia súčasného stavu vyhľadávania;
5. aplikácia pravidiel na optimalizáciu vyhľadávania;
6. implementácia;
7. testovanie.

Analýza súčasného stavu bude vykonaná z dvoch pohľadov. Prvým je preskúmanie správania návštevníkov skúmaného e-shopu za pomoci nástroja webovej analytiky Google Analytics (ďalej ako GA). Budú zisťované informácie o návštevnom správaní webu, o počte návštev s využitím vyhľadávania a bez neho a jeho vplyv na nákupné správanie. Ďalej sa budeme zameriavať na vyhľadávané slová a ich počet, množstvo upresnení vyhľadávania. Budeme skúmať vplyv času stráveného na webe, počty opustení stránok a hĺbku vyhľadávania na transakcie. Druhým pohľadom je užívateľský, ktorý vyplýva zo skúseností. Vyhľadávané slová získané pomocou GA budú použité pre zadávanie do vyhľadávania priamo na stránke a bude sa sledovať, aké výsledky dostáva užívateľ. Vzhľadom na neobjektívnosť získaných informácií týmto spôsobom je nutné vytvoriť dotazník pre užívateľov e-shopu, ktorým budú zisťované preferencie respondentov pri zobrazení produktov a výsledkov vyhľadávania na najčastejšie hľadané slová a frázy. Dotazník bude realizovaný elektronickou formou pomocou aplikácie survio.com s využitím obrázkov pre lepšiu predstavivosť. Tieto informácie budú spracované a využité v ďalšom kroku pri definícii obchodných pravidiel pre vyhľadávanie.

Po spracovaní dát a získaní informácií potrebných k tvorbe optimalizácie bude prístupné k navodeniu aktuálneho stavu za pomoci viacerých nástrojov. Bude potrebné nainštalovať WampServer 2.5, ktorý využíva webový server Apache 2.4.9. Taktiež podporuje databázový systém MySQL 5.6.17, skriptovací jazyk PHP 5.5.12 a obsahuje nástroj PHPMyAdmin 5.5.12 slúžiaci na administráciu MySQL s použitím webového prehliadača. Následne bude prístupné k použitiu fulltextového vyhľadávača Sphinx, ktorý bude pracovať s dátami uloženými v MySQL databáze. Dáta sú poskytnuté vo forme tabuliek v .xlsx súbore, tie budú importované ako obsah databázy. Okrem systémových nástrojov je však dôležité aj nastavenie vyhľadávania v konfiguračnom súbore pre vyhľadávanie Sphinx-om, ktorý bude taktiež poskytnutý.

Samotný algoritmus bude implementovaný v jazyku PHP za pomoci funkcií, ktoré poskytuje Sphinx API.

Testovanie prebehne na vzorke nezainteresovaných osôb, ktorým bude poskytnuté súčasné a inovované riešenie na porovnanie. Budú zhodnotené výsledky porovnaním sú-

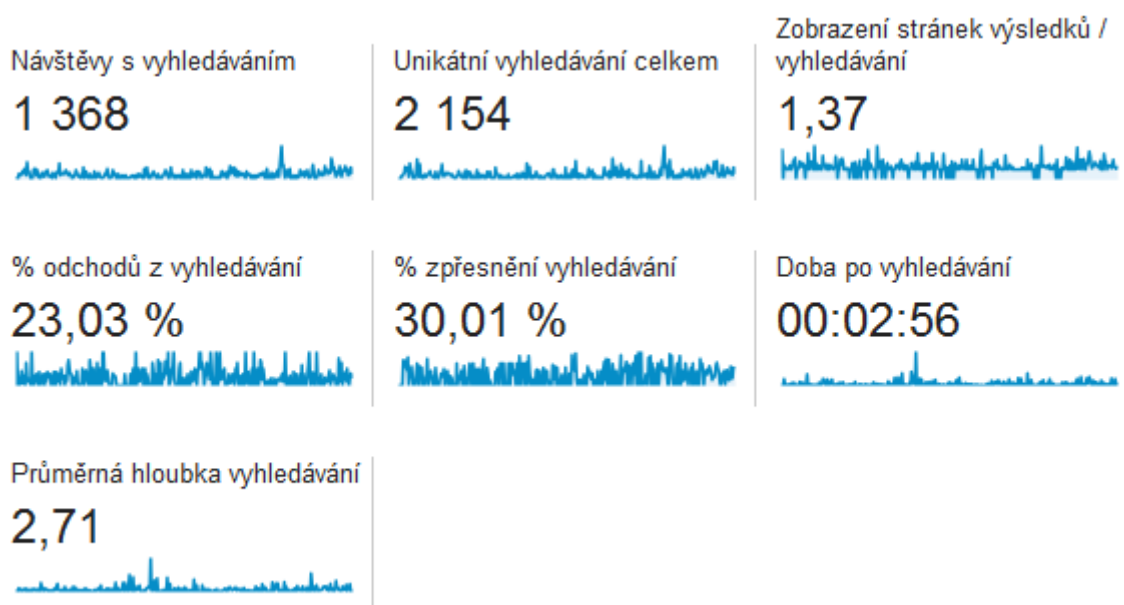
časného a nového vyhľadávania z hľadiska funkčnosti a užívateľskej prívetivosti podľa osôb zúčastnených na testovaní. v závere budú načrtnuté ďalšie možné rozšírenia práce.

4 Analýza súčasného stavu

4.1 Správanie zákazníkov

Lepšia užívateľská skúsenosť sa vo veľkej miere prejaví aj na cieľoch, ktoré má firma stanovené, a to hlavne v maximalizovanom zisku. Preto je dôležité sledovať aktivitu zákazníkov na webe. Tá nám prezradí, čo je vhodné vylepšiť, pri akých činnostiach zákazníci opúšťajú webové stránky, koľko času priemerne strávia surfovaním po webe, aké slová najčastejšie vyhľadávajú, a pod. Meranie skúseností s interným vyhľadávaním je v tomto prípade kľúčovou metrikou.

Na obrázku 3 sú zobrazené štatistiky za celý rok (obdobie od marca 2014 do marca 2015). Toto obdobie je zvolené z dôvodu uceleného pohľadu na web, aby boli eliminované sezónne výkyvy.



Obr. 3: Základné KPI vyjadrujúce vyhľadávanie

V prvej časti obrázku 3 sú návštevy s vyhľadávaním v absolútnom vyjadrení, percentuálne je to 6,9%. Je to počet návštev, pri ktorých bolo aspoň raz využité vyhľadávanie. Unikátne vyhľadávanie na vedľajšom obrázku hovorí o počte, koľkokrát užívatelia vyhľadávali na webových stránkach, pričom duplicitné vyhľadávanie v rámci jednej návštevy je vylúčené. Ďalšia časť, zobrazenie stránok výsledkov, vyjadruje priemerný počet, koľkokrát po vyhľadávaní návštevníci zobrazili nejakú zo stránok uvedených vo výsledkoch vyhľadávania. Číslo 1,37 teda signalizuje, že po vyhľadaní si návštevníci zobrazili výsledok vyhľadávania 1-2 krát, potom pravdepodobne zmenili vyhľadávaciu frázu alebo zvolili iný spôsob prehľadávania. Odchody z vyhľadávania (23%) znamenajú počet opustení webu na základe výsledkov interného vyhľadávania. Ďalšia časť obrázku hovorí o celkovom počte upresnení, ku ktorým dôjde v priebehu návštevy medzi internými vyhľadávacími kľúčovými slovami. Približne 30 % návštevníkov je ochotných upraviť svoju vyhľadávaciu frázu, pokiaľ nie je

s výsledkami spokojná. Ďalej je doba, ktorú návštevníci strávili na webe potom, čo získali výsledky vyhľadávacieho dotazu. Priemerne je tento čas dve a pol minúty. Priemerná hĺbka vyhľadávania udáva počet stránok (3 stránky v analyzovanom prípade), ktoré návštevníci prehliadali na webe po tom, čo získali výsledky vyhľadávacieho dotazu.

To nám vlastne hovorí, že veľmi malé percento užívateľov využíva vyhľadávanie, čo môže byť spôsobené viacerými faktormi. Web je veľmi dobre prepracovaný, zákazník vždy nájde, čo potrebuje, a nie je preto motivovaný ani nútený využiť vyhľadávanie. Na webe sa nachádza málo produktov, v ktorých sa dá dobre orientovať, čo vedie k prvej variante. Alebo môže byť chyba vo vyhľadávaní, ktoré nefunguje ideálne, a zákazník sa radšej prekliká stromom produktov a nájde si, čo hľadá týmto spôsobom. V kombinácii s percentom opustení stránok po vykonaní vyhľadávania a s množstvom spresnení vyhľadávanej frázy môžeme usúdiť, že dotaz v tretine prípadov musí zákazník upraviť, aby bol spokojný, alebo to vzdá a opustí web. To je samozrejme najhoršia varianta, ktorej sa treba vyhnúť.

Medzi ďalšie významné ukazovatele KPI pre hľadanie patrí aj to, ako sa návštevníci využívajúci interné vyhľadávanie líšia od tých, ktorí túto službu nevyužívajú.

Tabuľka 1: Rozdiel medzi zákazníkmi podľa využitia možností interného vyhľadávania

Stav vyhľadávania na webe	Akvizícia		Správanie			Konverzia	
	Návštevy	Noví užívatelia	Miera okamžitého opustenia	Počet stránok na jednu návštevu	Priemerná doba trvania jednej návštevy	Tranzakcia	Konverzný pomer e-obchodu
Celkom/ Priemer	19 753	12 998	33,92 %	5,87	0:03:27	264	1,34 %
Návštevy BEZ vyhľadávania	93,07 %	94,05 %	36,23 %	5,56	0:03:10	94,70 %	1,36 %
Návštevy S vyhľadávaním	6,93 %	5,95 %	2,85 %	10,1	0:07:16	5,30 %	1,02 %

Podľa tabuľky 1 je vidieť, že zákazníci využívajúci vyhľadávanie majú zrejme jasnú predstavu o tom, aké produkty hľadajú. Na jednu návštevu prehľadajú takmer dvojnásobný počet stránok ako zákazníci bez vyhľadávania a trvanie návštevy je tiež približne dvojnásobné. To znamená, že sú zrejme odhodlaní nájsť, kvôli čomu web v prvom rade navštívili. Tomu nasvedčuje aj miera okamžitého opustenia, iba necelé tri percentá návštev skončia opustením webu hneď po zobrazení prvej stránky.

Tabuľka 2: Konverzie nákupu podľa využitia vyhľadávania

Stav vyhľadávania na webe	Návštevy	Transakcie	Tržby	Priem. tržba za návštevu
Bez vyhľadávania	18 385	250	1,36 %	317 160 Kč
s vyhľadávaním	1 368	14	1,02 %	10 657 Kč

Tabuľka 2 poskytuje informáciu o tom, že využitie vyhľadávania nemá výrazný vplyv na percentá transakcie. Transakcia v tomto prípade znamená dokončenie nákupu v skúma-

nom e-shope. Zákazníci, ktorí požadované produkty našli iným spôsobom ako interným vyhľadávaním, ukončili svoju návštevu vykonaním objednávky iba v o 0,34 % prípadoch viac ako tí, ktorí vyhľadávanie využili.

Ďalším významným ukazovateľom vo webovej analytike je index USD. Je to stránková metrika, ktorá určuje hodnotu jednej stránky vzhľadom k cieľovej konverzii. Podľa Němca inak povedané, vysoká hodnota tohto indexu znamená, že návštevníci prešli touto stránkou a následne vykonali konverziu, najčastejšie vykonali objednávku. (Němec, 2009)

Konverzia znamená činnosť, ktorá je žiadaná. V závislosti od typu stránok to môže byť napríklad objednávka, zaslanie komentára, zaslanie ponuky, stiahnutie určitých dokumentov a podobne. Výpočet je definovaný nasledujúcim spôsobom:

$$\text{Index USD} = \frac{\text{hodnota cieľa} + \text{príjem z elektronického obchodovania}}{\text{počet jedinečných zobrazení stránky}} \quad (1)$$

Samotný príjem nemusí byť relevantný, pretože na každom druhu tovaru môže byť iná marža, preto sa v čitateli v zlomku používa súčet daných hodnôt. Index USD je aspoň čiastočne schopný vyrovnávať tieto rozdiely. V ideálnom prípade je vhodné použiť všetky hodnoty pre uskutočnenie výpočtu. Čím väčší dôraz je kladený na nastavenie hodnôt, tým bude mať index vyššiu vypovedajúcu schopnosť. V realite však nemusíme vedieť, ako nastaviť hodnotu cieľa, prípadne ak neexistuje elektronický obchod, ťažko sa bude počítať s hodnotou jeho príjmu. V skúmanom prípade je hodnota cieľa nastavená na nulu, takže hodnota indexu USD vychádza iba z príjmov elektronického obchodu.

Táto metrika sa využíva k identifikácii stránok, ktoré sú najvhodnejšie pre optimalizáciu. Logicky sa ponúka variant, kedy je vhodné optimalizovať stránky podľa najnižšieho indexu USD. To však nemusí byť vždy najlepšie, pretože tieto stránky sa s najväčšou pravdepodobnosťou ani nenachádzajú na návštevníkovej ceste k požadovanej činnosti. Preto je lepšie sa zamerať na stránky, ktoré majú najvyšší index, čo znamená, že sa návštevníkovi zobrazujú často (napr. pri dokončení objednávky), alebo cez ne často prechádza. Upraviť ich treba podľa uváženia, ak sú to napr. stránky so správou o neúspešnej platbe, môžu signalizovať určité problémy s formulárom objednávky. Vhodný začiatok tiež môže byť so stránkami s vysokými hodnotami, ktoré majú vysokú mieru okamžitého opustenia, teda že návštevník opustil web po vzhliadnutí stránky bez toho, aby vykonal akúkoľvek interakciu.

S prihliadnutím na fakt, že hodnota cieľa je v našom prípade nulová, bude to zohľadnené. V prehľade obsahu stránok nastavíme parametre tak, aby sme sa zamerali na stránky, ktoré smerujú k opusteniu webu. Filter teda bude nastavený na 100% mieru okamžitého opustenia, údaje budú zoradené podľa počtu zobrazení stránok a cieľom budú stránky, ktoré majú najvyššiu hodnotu, teda index USD. Tie, ktoré zostali v prehľade po aplikácii týchto filtrov obsahujú údaje o produktoch. Najčastejšie ide o zobrazenie stránky s konkrétnym produktom z vyššej cenovej kategórie a priemerný čas strávený na stránkach je 45 sekúnd, z čoho sa dá usúdiť, že dôvodom pre opustenie pravdepodobne nebolo zobrazenie danej stránky, ale iný faktor, napríklad cena produktu či chýbajúca fotografia. V tomto prípade nie je dôvod dané stránky nejakým spôsobom upravovať.

V kombinácii, ktorá využíva hodnoty stránky v relácii s počtom zobrazení, sa najčastejšie vyfiltrujú stránky, ktoré sú priamo spojené s cieľovou konverziou, teda súčasti nákupu. Takými sú stránky s obsahom košíka, platbou, vyplnením kontaktných informácií, dodacej adresy, prehľad zákazníckych objednávok, prihlásenie, registrácia, potvrdenie

úspešnej objednávky a pod. Z toho vyplýva, že štruktúra elektronického obchodu by mala fungovať správne, na popredných priečkach sa neobjavujú žiadne nežiaduce stránky.

Stránky, ktoré nie sú priamo súčasťou objednávkového procesu, obsahujú konkrétne produkty. Najzaujímavejším produktom čo sa týka počtu návštev, hodnoty stránky a percentom odchodov sú zimné športové boby, najmä však boby z radu šoférovateľné. Vyššie percento odchodov však pri tomto produkte môže byť spôsobené cenou, návštevou zo záujmu či zvedavosti.

Podľa vyhľadávaných fráz v internom vyhľadávaní (zobrazených na obr. 3) je vidno, že užívatelia vyhľadávajú skôr všeobecné názvy. Vedia, akú produktovú radu hľadajú, ale nevedia tvar, typ, farbu či rozmery. Je to zrejme najjednoduchší spôsob v prípade, že nemajú presnú predstavu. Nie je potrebné preklikávať strom produktov, kým užívateľ nájde, do ktorej kategórie je ním hľadaný produkt zaradený. Pri výsledku vyhľadávania, ktorý vráti produktovú radu je jednoduché vybrať si typ, a zvoliť farbu a rozmery, ktoré sú dostupné.

Tabuľka 3: Prehľad vyhľadávacích fráz v internom vyhľadávaní TOP 10

Vyhľadávaná fráza	Unikátne vyhľadávanie celkom		Zobrazenie stránok výsledkov	Vyhľadávanie			
				% odcho- dov	% upres- nenia	Doba po hľadaní	Priemerná hĺbka
	2154		1,37	23,03 %	30,01 %	0:02:56	2,71
1. elise	93 (4,32 %)		1,73	17,20 %	11,80 %	0:04:58	4,63
2. misky	58 (2,69 %)		1,02	15,52 %	5,08 %	0:00:45	2,10
3. fantazie	48 (2,23 %)		1,56	12,50 %	38,67 %	0:07:01	4,12
4. stratos	45 (2,09 %)		1,16	46,67 %	15,38 %	0:03:31	2,02
5. gardenie	43 (2,00 %)		1,91	32,56 %	25,60 %	0:04:55	4,12
6. quattro	29 (1,35 %)		1,14	24,14 %	30,30 %	0:01:20	3,45
7. balconia	25 (1,16 %)		1,08	40,00 %	33,33 %	0:02:04	1,44
8. truhlík	24 (1,11 %)		3,33	29,17 %	20,00 %	0:01:57	2,42
9. cubico	23 (1,07 %)		3,00	39,13 %	7,25 %	0:07:47	5,09
10. miska	23 (1,07 %)		1,04	47,83 %	16,67 %	0:01:31	1,74

Z toho logicky vyplýva, že na prvých priečkach výsledkov vyhľadávania by mali byť produkty, ktoré užívateľ hľadá s najväčšou pravdepodobnosťou, aby mal čo najjednoduchšiu orientáciu. Najmenej často vyhľadávané frázy nie sú pre prehľadnosť v tabuľke znázornené, pretože majú veľmi nízky počet unikátnych vyhľadání. Nachádzajú sa tu napríklad konkrétne produkty jednoznačne definované názvom a rozmermi, prípadne EAN kódom.

Čo sa však ukazuje ako zaujímavý poznatok pri tejto analýze, ľudia často hľadajú aj produkty, ktoré e-shop neposkytuje. Keď je návštevník v obchode s plastovými výrobkami, skúša hľadať rôzne produkty, ktoré by sa tu mohli nachádzať. To je celkom logické, na to slúži vyhľadávanie. Za zamyslenie však stojí informácia, že druhý najhľadanejší produkt sa v tomto elektronickom obchode vôbec nenachádza. Otázkou je, či to nie je tzv. diera na trhu. Potenciál obchodu by bol, zákazníci tu daný tovar očakávajú, len tovar neexistuje.

So znalosťou kľúčových slov v internom vyhľadávaní sa dá pracovať na viacerých úrovniach, Jednou z nich je úprava interného vyhľadávania a stránok produktov podľa požiadaviek firmy či potrieb zákazníkov, druhá je využiť najviac hľadané slová do externého

vyhľadávania (Google, Bing a pod.). Slová s najvyšším počtom hľadání je vhodné zaradiť medzi kľúčové slová do plateného vyhľadávania a naopak, vďaka tomuto ukazovateľu je možné určiť negatívne kľúčové slová, ktoré je vhodné z plateného vyhľadávania odstrániť.

Ukazovateľ hodnota za jedno vyhľadávanie meria hodnotu výrazu hľadaného na stránkach. Ukazuje, či návštevníci hľadajúci daný výraz pokračujú v dosiahnutí nejakej akcie, ktorá je prínosná. Čím je táto hodnota vyššia, tým väčší prínos má daný výraz pre analyzovaný obchod. (Clifton, 2009) Ako ukazuje tab. 4, podľa tohto ukazovateľa má najvyšší prínos *quadrato*, *fantazie*, *cubico*. Výrazy na obrázku sú zoradené podľa hodnoty za vyhľadávanie, tá je však v prvých troch prípadoch najvyššia vďaka jednej/dvom návštevám, takže sa nejedná o štatistiky významné slovo.

Tabuľka 4: Ukazovateľ hodnota za jedno vyhľadávanie TOP 10

Vyhľadávaná fráza	Unikátne vyhľadávanie celkom	Hodnota na vyhľadávanie
	2154	4,70 Kč
1. květináč na svod	1	795,00 Kč
2. 28x14x21	1	695,00 Kč
3. Smart	2	358,00 Kč
4. quadrato	13	157,77 Kč
5. fantazie	48	58,75 Kč
6. bambi bob	5	55,00 Kč
7. cubico	23	51,83 Kč
8. ella	19	30,00 Kč
9. truhlík	24	20,92 Kč
10. elise	93	5,52 Kč

4.2 Aktuálne využívané vyhľadávanie

Vyhľadávanie na stránkach je v súčasnej dobe pri rozsiahlych weboch tak dôležité, že sa stalo súčasťou navigačného systému. Taktiež pri menších weboch môže kvalitné interné vyhľadávanie zlepšiť užívateľskú skúsenosť. Podľa tabuľky 2 je zrejmé, že v skúmanom období priemerne zákazník používajúci vyhľadávanie vykonal nákup za nižšiu sumu.

Princíp momentálne používaného vyhľadávania je jednoduchý. Z hľadiska logiky sa pre každé vyhľadávané kľúčové slovo zobrazia všetky výsledky, ktoré dané slovo obsahujú. Primárne sú všetky výsledky usporiadané podľa ceny vzostupne a pri rovnakej cene sú radené podľa interného kódu ID produktu. Dokumentom, ktoré sú výsledkom vyhľadávania, môžeme pridať váhy. Je to číselná hodnota, ktorá označuje, ako relevantný je výsledok k zadávanému dotazu. Najviac zhodované produkty k zadávanému dotazu by teda mali byť medzi prvými. Váhy sú v súčasnosti nastavené nasledovne:

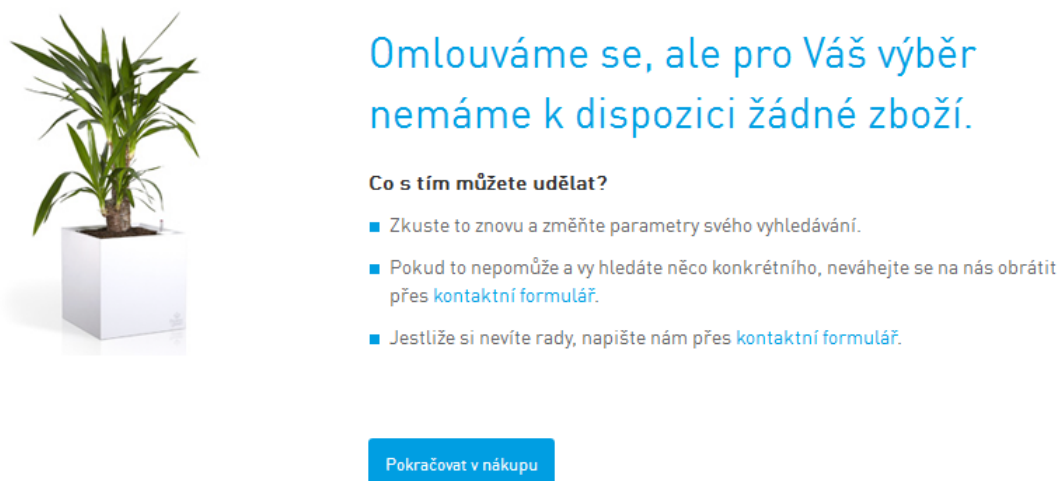
```
'name' => 100,
'name_url' => 90,
'short_description' => 80,
'product_user_information' => 80,
'seo_title' => 70,
'seo_description' => 70,
```

```
'seo_keywords' => 70,  
'seo_url' => 70,  
'code' => 60,  
'ean' => 50
```

To znamená, že najväčší dôraz je kladený na to, aby sa vyhľadávané slovo nachádzalo v názve, potom v URL odkaze, v popise a tak ďalej s klesajúcou dôležitosťou. Avšak ak je predradené usporiadanie podľa ceny, tieto váhy sú pravdepodobne ignorované. Vyplývajú z predošlého popisu tak vznikne stav, kedy návštevník zadá ako kľúčové slovo nejakú farbu, a vďaka cene a identifikátoru je zobrazené poradie logicky neucelené.

V nasledujúcich odstavcoch sa budeme venovať vyhľadávaniu, zadávaným frázam a zodpovedajúcim výsledkom. Ako vidno z tab. 3, najvyhľadávanejším produktom je elise. Patrí medzi kvetináče a nie je k nemu odporúčané žiadne príslušenstvo, zároveň sa slovo elise nezhoduje so žiadnym iným podobným v češtine, takže vyhľadávanie funguje ideálne. Zobrazí všetky produkty, ktoré patria do produktovej rady elise a usporiadané sú podľa východiskového nastavenia – podľa ceny vzostupne. Obdobne sú na tom aj iné produktové rady bez príslušenstva, napr. cubico.

Druhým najvyhľadávanejším výrazom je slovo misky. Vzhľadom na neexistenciu tohto produktu medzi ponúkaným tovarom je pochopiteľné, že výsledky neobsahujú žiaden produkt. Užívateľ uvidí nasledujúci oznam:



Obr. 4: Výsledok neúspešného vyhľadávania

Z hľadiska vyhľadávania je to dostačujúce upozornenie, no z pohľadu obchodnej logiky je to nevyužitá príležitosť pre zistenie kvalitatívnych informácií od zákazníka. Ľudia vo všeobecnosti nemajú radi dotazníky, pretože s rastúcou konkurenciou je ich na trhu stále viac a viac a neprinášajú ľuďom žiadny priamy úžitok, preto nie sú motivovaní strácať čas s jeho vyplňaním. V tomto prípade však zákazník hľadá produkt, o ktorý by mal pravdepodobne záujem, takže zodpovedať pár krátkych a rýchlych otázok by bol motivovaný a na druhú stranu by takáto forma takmer nič nestála. Prívetivý mini dotazník by vyplnilo viac zákazníkov, pretože sa objavil v situácii, kedy by zákazník mal chuť prispieť svojim názorom. Napríklad:

1. Myslíte si, že by sa Vami hľadaný tovar mal nachádzať v ponuke?

2. Mali by ste záujem o zaradenie Vami požadovaného tovaru do ponuky?
3. Hľadáte produkt, ktorý sa v ponuke v minulosti nachádzal a už nie je dostupný?
4. V akej cenovej relácii by bol pre Vás zadaný produkt zaujímavý?

Čo sa bude diať s odpoveďami je už otázkou na vyšší management spoločnosti a na ďalšie analýzy vhodnosti a výhodnosti o zaradenie produktu. Na zákazníka to však môže pôsobiť ako zvýšený záujem o jeho potreby, pričom v neustále rastúcej konkurencii sa každá snaha počíta a zákazník je schopný ju oceniť.

Vypĺňať kontaktný formulár je taktiež menej efektívna forma, pretože zákazník musí byť oveľa viac odhodlaný niečo zmeniť, vypĺňať viac informácií, stojí ho to viac času ale hlavne, musí svoje myšlienky formulovať do prázdneho textového poľa.

Tretím najviac vyhľadávaným výrazom je názov produktovej rady *fantazie*. Táto kategória zahŕňa viac produktov, a to kvetináče (truhlíky) a k nim vhodné držiaky na upevnenie. Hlavným tovarom je pre obchod samozrejme kvetináč, a držiak je až v kategórii príslušenstvo. V súčasnom stave vyhľadávania sa medzi prvé výsledky zaraďujú všetky držiaky, a až po nich sú zobrazené všetky truhlíky dostupných veľkostí a farieb, čo môže na zákazníka pôsobiť mätúco alebo neprehľadne. Do tejto kategórie s rovnako zobrazenými výsledkami môžeme zaradiť aj produktovú radu *gardenie*.

Ďalším slovom je *stratos*, názov zimnej športovej pomôcky, tzv. boby. Opäť pri vyhľadávaní nenastáva žiadna komplikácia, zobrazia sa všetky dostupné farby. Vzhľadom na veľkosť uni sú tieto produkty 4, čo pôsobí prehľadne. Často vyhľadávaný výraz *quattro* pravdepodobne vyjadruje preklep a užívateľ má na mysli produktovú radu *quadrato*. Z užívateľského pohľadu by bolo vhodné, keby bola k dispozícii pomôcka alebo návrh existujúceho slova, s ktorým si daný výraz mohol spliesť. V prípade hľadania *quadrato* sa všetky produkty zobrazia správne, žiadny produkt, ktorý by nevyhovoval vyhľadávaniu sa tu nenachádza. *Truhlík* je výnimočne zaujímavé kľúčové slovo. Je to všeobecné pomenovanie typu kvetináča, ktorý sa nachádza vo viacerých produktových radách. Medzi výsledkami vyhľadávania sa zobrazujú aj držiaky na tieto truhlíky. Návštevník však pri hľadaní truhlíka pravdepodobne nechce držiak. Prekvapením však je, že sa zobrazia aj produkty, ktoré nemajú s týmto výrazom absolútne nič spoločné, napr. klzák, tanier na sánkovanie či kvetináč na orchideu. V názve ani vo svojom popise neobsahujú hľadané slovo.

Zákazníkov tiež napadlo hľadať produkt *koš*. Takýto produkt neexistuje, avšak vyhľadávanie pre lepšiu užívateľskú skúsenosť ignoruje interpunkciu. Vzhľadom na produkty z rady *cubico*, ktoré majú tvar „kostky“, zobrazené výsledky obsahujú všetky farby a veľkosti týchto kvetináčov. Medzi nimi sa tiež nachádza tanier na sánkovanie. Návštevník tak musí prezrieť takmer 50 produktov aby zistil, že *koš*, ktorý hľadá, sa tu nenachádza.

V prípade prázdneho hľadania, teda bez zadaného kľúčového slova sa zobrazí hlásenie rovnaké ako na obr. 6. Ak má zákazník zaškrtnutý nejaký výber vo filtri a náhodou sa mu podarí vykonať nechcené prázdne hľadanie, zobrazené produkty zmiznú a je nútený si filter nastaviť znova. Alebo sa vrátiť na pôvodnú stránku tlačidlom „späť“ v prehliadači. Jeho výber zostane uložený až na výnimku poslednej označenej položky.

4.3 Zostavenie dotazníka a analýza výsledkov

Pri vytváraní softvéru alebo webových služieb je na začiatku potrebné pochopiť užívateľov, čo potrebujú, čo majú radi, prečo používajú to, čo používajú a ako s tým môžu pracovať. Čím viac informácií o užívateľoch máme, tým sme schopní vytvoriť lepšiu funkcionálnosť podľa ich potrieb. Poskytnuté služby, ktoré spĺňajú požiadavky, znamenajú spokojného zákazníka.

Podstata problému spočíva v znalosti cieľa, ktorí chcú užívatelia dosiahnuť. Mnohokrát to však ani sami nevedia. Pravidlá dizajnu podľa Tidwellovej hovoria, že užívateľom treba vytvoriť cestu k stanoveným cieľom (hlavne nákupu) priamu a čo najjednoduchšiu. (Tidwell, c2011) Nezaťažovať ich zbytočnosťami, ktoré môžu odradiť od pokračovania v práci alebo vyhľadávanií.

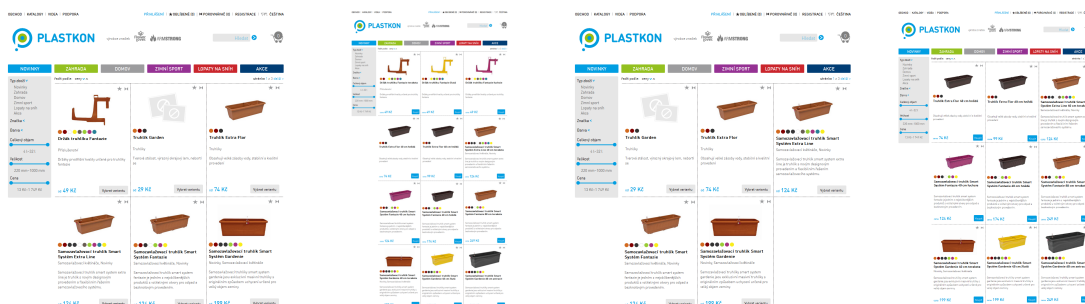
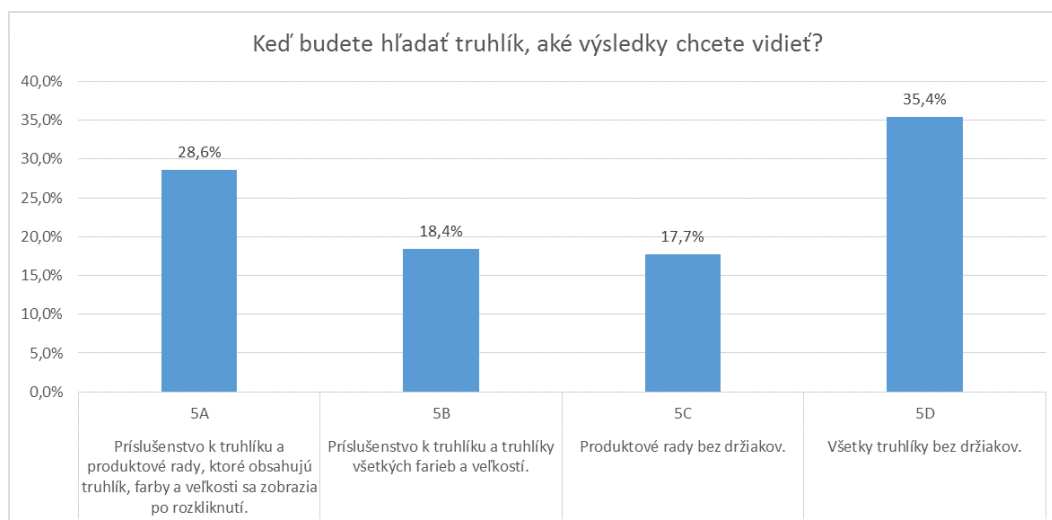
Pri návrhu vylepšenia aplikácie však treba zahrnúť viacero faktorov, ktoré ovplyvňujú správanie: demografické údaje, podvedomé reakcie, sociálne okolie, hodnoty atď. Preto je potrebné zistiť, čo užívatelia naozaj chcú, čo využívajú a čo by mali radi k dispozícii. Najlepším spôsobom je spoznať užívateľov. To je však v niektorých prípadoch nemožné, preto existujú dotazníky, v ktorých sa ich môžeme priamo spýtať. Dôležité je rozlíšiť, čo je všeobecne pravdivé a čo je iba náhodný a nepredvídateľný jav. Hlavný rozdiel medzi vnímaním spomínaných pojmov je v tom, že tvorca softvéru sa snaží pochopiť požiadavky užívateľa, ktorý s ním pracuje, pričom marketér chce pochopiť tých, ktorí daný softvér kupujú.

Na základe správania zákazníkov podľa kapitoly 4.1 a vyhodnotenia ich reálnych požiadaviek z kvantitatívneho hľadiska bolo pristúpené k tvorbe dotazníka ako k doplnku dát kvalitatívneho charakteru. Dotazník sa zameriava na návštevníkove aktuálne potreby, ale aj na získanie jeho predstáv o fungovaní vyhľadávania a zobrazovania zo subjektívneho hľadiska, či už ide o nápady na vylepšenie, pripomienky alebo porovnanie s konkurenciou. Dotazník obsahuje 15 otázok, z väčšej časti tvorené obrazovými možnosťami pre konkretizáciu predstavy a vizualizáciu riešenia. Umiestnený bol na internetových stránkach sledovaného obchodu počas doby približne štyroch mesiacov a skupina respondentov bola tvorená návštevníkmi/zákazníkmi tohto obchodu.

Počet návštev na stránke odkazu na dotazník bolo 345, pričom dokončených a správne vyplnených dotazníkov bolo 147 kusov. Doba vyplnenia dotazníku bola v rozpätí 2-5 minút v takmer 58 % prípadov, do 10 minút to bolo 24 % z celkového počtu dotazníkov, 10-30 minút v 16 % prípadov a zvyšné 2 % sú menej ako 2 minúty alebo viac ako 30 minút. Najčastejšia veková kategória sa vyskytovala v rozmedzí 18-30 rokov, čo tvorilo 85 % respondentov, ďalších 10 % tvorili návštevníci vo veku 31-50 rokov. Okrajové kategórie menej ako 18 rokov a viac ako 51 rokov tvorili zvyšných necelých 5%. Ženy vyplnili iba o 6 % dotazníkov viac ako muži (8 ks). Vysokú školu prvého stupňa uviedlo ako svoje najvyššie dosiahnuté vzdelanie 56 %, na druhom mieste s 20 % sú na tom rovnako stredná škola s maturitou a vysoká škola druhého stupňa, 2 % respondentov svoje vzdelanie ukončili výučným listom a zvyšné necelé 2 % tvorí kategória iné. Najviac respondentov teda tvorili vysokoškolsky vzdelané osoby vo veku 18 až 30 rokov.

Dotazník bol vytváraný tak, aby jeho znenie nebolo závislé na znalosti a zobrazení webu a mohol ho tak vyplniť aj nezainteresovaný človek. Vzhľadom na analýzu súčasného stavu vyhľadávania boli do dotazníku umiestnené dotazy na vyhľadávané frázy. Dotazom na truhlíky sa vo výsledkoch ako prvé zobrazovali držiaky na truhlíky, preto nás zaujímal

užívateľský názor na výsledky toho typu – na obr. 5. Hľadáme všeobecný produkt, ktorý sa nachádza v rôznych obmenách v rozličných produktových radách, líšiaci sa farbou a súvisiaci s príslušenstvom. Obrázky pod grafom sú iba ilustračné, ukazujú možnosti, aké videli užívatelia v dotazníku. v čitateľnej forme sú umiestnené v elektronickej podobe na CD (príloha B – v dotazníku).



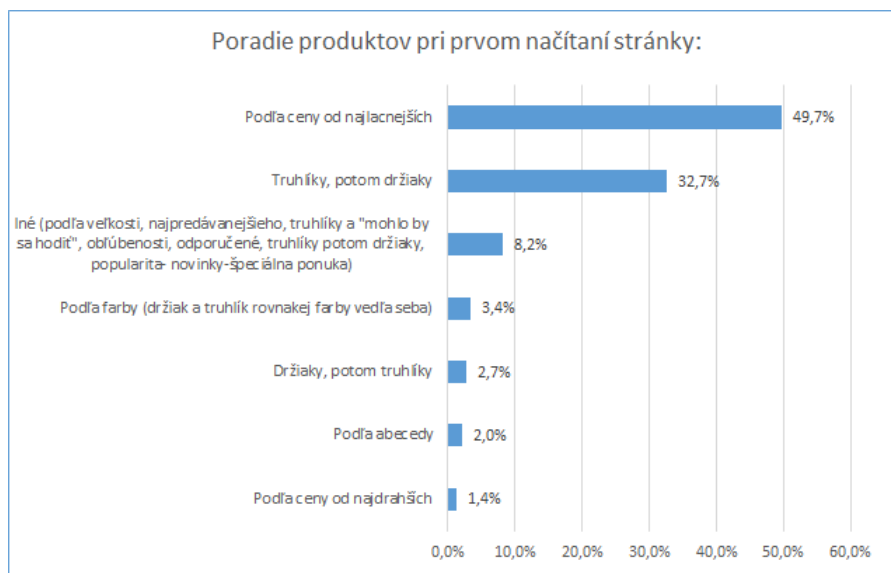
Obr. 5: Vyhľadávanie termínu Truhlík

Z grafu na obr. 5 vidno, že v tejto otázke sa predstavy respondentov výrazne odlišujú, je iba malý rozdiel medzi voľbou produktového (zhlukového) zobrazenia a zobrazenia podľa kategórií. Významovo ekvivalentná je otázka na produktovú radu Fantazie, ktorá obsahuje truhlíky aj držiaky. Podobná z toho dôvodu, že sa opäť vyskytuje otázka, či zobrazenie nastaviť zhlukovo alebo kategoricky. Kategorické zobrazenie znamená minimálne jedno kliknutie navyše. Zoznam produktov by chcelo vidieť 56 % opýtaných, ďalšie delenie by si prišlo 43 % opýtaných. Viz. obrázok 11 v prílohe A.

Keďže pri otázkach typu dotazu na všeobecný tovar alebo názov produktovej rady sú voľby zobrazení od respondentov vyvážené, máme buď možnosť dať im na výber, ktoré zobrazenie by chceli, alebo rozhodnúť za nich. Každé rozhodovanie však človeka podvedome stojí úsilie a energiu, a je jednoduchšie prijať, čo je ponúkané. Navyše podľa prednášky z pokročilých užívateľských rozhraní (Procházka, 2011) treba, aby sa užívateľ dostal k cieľu svojho snaženia čo možno najkratšou cestou a musel vykonať minimálny počet preklikov.

Ďalšou významnou otázkou bolo zobrazenie spomínaných výsledkov pri prvom načí-

taní stránky. Podľa očakávania najviac odpovedí tvorilo zoradenie od najlacnejších. Toto je však z podnikateľského hľadiska neefektívne, a navyše vieme, že väčšina zákazníkov je ochotná toto zoradenie vykonať ručne. Znamená to, že túto možnosť je vhodné poskytnúť na každej zobrazenej stránke pre uspokojenie potrieb, avšak na predvolené zobrazenie bude výhodnejšie zvoliť inú variantu.



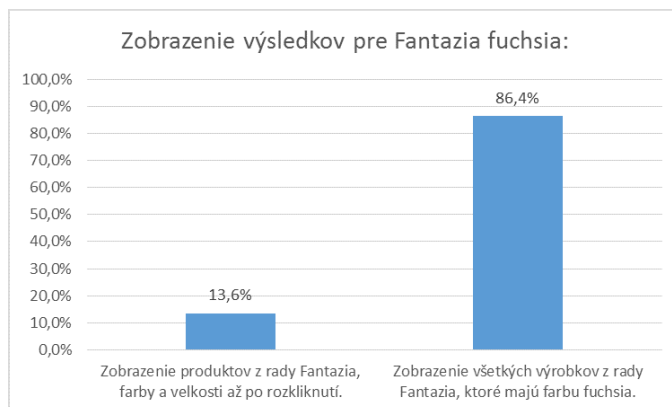
Obr. 6: Predvolené poradie produktov podľa respondentov

Druhá najpočetnejšia odpoveď na otázku z obr. 6 sa ponúka na zobrazenie hlavného tovaru a následne jeho príslušenstva. Za zmienku však stojí aj možnosť *iné*, ktorú sa rozhodlo vyplniť 12 respondentov (8%), ktorá obsahuje voľby ako najpredávanejšie, obľúbené, odporúčané, novinky alebo špeciálna ponuka.



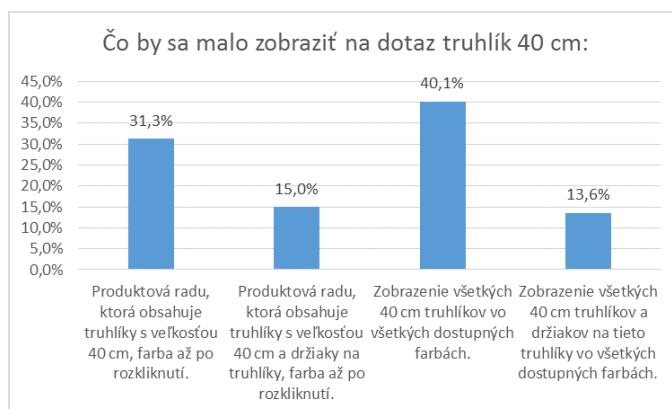
Obr. 7: Predvolené usporiadanie podľa respondentov

Elektronické obchody ponúkajú viacero možností, vďaka ktorým si užívateľ môže zobrazené výsledky usporiadať podľa svojho uváženia. Podľa ceny je najbežnejšia možnosť a usporiadanie podľa veľkosti by bolo v prípade niektorých produktov neuskutočniteľné. Najpredávanejší tovar a novinky sa väčšinou zobrazujú ako prvé. Alebo v špeciálnej lište ako tipy na nákup.



Obr. 8: Vyhľadávanie termínu Fantazia fuchsia

Otázka, ktorá má odpoveď s jednoznačnou prevahou v počte odpovedí je s frázou na vyhľadanie Fantazia fuchsia na obr. 8, teda produktová rada vybranej farby. Možnosti opäť obsahovali výber medzi produktovým zobrazením a zobrazením v kategóriách.



Obr. 9: Vyhľadávanie termínu Truhlík 40

Posledná otázka obsahuje termín truhlík s rozmerom, pričom odpovede nie sú výrazne odlišné od predchádzajúcich. Väčšia časť užívateľov znovu zvolila zobrazenie všetkých truhlíkov vo všetkých farbách, a o 7 % menej respondentov zvolilo možnosť kategorického zobrazenia produktov.

5 Návrh riešenia a implementácia

5.1 Špecifikácia obchodných pravidiel pre vyhľadávanie

Napriek faktu, že každý človek je unikátny, je ľudské správanie predvídateľné. Tvorcovia softvéru sledovali návštevnosť a pozorovali užívateľov roky, vedci a výskumníci skúmali, ako ľudia robia veci a ako premýšľajú nad tým, čo robia. Existujú vzory, ktorými je vhodné sa pri budovaní softvéru inšpirovať. (Tidwell, c2011)

- Viditeľnosť – vyhľadávacie pole umiestnené na mieste, kde ho užívatelia očakávajú. Ideálne nad navigačnú lištu.
- Neustále potešenie – ľudia radi vidia okamžité výsledky vykonaných akcií.
- Uspokojenie – užívatelia kliknú na prvé, čo zbadajú, aj keď to nemusí byť správne. Väčšinou sa uspokojia s prvou fungujúcou variantou a nechcú sa učiť nič nové.
- Zmeny uprostred činnosti – ľudia často menia svoj zámer v priebehu činnosti, preto by jednotlivé časti systému mali byť vzájomne prepojené, aby sa užívatelia aj v strede činnosti mohli rozhodnúť pre zmenu postupu. Napr. stále viditeľné menu či možnosť kedykoľvek sa vrátiť na hlavnú stránku.
- Postupné budovanie – umožniť užívateľom zmenu dizajnu či jazyka, poskytnúť im možnosť prispôbenia.
- Návyk – ľudia sú zvyknutí na využívanie napr. niektorých klávesových skratiek, ktorých význam a funkcionality je vhodné zachovať.
- Priestorová pamäť – užívatelia si pamätajú jednoduchšie miesto, kde je umiestnená hľadaná vec ako jej názov. Pri vylepšeniach dizajnu nie je vhodné meniť polohu významných prvkov, na ktoré sú užívatelia zvyknutí.
- Budúca pamäť – možnosť ponechať určité prvky viditeľné ako pripomienku nejakej budúcej akcie.
- Efektívne opakovanie – niektoré činnosti musia užívatelia opakovať často, preto je dobré vytvoriť danú funkcionality čo najjednoduchšiu a s minimálnym počtom klikov či použitých kláves.
- Rada inej osoby – ide o zdieľanie užívateľskej skúsenosti, o priestor, kde si ľudia môžu vzájomne vymeniť názory, typicky nejaké fórum, alebo komentáre k produktom.
- Osobné odporúčania – ľudia sú veľmi ovplyvnitelní, a radšej dajú na odporúčanie známeho. Vhodné je teda poskytnutie možnosti zdieľania obsahu medzi ďalších ľudí.

Ďalej by webové stránky mali zobrazovať iba jednu časť svojho obsahu – bez komentárov alebo iných prvkov, ktoré môžu narušovať dojem a sústredenie návštevníka. Môže sa tu nachádzať zopár navigačných prvkov, ktoré nepôsobia rušivo, ako sú posuvné lišty, prihlasovacie okno, menu, záhlavie a zápätie. Ideálne je využiť rozloženie obsahu stránky do dlhého, vertikálne posúvateľného zoznamu produktov, textu, obrázkov. (Tidwell, c2011) Tento zoznam môže byť tvorený hierarchicky, ďalším menu, tabuľkou, mriežkou, stromom a pod. K rovnakému výsledku nás dovedli aj odpovede z dotazníka, sami respondenti sú

o niečo viac naklonení k zobrazovaniu všetkých produktov. Na podporu dizajnu je vhodné využiť techniku alternatívnych zobrazení, ktorá umožňuje zobrazenie viacerými spôsobmi. Môže ísť napr. o zjednodušené zobrazenie názvu a obrázku produktu, zobrazenie aj s podrobnosťami, jednoduchý zoznam produktov, úplne bez obrázkov alebo iba so zmenšeninami slúžiacimi k náhľadu.

V svete online nakupovania je kvalita interného vyhľadávania významným faktorom v procese zmeny návštevníka na zákazníka.

Podľa štúdie MarketingSherpa 43 % návštevníkov online predajní tvrdí, že prvá vec, ktorú zadajú do vyhľadávania je názov produktu alebo jeho kategória. Ďalších 39 % využíva navigáciu na stránke, teda menu, záložky a podobne. To znamená, že na detaily okolo vyhľadávania je nutné vynaložiť rovnako veľa času a úsilia ako na dizajn a funkcionality domovskej stránky. Rýchlosť a presnosť výsledkov vyhľadávania spôsobuje aj reálne rozdiely v angažovanosti návštevníka. Tomu trvá priemerne 8 sekúnd, kým sa rozhodne či na stránke zostane alebo ju opustí. (MarketingSherpa, 2007)

Vzhľadom na fakt, že užívatelia majú možnosť si usporiadanie produktov prispôbiť podľa ceny, pravidlá pre vyhľadávanie sa tak týkajú prvotného zobrazenia produktov ihneď po ukončení dotazu. Ak im toto zobrazenie bude vyhovovať, nebudú mať dôvod využívať iné usporiadanie. Je to jedinečná možnosť, ako im navrhnúť, čo by sa im hodilo alebo čo možno hľadali.

Z výsledkov doterajších analýz uvedených v tejto práci vyplývajú všeobecné pravidlá:

- Bude zrušené primárne usporiadanie produktov podľa ceny a identifikačného čísla produktu.
- Položky po vyhľadávaní budú primárne zobrazené podľa relevantnosti k vyhľadávanému výrazu, teda hlavný tovar bude na prvých zobrazených priečkach.
- V prípade nenájdenia frázy bude fráza rozložená na menšie celky. Fráza bude rozdelená na menšie celky tak, aby bolo možné vyhľadať v rámci jedného záznamu kombináciu zadaných slov. V prvom rade bude vyhľadaná zadaná fráza ako celok, v druhom všetky slová obsiahnuté jednotlivito v jednom hľadanom dokumente (zázname), a v treťom prípade ak sa nenájde ani takáto kombinácia, vyhľadá sa každé slovo obsiahnuté zvlášť.
- Vyhľadávanie bude fungovať aj na zadané časti slov, aby sa eliminovalo množstvo výsledkov s informáciou, že daný produkt nebol nájdený. Napríklad: *tru*—*růž*
- Zobrazenie výsledkov vyhľadávania bude tvorené zoznamom položiek, bez delenia na kategórie. Ďalšie delenie by návštevníkov mohlo odradiť, pretože po zadanom výraze očakávajú priamo výsledky.
- Indexované celé slová budú mať minimálnu dĺžku 2 znaky (slovo cm, 40 a pod.). v slovách, ktoré sú dlhšie, sa bude vyhľadávať reťazec s minimálnou dĺžkou 3 znaky. Pri 3 písmenách je už určitá šanca, že budú označovať celý produkt alebo aspoň časť jeho názvu. (Napríklad názov Ufo.)
- Bude možné vyhľadávať aj podreťazce a frázy bez interpunkcie.
- Bude umožnené zobrazenie správneho výsledku pre vyhľadávanú frázu s preklepom.

5.2 Implementácia

Pred začatím popisu implementácie si pripomeňme hlavné ciele práce, čo nám optimalizácia vyhľadávania umožní a čo by mala rešpektovať. Ako bolo uvedené v kapitole 1.2, na základe analýzy prístupov, vlastného prieskumu užívateľských preferencií a rozboru súčasného stavu vyhľadávania boli vytvorené pravidlá, ktoré slúžia ako podklad pre implementačnú časť práce.

Z návštevnosti sledovaného elektronického obchodu podľa tabuľky 1 vidíme, že zákazníci využívajúci vyhľadávanie sú viac odhodlaní k nákupu – majú nižšiu mieru opustenia stránok, vyšší počet prehliadaných stránok a dlhší čas na nich strávený. Majú presnejšiu predstavu o hľadanom produkte ako iný návštevník, ktorý produkty iba prehliada.

Tiež z webovej analytiky vieme, že naše zameranie by malo byť vyššie na názvy produktových rád, ktoré sú vyhľadávané najčastejšie. Patria zároveň medzi produkty s najvyšším prínosom, o čom hovorí ukazovateľ hodnota za vyhľadávanie.

Z dotazníka sme ďalej zistili, že užívatelia chcú pri použití vyhľadávania zobrazené produkty vidieť radšej v dlhšom zozname, ako by sa mali preklikávať ďalším delením na kategórie.

Rozbor súčasnej situácie poskytol informáciu o predradenom filtri zobrazenia, ktorý produkty usporiadal vždy podľa cien od najlacnejších. Na prvý pohľad je to nastavenie vyhovujúce zákazníkom. Pri určitých produktových radoch, ktoré majú vysokú variabilitu v cenách produktov sa tak ľahko stane, že produkty rovnakého typu budú zobrazené na rôznych stránkach. Napríklad truhlíky rozmanitých cien a typov budú mať medzi sebou držiaky. To môže byť miestami neprehľadné. Samozrejme usporiadanie by malo vyhovovať aj predajcovi a byť v súlade s predajnou stratégiou. Toto tvrdenie podporuje aj fakt, že vyhľadávajúci zákazník je odhodlanejší. V prípade záujmu si produkty usporiada podľa vlastného uváženia.

Prínosom práce bude vylepšenie vyhľadávacieho algoritmu podľa pravidiel špecifikovaných v kapitole 4.4. Tieto pravidlá vznikli v nadväznosti na odhalenie problémov, ktoré má súčasné vyhľadávanie v kombinácii s výsledkami z predošlých analýz. Súčasnú vyhľadávanie hľadá celé frázy a časti slov. Ak užívateľ zadá kombináciu slov, ktoré sa nenachádzajú v jednom zázname, dostane ako výsledok informáciu o neúspešnom hľadaní.

V prípade vylepšeného algoritmu bude užívateľ informovaný o nenájdení zadanej kombinácie. v rámci eliminácie prázdnych výsledkov mu budú zobrazené aspoň výsledky, ktoré zodpovedajú zadaným slovám samostatne. Bude upravené usporiadanie produktov a vytvorená korekcia zadaných slov v prípade preklepov. Podrobnejší popis a ďalšie vylepšenia budú uvedené v tejto kapitole.

5.2.1 Prostredie, príprava a kódovanie

Aby bolo možné optimalizovať vyhľadávací algoritmus, muselo byť nastavené súčasné prostredie vyhľadávania. Na vytvorenie simulovaného prostredia reálneho e-shopu pre vývojové účely je využitý WampServer 2.5 s predinštalovaným webovým serverom Apache 2.4.9. Využívaná je databáza MySQL 5.6.17, s ktorou sa pracuje za pomoci skriptovacieho jazyka PHP 5.5.12. Na indexovanie a vyhľadávanie je využívaný nástroj Sphinx 2.2.5 s podporou databázy MySQL.

Zdrojové dáta poskytnuté zadávateľom boli uložené v XML formáte v českom jazyku a doplňujúce informácie s podrobnosťami o produktoch boli poskytnuté v súbore XLSX

v anglickom jazyku. Pre potrebu kombinácie dát z oboch zdrojov bolo nutné tieto dva zdroje spojiť a potrebné položky preložiť z anglického jazyka do českého a následne dáta nainportovať do databázy. Vzhľadom na fakt, že Sphinx vie pracovať iba s údajmi v kódovaní UTF-8, bolo potrebné importovaný súbor nastaviť do tohto formátu, v HTML kóde nastaviť toto kódovanie, v časti PHP definovať pripojenie na databázu v UTF-8 kódovaní, a využiť funkciu

```
mysqli_real_escape_string($conn, $_GET["value"])
```

na definovanie, že zadávaná fráza je v rovnakom kódovaní ako zvyšné údaje, teda UTF-8. Sphinx zatiaľ natívne podporuje iba tabuľku znakov k anglickému a ruskému jazyku, preto bolo v konfiguračnom súbore treba znova špecifikovať, o aké kódovanie sa jedná

```
sql_query_pre = SET CHARACTER_SET_RESULTS = utf8
sql_query_pre = SET NAMES utf8
```

a definovať tabuľku znakov pre českú diakritiku. Toto nastavenie je dôležité pre správne indexovanie. V konfiguračnom súbore sa dá pre nastavenie indexovania nastaviť minimálna dĺžka indexovaného slova. Tá bola nastavená na dva znaky. Bez tohto nastavenia by sa nedali nájsť produkty dvojciferných rozmerov (napr. truhlík veľkosti 40, 60 cm). Ide o samostatné slová dĺžky dva znaky. Ďalším parametrom je dĺžka prefixu – definovaná na tri znaky. Tá hovorí o tom, koľko znakov musí byť zadaných, aby sa hľadali podreťazce slov. Ďalej sa dá nastaviť používanie zástupných znakov vo vyhľadávaní, čo by ale vyžadovalo určitú pokročilú znalosť práce užívateľov s vyhľadávačmi. z praxe je jasné, že bežný návštevník by takúto funkcionality nevyužíval, preto je efektívnejšie implementovať tieto výhody do kódu a nezaťažovať ho ďalšími možnosťami. Predpokladá sa teda, že presne použité znaky od užívateľa do vyhľadávania sú to, čo hľadá.

Samotné indexovanie prebieha tak, že indexer všetky znaky, ktoré nepozná, označí za rozdeľovače. Sú to znaky, ktoré nemá definované v tabuľke znakov v konfiguračnom súbore. Pri jazyku obsahujúcom diakritiku to samozrejme znamená absolútne nesprávne indexovanie, a vyhľadávanie funguje iba na zadávané slová a frázy v anglickej abecede a-Z.

Pred spustením indexovania si treba premyslieť, ktoré atribúty majú byť indexované. v zaindexovaných atribútoch potom bude prebiehať vyhľadávanie a ich hodnoty sa budú ponúkať ako výsledky. Ďalej si treba uvedomiť, podľa ktorých atribútov bude nastavené usporiadanie výsledkovej listiny. Takéto funkcie ponúka priamo Sphinx, ale vie používať iba zaindexované atribúty. Ak nechceme, aby v nich prebiehalo vyhľadávanie, treba ich definovať zvlášť za pomoci premenných *sql_attr*. Môže to byť napríklad bezznamienkový integer, teda nezáporná číselná hodnota alebo string, teda reťazec znakov.

Nasledujúca časť kódu určuje, ktoré atribúty majú byť vybrané k práci so Sphinxom. Premenná *sql_query* určuje atribúty určené k zaindexovaniu. Id je povinná hodnota identifikátora záznamu v databáze, aby sa pri vyhľadávaní pomocou Sphinxu dalo zistiť, ktorý záznam má byť vrátený. Ďalšie atribúty sú priamo názvy stĺpcov z databázy, teda označenie zdroja hodnôt. CODE je interný kód produktu, PRODUCTNAME je celý názov produktu, DESCRIPTION je popis produktu, CATEGORY je kategória produktu, PRICE je cena, URL je webová adresa a EAN je číslo produktu.

```
sql_query = SELECT id, CODE, PRODUCTNAME, DESCRIPTION, CATEGORY,
PRICE, URL, EAN FROM produkty
```

```
sql_attr_uint = PRICE
```

```
sql_attr_string = CATEGORY
```

To znamená, že zaindexované atribúty zúčastňujúce sa vyhľadávania sú kód produktu, popisný text, kategória, cena, url adresa a EAN kód. Pritom v cenách a kategóriách sa nedá vyhľadávať, slúžia iba pre usporiadanie.

5.2.2 Základný princíp optimalizovaného algoritmu

Fungovanie algoritmu je najlepšie popísať pomocou vývojového diagramu. Na začiatku vyhľadávania musí byť zadaná fráza (teda slovo, slová alebo slovné spojenie), ktoré si užívateľ praje vyhľadať. Predpokladáme, že čo bude zadané, to by malo byť vyhladané.

Môžu nastať prípady, kedy bude zadané jedno slovo, iba časť slova, viac slov. Ak je tam viac slov, môžu nastať dva prípady. v prvom ide o doslovnú frázu, kedy potrebujeme vyhľadať presnú zhodu medzi zadávaným a vyhladaným výrazom. v druhom prípade ide o slová, ktoré popisujú produkt a jeho vlastnosti, pričom nezáleží na ich poradí, sú na sebe nezávislé. Zadávaná fráza môže mať ľubovoľný počet slov, vždy je tu ale možnosť nejakého preklepu, prípadne otázka vyhľadávania s ohľadom na diakritiku.

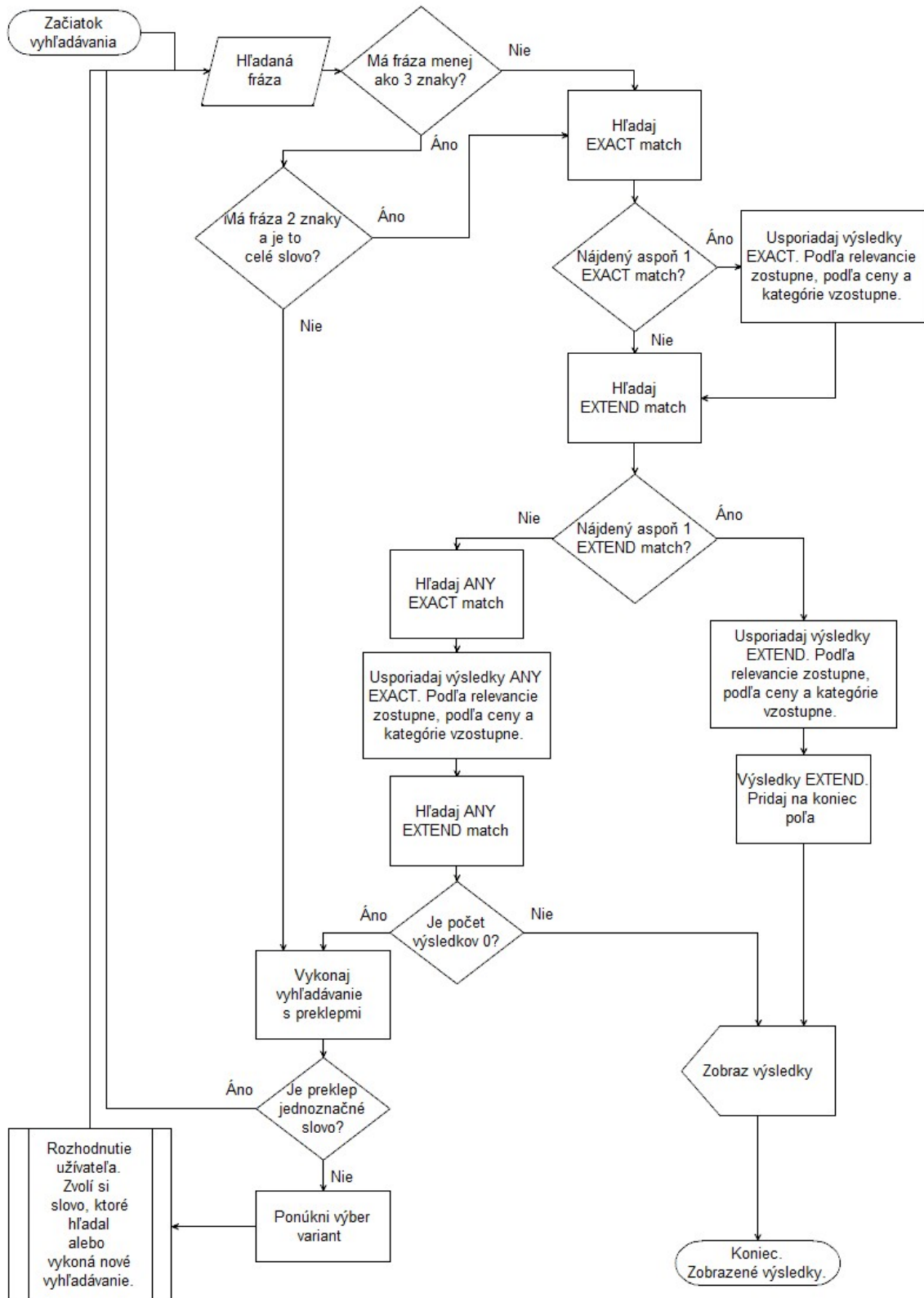
Podľa diagramu zobrazeného nižšie na obrázku 10 vidíme, že vyhľadávanie začína zadaním hľadanej frázy. Predpokladáme, že užívateľ hľadá konkrétny produkt – zadáva primárne jeho názov, alebo slová, ktoré sú preň charakteristické. Teda nastáva prípad, že je zadané slovo minimálnej dĺžky 3 znaky. Najskôr chceme úplne vyhovieť požiadavkám, takže ako prvé sa vyhľadá doslovná zhoda (exact match) frázy presne v zadanom tvare. v prípade, že sú nájdené výsledky, budú usporiadané a uložené do poľa podľa parametrov významnosti, ceny a kategórie (podrobnejší popis v kapitole 5.2.3).

Ak neboli nájdené žiadne výsledky, nemá sa čo usporiadať. Táto možnosť môže nastať, ak ide o preklep, alebo o zadanú *iba časť slova*. Obe vetvy sa znova spoja v nasledujúcom kroku.

Vyhľadávanie pokračuje nájdením rozšírených možností (extend match). Medzi slová frázy sú vložené zástupné znaky pre označenie ľubovoľného počtu znakov. To je z dôvodu, aby sa nehľadala doslovná fráza, alebo boli prípustné možnosti, kedy sa zadané slová nachádzajú samostatne v indexovaných atribútoch. Stále však charakterizujú jeden konkrétny produkt. Tento zástupný znak však môže znamenať aj iný tvar slova, napríklad možnú predponu alebo príponu hľadaného slova. Napr. to môže byť pre zadaný „truhlík“: (držiak) truhlík-u.

Ak boli nájdené aj výsledky z tohto rozšíreného hľadania, budú usporiadané podľa zadaných parametrov v algoritme. Keďže sa podarilo nájsť užívateľom zadanú frázu v podobe minimálne jedného produktu, zadanie je splnené a nič viac hľadať netreba. Zobrazia sa výsledky a vyhľadávanie skončí.

V prípade, že žiadne výsledky neposkytne ani *exact*, ani *extend*, prechádzame k ďalšej časti. Budeme hľadať tzv. *any exact*. Pri jednoslovnej fráze to nemá veľký význam, chová sa rovnako ako základný *exact* a následne *extend*. Podstata funkcionality je využiteľná až pri viacslovných zadaniach. Hľadá sa totiž doslovná zhoda ktoréhokoľvek slova zo zadania so slovami v názve, popise a ďalších atribútoch, v ktorých prebieha vyhľadávanie. Ak existujú vyhovujúce výsledky, sú usporiadané a ďalej sa hľadá *any extend*, teda ktorékoľvek slovo rozšírené o možné okolité znaky. Časť algoritmu obsahujúca *any* poskytne výsledky, ak zadaná fráza obsahuje napríklad druh produktu a farbu, pričom taká kombinácia neexistuje. Užívateľ tak dostane informáciu o nenájdenej presnej zhode, dostane však aspoň



Obr. 10: Vývojový diagram popisujúci základné fungovanie algoritmu

výsledky, ktoré čiastočne vyhovujú jeho zadaniu. Pri nenulovom počte výsledkov budú nájdené produkty zobrazené a vyhľadávanie tak úspešne končí.

Znova ale môže nastať situácia, že výsledky ani za pomoci *any* nebudú nájdené. Prichádzame tak k zadanému slovu, v ktorom užívateľ spravil preklep alebo hľadá slovo charakterizujúce neexistujúci produkt. Súčasne sem spadá aj vetva zo začiatku, ak je slovo kratšie ako 3 znaky. Vykoná sa vyhľadávanie s preklepmi. Detekcia preklepu je zúžená iba na hľadanie v atribúte názvu produktu. v závislosti od množstva chýbajúcich či nesprávne usporiadaných písmen v slove a na pozícií tejto chyby môže ísť o preklep s jednoznačnou detekciou zamýšľaného slova, alebo spornou. Pri jednoznačnom určení slova sa opraví zadávaná fráza a zobrazia sa výsledky vyhľadávania.

Ak ide o sporný preklep a mohlo by sa jednať o viac rôznych slov, pričom algoritmus nie je schopný detekovať, ktoré je správne, užívateľ dostane na výber z potenciálne hľadaných slov. Dalo by sa povedať, že ide o našepkávanie. Užívateľ tak môže kliknúť na slovo, ktoré mal na mysli, a následne sa vykoná vyhľadanie opravenej frázy zopakovaním celého algoritmu.

Prakticky je tak možné dosiahnuť prázdnu výsledkovú množinu až pri zložitých preklepoch alebo neexistujúcich slovách.

5.2.3 Vyhľadávacie módy a možnosti usporiadania

V kapitole 5.2.2 bol popísaný základný princíp fungovania optimalizovaného algoritmu, môže byť teda plynule pristúpené ku konkrétnemu prevedeniu.

Základom je do premennej *\$phrase* uložiť užívateľom hľadanú hodnotu, vytvorí pripojenie na databázu so správnym kódovaním a umožniť využívanie predpripravených funkcií Sphinxu pomocou súboru sphinxapi.php.

Tieto funkcie (módy) charakterizované v kapitole 2.5.3 ponúkajú veľa možností a kombinácií vyhľadávania komplikovaných výrazov, ktoré by ale musel užívateľ vedieť správne zapísať. To by neprinieslo požadované zjednodušenie. Preto sú v algoritme zohľadnené niektoré z týchto možností, ktoré bolo možné zapracovať do vyhľadávania bez vedomia užívateľa.

Funkcia *MainAlgorithm* je hlavnou časťou zdrojového kódu, ktorá sa stará o správnu postupnosť vykonávania vyhľadávacích sekvencií uvedených v diagrame na obrázku 10. V parametri preberá inštanciu definovaného *SphinxClienta*, ktorý slúži na sprístupnenie funkcií Sphinxu. Druhým parametrom je vyhľadávaná fráza presne tak, ako bola zadaná užívateľom do vyhľadávacieho poľa.

MainAlgorithm využíva ďalšiu vlastnú funkciu a tou je *QuerySpecificMode*. Určuje, ktorým vyhľadávacím módom má byť fráza vyhľadaná, ako majú byť usporiadané výsledky a v prípade potreby definuje úpravu frázy o zástupné znaky. V parametri preberá inštanciu klienta, názov módu, v ktorom má vyhľadávať, hľadanú frázu a boolean hodnotu určujúcu, akým spôsobom sa má pracovať s hľadaným reťazcom.

```
QuerySpecificMode($sphinxclient, "exact", $searchedPhrase, true);
```

Jej náplňou je vrátiť usporiadané pole identifikátorov záznamov, ktoré vyhovujú definovanému vyhľadávaniu. Hlavným dôvodom vzniku bola potreba v priebehu vyhľadávania meniť jeho módy. Využívame pri tom SPH_MATCH_PHRASE, SPH_MATCH_EXTENDED2 a SPH_MATCH_ANY. Táto funkcia je rozdelená príkazom *case* na „exact“, „extend“ a „any“.

Rovnako ako módy je tu možnosť nastaviť usporiadanie produktov. Ak by pri každej zmene vyhľadávacieho módu mala nastať aj zmena poradia zobrazenia produktov, malo by to neželaný efekt na užívateľa.

Preto je usporiadanie produktov principiálne v každej časti rovnaké. V prvom rade podľa váhy ako interného atribútu Sphinxu zostupne. Zohľadňuje, koľkokrát sa hľadaná fráza nachádza v indexovaných stĺpcoch v rámci jedného záznamu.

V druhom rade je usporiadanie zoradené podľa ceny produktu zostupne. Táto voľba vychádza z princípu kontrastov, ktoré popísal Cialdini vo svojej knihe. (Cialdini, 2012) Keď zákazníkovi ukážeme najskôr najdrahšie produkty, nastaví si podvedome istú hranicu vnímania ceny. Ak sa potom pozrie na iné produkty, ktoré by sa mu inokedy zdali byť predražené, nebude túto sumu už vnímať tak negatívne.

V poslednom rade sú produkty usporiadané podľa kategórie vzostupne. Keďže pri niektorých otázkach výskumu respondenti odpovedali, že by prijali ďalšie delenie na hlavné kategórie a až potom videli zoznam produktov (napríklad obr. 8), do voľby usporiadania bol vložený tretí parameter kategória, podľa ktorého budú radené produkty rovnakej cenovej hladiny vzostupne. Tým by sa výsledky mali zobrazovať v pomyselných podskupinách, významovo podobné blízko seba.

V časti *exact* funkcie *QuerySpecificMode* je hľadaná doslovná fráza zadaného výrazu tak, ako bola prebraná parametrom zo vstupu od užívateľa a nastaví sa usporiadanie produktov.

```
case "exact":
    $sphinxclient->SetSortMode(SPH_SORT_EXTENDED, "@weight DESC
    , PRICE DESC, CATEGORY ASC");
    $sphinxclient->setMatchMode(SPH_MATCH_PHRASE);
    $res = $sphinxclient->Query($searchedPhrase);
    break;
```

Časť *extend* má iný mód a je obohatená o úpravu reťazca zástupným znakom *, ktorý nahradzuje ľubovoľný počet znakov.

```
$sphinxclient->setMatchMode(SPH_MATCH_EXTENDED2);
$changedPhrase = str_replace(' ', '* ', $searchedPhrase);
$changedPhrase = "*" . $changedPhrase . "*";
$res = $sphinxclient->Query($changedPhrase);
```

V prípade že je to jedno slovo základného tvaru *truhlík*, stane sa z neho **truhlík**, a teda zahŕňa iné tvary slova, napr. truhlíku alebo truhlíky. V prípade, že je zadaná iba časť slova *tru* toto nastavenie zabezpečí vyhľadanie celého slova obsahujúceho zadaný podreťazec. Ak fráza obsahuje viac slov, medzera, ktorá ich oddeľuje je nahradená dvoma * s medzerou. Každé slovo tak bude rozšírené o potenciálne ďalšie znaky z oboch strán. Zmena vyhľadávacieho módu zabezpečí, že slová frázy nemusia byť umiestnené vedľa seba, stačí ak budú v jednom zázname.

Posledným prípadom je *any*, kedy využívame možnosť vyhľadania ktoréhokoľvek slova zo zadaného výrazu. Vyhľadávací mód je *SPH_MATCH_ANY*. (Ekvivalentná úprava by bola pri použití módu *SPH_MATCH_EXTENDED2*, pričom medzera by musela byť nahradená znakom „—“, ktorý označuje OR.)

Táto vetva vyhľadávania sa člení na *any exact* a *any extend*. Rozdiel je určený boolean parametrom *\$stringType*. Ak je jeho hodnota *false*, ide o *any exact* a ak je hodnota *true*,

jedná sa o *any extend*. Vyhľadávanie funguje principiálne rovnako ako vo vyššie popísaných prípadoch, v *extend* je fráza doplnená o zástupný znak.

```
$sphinxclient->setMatchMode(SPH_MATCH_ANY);
if ($stringType) {
    $changedPhrase = str_replace('_', '*_*', $searchedPhrase)
    ;
    $changedPhrase = "*" . $changedPhrase . "*";
    $res = $sphinxclient->Query($changedPhrase);
} else {
    $res = $sphinxclient->Query($searchedPhrase);
}
```

MainAlgorithm teda využíva vyššie spomínanú funkciu a módy, ktoré zabezpečujú vyhľadanie tých správnych údajov.

Využívame prácu s poľami. Na začiatku si jedno prázne pole definujeme, a následne sa ho snažíme pomocou vyššie uvedených funkcií naplniť správnymi hodnotami. Začneme postupom popísaným v diagrame na obr. 10, zavoláme funkciu s vyhľadávaním doslovnej frázy.

```
$pomResult = QuerySpecificMode($sphinxclient, "exact",
    $searchedPhrase, true);
```

Ak boli nájdené výsledky, zavolá sa rozšírené hľadanie *extend*. Tieto výsledky sa uložia do iného poľa. Keďže výsledky *extendu* nájdu aj rovnaké výsledky ako *exact*, je potrebné zbaviť sa duplicitných záznamov.

Na to sa využíva funkcia *removeDuplicities*, ktorá preberá v parametri dve inštancie poľa – s výsledkami vyhľadávania režimu *exact* a *extend*. Výsledné pole sa naplní hodnotami z *exactu*. v dvoch vnorených cykloch sa potom prechádzajú obe polia z parametrov, pričom ak sa záznamy zhodujú, pokračuje sa ďalej. Ak sa zistí, že sú záznamy rozdielne, do výsledného poľa sa na koniec pridá nový záznam, ktorý tam zatiaľ obsiahnutý nie je.

V režime *extend* tak boli nájdené výsledky, ktoré nie sú presné, ale základ zadaného slova je rovnaký, teda sú stále pre užívateľa relevantné. Nie sú však hodné umiestnenia na prvé pozície vo vyhľadávaní. *MainAlgorithm* tak vráti výsledky a vyhľadávanie je ukončené.

Vráťme sa o odstavce vyššie. Režim *exact* žiadne výsledky nenašiel, v tom prípade nájdeme a zobrazíme výsledky *extendu*. Ak však ani *extend* nič nenájde, znamená to, že fráza obsahuje slová, ktoré sa nenachádzajú v jednom zázname alebo je zadané neexistujúce slovo či preklep. V tejto časti počítame s možnosťou kombinácie napr. produktu a farby, ktoré sa v takejto dvojici nenachádzajú v ponuke. Pristupujeme teda k vyhľadaniu každého slova samostatne pomocou *any*. Pri tom by návštevník mal byť upozornený, že to, čo hľadá, neexistuje v presne zadanej kombinácii, ale mohlo by ho to aj tak zaujímať, pretože výsledky aspoň čiastočne vyhovujú jeho zadaniu. Hľadáme tak *any exact* a *any extend*. Odstránime prípadné duplicity, alebo zobrazíme iba výsledky z *any extend*.

Tým je úspešne ukončené vyhľadávanie slov a fráz vrátane častí slov, ktoré sú korektné zadané, teda hľadané slová naozaj existujú, dajú sa vyhľadať a usporiadané sa zobrazia užívateľovi.

Návratovú hodnotu funkcie tvorí pole usporiadaných identifikátorov záznamov vyhovujúcich zadaniu návštevníka. Nad týmito záznamami sa následne vykoná dotaz do databázy a zobrazia sa ľubovoľné atribúty záznamov s nájdenými identifikátormi.

5.2.4 Frázy obsahujúce preklepy

Poslednou možnosťou, ako zabrániť zobrazeniu informácie o neúspešnosti vyhľadávania, je detekcia preklepov. Tá je obsiahnutá vo funkcii *isFalseString*. Využitie jej funkčnosti nastáva až po neúspešnom vyhľadávaní počas hlavného algoritmu. Podstatou tejto metódy je nájsť existujúce slovo z názvu produktu, ktoré by mohlo vyhovovať ako zadávané slovo bez preklepov. a ak je týchto slov viac, ponúknuť užívateľovi možnosť výberu toho správneho, ktoré mal na mysli.

Ako každá funkcia, aj táto preberá v parametri frázu pre vyhľadanie. s tou sa následne pracuje na úrovni podreťazcov, pričom je zohľadnená možnosť viacslovného zadania.

Na začiatku sa musí fráza rozdeliť na slová, tie sa priradia do poľa. Pracuje sa s každým slovom zvlášť, preto sa bude vyhľadávať v cykle „pre každé slovo hľadanej frázy vykonaj“. Pre potreby ďalšej práce so zadanou frázou je ju nutné najskôr zbaviť diakritiky. Znaký kódovania UTF8 sa zkonvertujú na znaky ASCII. Pritom zo znaku s diakritikou vznikne znak ' a písmeno (č → 'c), preto všetky výskyty znaku ' musia byť nahradené prázdny reťazcom. Potom je potrebné zistiť dĺžku slova. Podľa tejto hodnoty budú z hľadaného slova vytvorené dva alebo tri podreťazce. Hranica určujúca počet podreťazcov je tvorená slovami s dĺžkou 6 znakov.

```
if ($length < 6) {
    $string1 = substr($item, 0, 3);
    $string2 = substr($item, -3);
} else {
    $string1 = substr($item, 0, ($length/2));
    $string2 = substr($item, -($length - ($length/2)));
    $string3 = substr($item, 2, ($length-4));
}
```

Ak je slovo kratšie ako 6 znakov, vytvoria sa podreťazce z prvých troch písmen a z posledných troch písmen. Tri znaky sú minimum pre vyhľadanie časti indexovaného slova a zároveň už môžu tvoriť základ určitého slova. Pri hľadaní iba dvoch znakov by týchto možností mohlo byť príliš veľa a zahŕňali by aj nerelevantné výsledky.

Ak je slovo **dlhé aspoň 6 znakov**, vytvoria sa tri substringy – od začiatku do polovice slova, od konca do polovice slova a tretí bude pre presný stred slova tak, aby na začiatku aj na konci dva znaky zostali voľné. Pri rozdelení hľadaného slova na podreťazce sa predpokladá, že aspoň jedna časť slova bude bez preklepu.

Na základe vytvorených podreťazcov (začiatku a konca) hľadaného slova s preklepom sa dotazom priamo do databázy cez SQL zisťuje ich existencia v názve produktu.

```
$results1 = $conn->query('select PRODUCTNAME from produkty where
PRODUCTNAME LIKE "%'.$string1.'%"');
```

Ak existuje hľadaný podreťazec v slove z názvu produktu, je identifikátor tohto záznamu uložený do poľa s potenciálne hľadaným produktom. Vzniknú teda polia s výsledkami, ktoré vyhovujú podreťazcu začiatku a konca slova.

```
if (($results1->num_rows > 0) && ($results2->num_rows > 0)){
    $results3 = $conn->query('select PRODUCTNAME from produkty
where PRODUCTNAME LIKE "%'.$string3.'%"');
    while($product1 = $results1->fetch_assoc()){
```

```

while($product2 = $results2->fetch_assoc()){
    if (strcasecmp($product1["PRODUCTNAME"], $product2["
PRODUCTNAME"]) == 0 ) {
        $isSame = true; //overime, ci najdene slovo so
        substringom zaciatku a konca je rovnake->preklep
        je v strede
        break;
    } else { //ak zaciatok a koniec nenasli rovnake slovo,
        tak to asi obsahuje preklep
        while($product3 = $results3->fetch_assoc()){
            if (strcasecmp($product1["PRODUCTNAME"], $product3["
PRODUCTNAME"]) == 0 ) {
                $isSame = true; //zaciatok a stred dava spravne
                slovo, preklep je v poslednej casti
                break;
            } else if (strcasecmp($product2["PRODUCTNAME"],
                $product3["PRODUCTNAME"]) == 0) {
                $isSame2 = true; //koniec a stred dava rovnake
                slovo, preklep je na zaciatku
                $pomWord = $product2;
                break;
            }
        }
    }
}

if ($isSame) { //ak je preklep v strede slova alebo na
    konci, do vysledneho pola prirad slovo najdene
    zaciatocnym substringom
    $results[] = $product1;
    $isSame = false;
} else if ($isSame2){ //ak je preklep na zaciatku, prirad
    do vysledneho pola slovo najdene substringom konca
    $results[] = $pomWord;
    $isSame2 = false;
}
}
}

```

Preklep v strede hľadaného slova

V prípade nájdenia minimálne jedného názvu produktu, ktorý obsahuje podreťazec začiatku alebo konca hľadaného slova sa vykoná rovnaký dotaz do databázy aj pre podreťazec stredu hľadaného slova. Potom sa porovnávajú polia produktov, ktoré vyhovujú výsledkom, na obsah podreťazcov začiatku a konca slova. V prípade, že sa jedná o rovnaký produkt to znamená, že preklep bol spravený v strede slova a navrhované slová sú jednoznačné.

Preklep na konci hľadaného slova

Ak sa ale podreťazce začiatku a konca hľadaného slova nenachádzajú v rovnakom produkte, znamená to, že niektorá časť obsahuje preklep a je potrebné vykonať ešte kontrolu nad názvami produktov obsahujúcimi podreťazec stredu slova. Takže sa porovnávajú názvy produktov obsahujúce podreťazec začiatku a stredu slova. Ak sú rovnaké, bol nájdený správny produkt a preklep je odhalený v poslednej časti slova.

Preklep na začiatku hľadaného slova

Zostáva posledný prípad, porovnať koniec a stred. Ak dáva rovnaké produkty ako výsledky, preklep bol odhalený na začiatku zadávaného slova. Do výsledného poľa je priradený produkt nájdený vďaka začiatočnému podreťazcu pretože preklep bol objavený v strede alebo na konci slova. Ak bol preklep na začiatku slova, do výsledného poľa sa priradí produkt nájdený pomocou podreťazca konca hľadaného slova.

U slov **kratších ako 6 znakov** sa vyhľadáva iba podľa rutín hľadajúcich chybu na začiatku a konci slova. Ak sa dotazom do databázy našli produkty, ktoré obsahujú podreťazec počiatku slova, priradia sa ako potenciálne hľadané produkty. Rovnako je to aj pre podreťazec na konci slova.

```
else if ($results1->num_rows > 0) { //ak sa v db nasli vysledky
    iba s pociatocnym substringom, tak vypis tieto
while($product1 = $results1->fetch_assoc()){
    $results[] = $product1;
}
```

Vo všetkých prípadoch platí, že v poli s výsledkami sú teraz uložené produkty, ktoré vo svojom názve môžu obsahovať hľadané slovo po detekcii preklepu. Úlohou však je ponúknuť užívateľovi konkrétne slovo, ktoré pripadá do úvahy ako zamýšľané pre vyhľadávanie. Je to vlastne návrh užívateľovi typu: „Mali ste na mysli...?“. Je teda potrebné z názvu produktu extrahovať to správne slovo, ktoré bude týmto návrhom. Je nutné porovnať slová z názvu produktu s navrhovaným slovom už bez preklepov.

- + Ak je v týchto slovách rozdiel, slovo z názvu produktu bude zbavené diakritiky a prebehne kontrola, či je podreťazec začiatku alebo konca užívateľom hľadaného slova obsiahnutý v slove z produktu.
 - + Pokiaľ je medzi slovom z názvu produktu a vyhľadávaným slovom zhoda, bude produkt pridaný do výsledkov vyhľadávania zobrazených užívateľovi.
 - Ak rozdiel medzi slovami z názvu produktu a navrhovaným bez preklepov neexistuje, znamená to, že už sa medzi navrhovanými výsledkami nachádza.

Finálne pole navrhovaných slov pre užívateľa je vypísané pomocou radiobutonov. Keď užívateľ vyberie jedno slovo z ponuky, jeho znenie sa skopíruje do vyhľadávacieho políčka a po stlačení tlačidla *vyhladať* alebo potvrdení klávesou *enter* sa spustí znova fulltextové vyhľadávanie Sphinxu.

6 Testovanie

Testovania sa zúčastnili osoby, ktoré skúmaný e-shop ani ponúkané portfólio produktov nepoznali. Vekovo spadali do kategórie 20-50, vysokoškolsky vzdelaní. Zvolená vzorka bola vybraná podľa najpočetnejších skupín respondentov z dotazníka.

Testovanie inovovaného algoritmu prebiehalo v dvoch fázach. v prvej fáze bol nezainteresovaným osobám v stručnosti predstavený e-shop, ktorý slúžil ako zdroj dát. Mali poskytnutý priestor na to, aby sa sami zoznámili s prostredím webových stránok a prezreli si produktové portfólio. Bolo im tak umožnené preskúmať produkty, ktoré následne mohli vyhľadávať. Na testovanie funkcionality dostali prístup k webovému obchodu s vyhľadávaním a k stránke na localhoste s novým algoritmom, aby mohli pri paralelnom vyhľadávaní porovnávať zobrazované výsledky.

S každým užívateľom sa pracovalo samostatne. Najskôr bol užívateľ ponechaný, aby samovoľne otestoval oba spôsoby vyhľadávania, na ľubovoľných slovách. Potom bol upozornený na veci, ktoré by si mal všimnúť, ako je usporiadanie produktov, zobrazenie relevantných výsledkov na konkrétnych pozíciách, na tvary zadávaných slov, na používanie fráz alebo iba slovných spojení, či využívanie diakritiky. Napriek tomu, že sa s osobami zúčastnenými testovania pracovalo samostatne, postupovali v týchto veciach rovnako. Nikto svoju pozornosť neupriamil na usporiadanie produktov. Na to reagovali až po upozornení.

Pri hľadaní celých a existujúcich slov si nevšimli žiadny rozdiel, vyhľadávanie funguje z užívateľského hľadiska rovnako v oboch testovaných prostrediach. v novom algoritme je ale zmenené poradie zobrazovaných výsledkov podľa relevancie a ceny zostupne.

Ďalším prípadom bolo zadanie frázy, ktorá ale nemala správny tvar slov. Napríklad hľadanie výrazu *truhlík hnědý*. Užívateľ v obchode dostane odpoveď, že sa daný produkt nenašiel. Pritom produkt existuje pod názvom *truhlík hnědá*.

Súčasný vyhľadávací algoritmus produkt nenájde ani pri hľadaní samostatného slova *hnědý*, ale iba pri zadaní samostatného slova *truhlík*. To je však uspokojujivé iba čiastočne, pretože užívateľ má síce produkt, ale stále nemá požadovanú farbu. s pribúdajúcim časom stráveným vyhľadávaním rastie pravdepodobnosť, že stránku opustí.

Inovovaný algoritmus tento problém rieši, robí to nasledovným spôsobom. Najskôr hľadá frázu, rozšírené vyhľadávanie, potom aspoň jedno slovo, a nakoniec preklep. v spomínanom prípade sú zobrazené výsledky slova zadaného bez chyby, teda pre slovo *truhlík*. v prípade, že je zadané slovo s preklepom, dostane užívateľ upozornenie na túto skutočnosť, môže zvoliť z vybraných možností, čo chcel napísať, a vyhľadávanie pokračuje.

Vyhľadávanie využívané v súčasnosti neumožňuje užívateľovi vidieť vyhľadávanú frázu. Ten potom nevidí, čo zadával a nevie, či hľadaný produkt naozaj neexistuje alebo ho napísal chybne. Optimalizovaný algoritmus mu zobrazuje hľadanú frázu v zadanom tvare, a poskytuje po neúspešnom hľadaní nápovedu. Vďaka tomu, že algoritmus využíva dva doplnkové spôsoby (korekciu preklepov a návrhy opravených slov) vedie k vyššiemu počtu úspešne dokončených vyhľadávaní. To kladne hodnotili aj osoby zúčastnené testovania.

Po dobu testovania sa prázdne výsledky zobrazili len pri zadaní skutočne neexistujúceho názvu produktu, ako napr. *kropáč*. Ďalej sa výsledky nezobrazili pre príliš komplikovaný preklep, kedy algoritmus nebol schopný nájsť zhodu so zaindexovaným slovom, napr. *kevtianc*.

Pri testovaní na hľadané výrazy získané analýzou prístupov z webovej analytiky sa zobrazovali výsledky v množstvách, ktoré popisuje tabuľka 5.

Tabuľka 5: Testované výrazy a počty ich výsledkov

Typ výrazu	Vyhľadávaná fráza	Počet nájdených produktov		Poznámka Navrhované slovo
		Súčasný stav	Nové vyhľadávanie	
fráza	květináč na svod	0	183	
slovné spojenie	truhlík 40 cm	106	160	
slovné spojenie	truhlík růžová	0	134	
slovné spojenie	truhlík červený	0	116	
fráza	červená	72	72	bez zmeny
fráza	40 cm	0	28	
slovné spojenie	28 x 14 x 21	0	6	
neexistujúce	misky	0	0	skyboard
neexistujúce	miska	0	0	dětská, eskada
neexistujúce	červený	0	0	červená, černá, černé
preklep	trhlík	0	0	truhlíku, truhlík
neexistujúce	quattro	0	0	bez zmeny

Tabuľka 5 popisuje, koľko relevantných produktov sa zobrazí užívateľovi po zadaní frázy. Modrou farbou sú uvedené príklady slov, pri ktorých aplikáciou algoritmu nenastala zmena.

Pri testovaní nebol nájdený záznam, pre ktorý by bol počet výsledkov horší ako v pôvodnom stave. Pre lepšiu orientáciu sú výrazy v tabuľke označené typom, o aký výraz sa jedná. v konečnom dôsledku záleží od užívateľa a zadávanej frázy, ako spoľahlivo bude vyhľadávanie fungovať. Vylepšený algoritmus minimalizuje počet zobrazení informácie, že žiadne výsledky neboli nájdené.

Čo sa týka usporiadania produktov, na prvý pohľad sú usporiadané podľa ceny zostupne. v kapitole bolo zdôvodnené, že je táto voľba zvolená kvôli princípu kontrastov. Pri výraze napríklad *červená*, teda výsledok *exact match* vetvy, sú výsledky primárne usporiadané podľa relevantnosti. Keďže sa slovo *červená* nachádza v zaindexovaných stĺpcoch nájdeného produktu 3-krát, je zobrazený ako prvý. Je to produkt bob za 1595 Kč. Na ďalších miestach sa nachádzajú produkty aj s vyššou cenou, ale hľadané slovo obsahujú už iba 2-krát.

V pôvodnom (súčasnom) nastavení sa ako prvý produkt zobrazil červený kvetináč s cenou 13 Kč. v prípade poradia pri *truhlík 40 cm* sú pôvodne zobrazené truhlíky všetkých veľkostí, a rozmer 40 cm je potrebné hľadať ručne v množstve výsledkov. Číselný údaj totiž vyhľadávanie ignoruje.

Nové vyhľadávanie umiestni 40 cm truhlíky na začiatok a potom zvyšné truhlíky s udanými rozmermi podľa relevancie a ceny, a na koniec umiestni produkty, ktoré najmenej spĺňajú kritériá, teda obsahujú napríklad iba jedno z hľadaných slov – *40* alebo *cm*.

7 Diskusia

7.1 Zhodnotenie dosiahnutých výsledkov a cieľov

Vyhľadávanie jedného slova bez preklepov

Pri bezchybnom zadaní iba jedného slova súčasné vyhľadávanie funguje dobre, ale výsledky radí nevhodne. Optimalizácia teda spočíva v upravení spôsobu usporiadania zoznamu výsledkov. Vzniká tu ale problém, keď sú niektoré kľúčové slová obsiahnuté aj v názvoch produktov (napr. truhlík, držiak truhlíku). Názvy produktov sú príliš dlhé a popisné, obsahujú rozmery, názov rady, tovaru a farbu, všetko približne v základnom tvare slova. Keď sú tieto slová použité aj v popise produktu, sú duplicitné a v prípade vyhľadania tohto slova sa daný produkt stáva viac relevantným voči zadávanej fráze.

Vyhľadávanie viacslovného výrazu

Súčasné vyhľadávanie nedokázalo nájsť prídavné mená v rôznych tvaroch, úspešne spracuje iba základný tvar slova (*hnědá*). Neporadilo si ani s preklepmi (truhlík, truhlíky, truhlíku, truhlíkoch, trhlík hnědý). v prípade existencie slov v rozličných produktoch boli výsledky tiež nenájdené (napríklad kombinácia slov *truhlík červená*, pričom *červený truhlík* neexistuje). Vo väčšine prípadov nebol nájdený žiadny výsledok. Napr. už spomínaný *truhlík hnědý*. Keďže nový algoritmus najskôr hľadá frázu, neskôr rozšírené vyhľadávanie, potom aspoň jedno slovo, a nakoniec preklep, nejaké výsledky budú zobrazené. v tomto prípade sú zobrazené výsledky slova zadaného bez chyby, teda pre slovo truhlík. Slovo hnědý je nezohľadnené, pretože nebolo nájdené.

Vyhľadávanie viacslovného výrazu s preklepmi

Ak sa celá fráza nenájde, a spadne do vetvy s detekciou preklepov, užívateľ dostane ako návrh opravenej frázy na výber iba jedno slovo, aj v prípade, že slov zadal viac. Tu nastáva problém kombinatoriky, do akej veľkej miery treba hľadať a porovnávať iba jedno slovo, dve, či všetky zadané. Kvôli lepším výsledkom by sa napríklad fráza *truhlík trakota* kombinatoricky dostala vďaka kombináciám bez opakovania na 10 možností. Výsledky vyhľadávania poskytnú návrhy na slová: truhlíku, Truhlík, terakota, antracit, Extra.

$$C_k(n) = \frac{n!}{k!(n-k)!} = \frac{5!}{2!(5-2)!} = 10 \quad (2)$$

Platí, že pre 5 nájdených slov by sa mali vytvoriť minimálne dvojice. To znamená že by mali vzniknúť dvojice každé slovo s každým okrem seba samého, aby výsledky boli presnejšie.

Usporiadanie vyhľadaných produktov

Čo sa týka usporiadania produktov, pre obchod je výhodnejšia varianta usporiadania podľa ceny zostupne. Dôvodom je to, že príslušenstvo k hlavnému produktu obsahuje aj názov produktu samotného (truhlík – držiak truhlíku), takže vyhľadávanie nájde aj toto príslušenstvo. Vďaka nižšej cene by bolo zaradené na začiatok. To je nežiadúce. Zákazník by túto možnosť tiež rád odfiltroval, aby videl iba hlavné produkty.

Nižším cenami bývajú tiež označené produkty, ktoré slúžia skôr ako doplnky k hlavnému produktu. Pokiaľ má zákazník záujem na usporiadaní podľa ceny vzostupne, je ochotný si toto nastavenie zmeniť sám.

Dosiahnuté výsledky sú podľa testovania užívateľmi lepšie ako v stave pred vylepšením. Pozitívne bolo hodnotené zobrazovanie návrhov slov v prípade neexistencie užívateľom zadávaných slov.

Miera vylepšenia súčasného spôsobu vyhľadávania závisí na presnosti a množstve slov v zadávanej fráze. Pri využití optimalizovaného algoritmu je podľa testovania o 50 % väčšia šanca, že užívateľ nájde, čo hľadá.

7.2 Ďalšie možnosti rozšírenia práce

Užívateľ by mohol dostať na výber, či si želá vyhľadávať v názve produktu, v popise produktu alebo všade. Pretože ako v pôvodnom, tak aj v novom vyhľadávaní sú medzi výsledkami hľadaného slova *fantazie* nájdené aj kvetináče, ktoré majú v popise „podľa vašej fantázie“. z hľadiska fulltextového vyhľadávania je to správny výsledok, no ak užívateľ hľadal produktovú radu *fantazie*, je tento výsledok nedôležitý.

Pri hľadaní iba v názve by sa používal zaindexovaný stĺpec názvu, nezohľadňovalo by sa použitie rovnakých slov v popise, a výsledky by mali byť relevantnejšie. Táto voľba by mala byť primárne doporučená. v prípade voľby hľadať všade by sa zákazníkovi dostalo viac výsledkov, no správnosť usporiadania by nebola jednoznačná.

Ďalšou možnosťou rozšírenia práce je vyhľadávanie preklepov pri zadaní hľadanej frázy viacerými slovami a to tak, aby návrhy na opravu zadávanej frázy zohľadňovali všetky zadávané slová a ponúkali možnosti návrhu ich kombinácií.

Pri zmene databázy produktov by mohol obchod dosahovať vyššie tržby, pretože by mohlo byť menej užívateľov, ktorí stránky opustia bez vykonania nákupu, ako vyplýva z prieskumu z tabuľky 2. Zložité a dlhé názvy produktov s farbami a rozmermi by sa mohli rozdeliť do viacerých atribútov, kde by vyhľadávanie fungovalo samostatne. Za zamyslenie by mohlo stáť vytvorenie pomocného atribútu obsahujúceho hlavné slovo produktu, ktorý by mal vyššiu váhu. Napríklad držiak truhlíku by mal ako hlavné identifikačné slovo držiak. Rovnako by tento atribút mohol obsahovať inak vyskloňované slovo, prípadne množné číslo, napr. držiaky, držiakom a podobne, a primárne by sa vyhľadávalo v tomto atribúte.

8 Záver

Cieľom práce bolo vylepšenie fulltextového vyhľadávania v elektronickom obchode tak, aby výsledky vyhľadávania lepšie uspokojovali potreby zákazníkov. Na dosiahnutie tohto cieľa bolo potrebné pochopiť, ako sa ľudia správajú pri nákupe a aký to má vplyv na vyhľadávanie a rozhodovanie. Následne boli preskúmané a analyzované historické dáta o návštevách a vyhľadávaní v e-shope. Tieto dáta boli získané z analýzy prístupov z Google Analytics za obdobie jedného roku, jar 2014 – jar 2015. Pojednávali o vyhľadávaných frázach, o prínose týchto fráz k zisku, početnosti využitia vyhľadávania a pod. Najčastejšie boli hľadané jednoslovné názvy produktov alebo produktových rád. Len v prípade 7 % zo všetkých návštev skúmaného elektronického obchodu bolo využité vyhľadávanie. Táto skupina ľudí bola však viac odhodlaná nájsť konkrétny produkt, strávili na jednej stránke dvakrát toľko času ako osoby, ktoré nepoužili vyhľadávanie. Na základe skutočného správania zákazníkov bolo odvodené, že sú ochotní svoj čas venovať spresneniu vyhľadávaných fráz, prípadne zmene usporiadania. z tohto prieskumu tiež vyplývalo, že užívatelia využívajúci vyhľadávanie sa stanú zákazníkmi s menšou pravdepodobnosťou ako tí, ktorí vyhľadávanie nepoužijú. Tieto analýzy prístupov boli vytvorené zo získaných dát z nástroja webovej analytiky Google Analytics.

V rámci analýz užívateľských preferencií sa v dotazníkovom prieskume zisťovalo, aký majú názor na usporiadanie produktov. Ďalej aké výsledky by radi videli pri zadaných konkrétnych frázach, v akej štruktúre by chceli vidieť zobrazené produkty a podobne. Pri vyhodnocovaní odpovedí bolo zistené, že väčšina užívateľov by chcela zobrazenie produktov v zozname. Ideálne usporiadanie týchto produktov je podľa ceny vzostupne. Dotazníkový prieskum užívateľských preferencií bol vytvorený v softvéri pre tvorbu dotazníkov Survio.

Ďalej bol vykonaný prieskum funkcionality súčasného vyhľadávania zadávaním najčastejšie vyhľadávaných fráz a sledovaním očakávaných a skutočne zobrazených výsledkov. Na základe tohto prieskumu boli sumarizované pravidlá pre vylepšenie vyhľadávania, teda čo by bolo vhodné vytvoriť, aby sa zákazníkovi ľahšie vyhľadávalo a tým jednoduchšie nakupovalo. Obchodné pravidlá boli stanovené ako všeobecné predpoklady, ktoré by mal spĺňať každý elektronický obchod.

Na vytvorenie vylepšeného algoritmu bolo treba najskôr zistiť, aký je východzí stav súčasného vyhľadávania – v ktorých atribútoch sa vyhľadáva a aké je nastavenie usporiadanie výsledkov.

Implementácia vzišla z vytvorených obchodných pravidiel. Algoritmus postupu pri vyhodnocovaní zadaného výrazu bol vytvorený tak, aby zahŕňal všetky možnosti, aké môžu nastať – aby v každom z týchto prípadov užívateľ dostal nejaké výsledky alebo návrh zadaného slova po prípadnej oprave, či rôzne kombinácie zobrazovaných výsledkov pri nejednoznačne zadanej fráze.

Výsledkom práce tak bolo upravené vyhľadávanie podľa získaných a analyzovaných informácií. Inovované vyhľadávanie poskytuje výsledky aj v prípade zadania dotazu s preklepmi alebo časťami slov, kedy sa v pôvodnom vyhľadávaní zobrazila informácia o nenájdení produktu. Namiesto nej je užívateľovi poskytnutý opravený návrh hľadaného slova. Výsledky vyhľadávania sú potom zobrazené podľa relevancie, takže na prvých pozíciách užívateľ nájde výsledky, ktoré najlepšie vyhovujú zadanému kritériu vyhľadávania. Upravené vyhľadávanie tak vedie k väčšiemu uspokojeniu užívateľov. Dá sa konštatovať, že zadané ciele práce boli úspešne splnené.

9 Literatúra

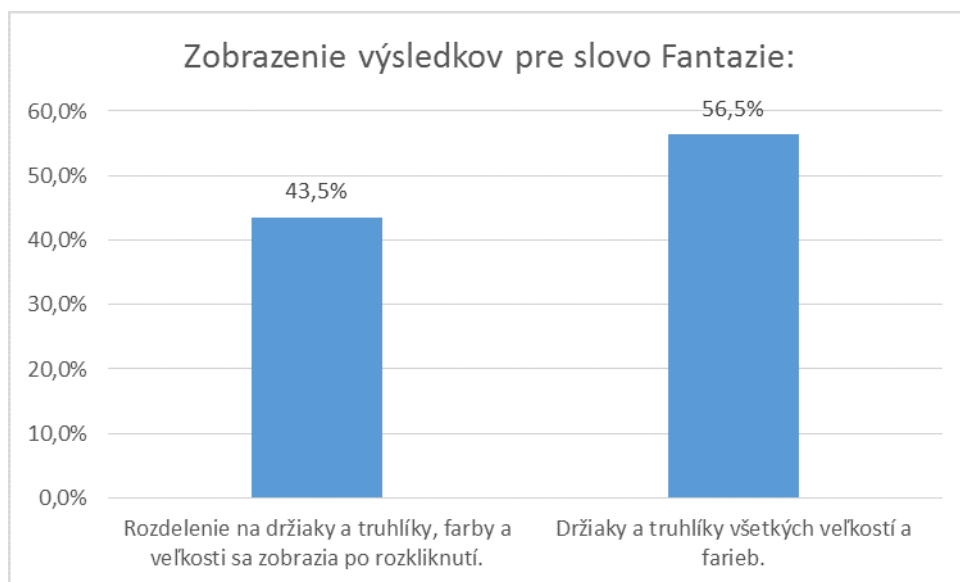
- 1.2. Sphinx features. Sphinx — Open Source Search Server [online]. Sphinx Technologies Inc., Copyright©2001-2015 [cit. 2015-12-06]. Dostupné z: <http://sphinxsearch.com/docs/current.html#features>.
- 12.9 Full-Text Search Functions. MySQL :: MySQL 5.7 Reference Manual [online]. Oracle Corporation and/or its affiliates, ©2015 [cit. 2015-12-06]. Dostupné z: <http://dev.mysql.com/doc/refman/5.7/en/fulltext-search.html>.
- 5.1. Matching modes. Sphinx — Open Source Search Server [online]. Sphinx Technologies Inc., Copyright©2001-2015 [cit. 2015-11-29]. Dostupné z: <http://sphinxsearch.com/docs/current.html#matching-modes>.
- AKSYONOFF, A., Sphinx 2.3.2-dev reference manual. Sphinx — Open Source Search Server [online]. 2015 [cit. 2015-08-04]. Dostupné z: <http://sphinxsearch.com/docs/current.html#compilingfromsource>.
- ALI, A., Sphinx search: beginner's guide : implement full-text search with lightning speed and accuracy using Sphinx. Olton, Birmingham: Packt Pub. Ltd., 2011, v, 226 p..
- CIALDINI, R., Zbraně vlivu: manipulativní techniky a jak se jim bránit. Vyd. 1. v Brně: Jan Melvil, 2012, 333 s. Žádná velká věda. ISBN 978-80-87270-32-5..
- CLIFTON, B., Google analytics: podrobný průvodce webovými statistikami. Vyd. 1. Překlad Lukáš Krejčí. Brno: Computer Press, 2009, 334 s. ISBN 978-80-251-2231-0..
- CUTRONI, J., Google Analytics. 1. ed. Beijing, China: O'Reilly, 2010. ISBN 978-059-6158-002.
- Ecommerce Benchmark Guide 2007 [online]. MarketingSherpa, 2007 [cit. 2015-08-03]. ISBN 978-1-932353-64-8. Dostupné z: <http://www.marketingsherpa.com/exs/Ecom07Summ.pdf>.
- Elasticsearch: The Definitive Guide [online]. Clinton Gormley, Zachary Tong, ©2015 [cit. 2015-12-06]. Dostupné z: <https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html>.
- Elektronický obchod — BusinessInfo.cz [online]. 2015 [cit. 2015-10-14]. Dostupné z: <http://www.businessinfo.cz/cs/clanky/elektronicky-obchod-ppbi-51052.html#!>.
- GONZALEZ, T., Dashboard Design: Key Performance Indicators and Metrics: Choosing the right data to display. In:BrightPoint Consulting, Inc. [online]. 2015 [cit. 2015-04-04]. Dostupné z: <http://www.brightpointinc.com/dashboard-design-key-performance-indicators-metrics/>.
- GOODWIN, D., Organic vs. Paid Search Results: Organic Wins 94% of Time. In: Search Engine Watch [online]. Aug 23, 2012 [cit. 2015-04-17]. Dostupné z: <http://searchenginewatch.com/sew/news/2200730/organic-vs-paid-search-results-organic-wins-94-of-time>.

- HEARST, M., BAEZA-YATES, R, Berthier de Araújo Neto RIBEIRO. Modern information retrieval: The Concepts and Technology behind Search [online]. New York: ACM Press, 1999, 21 - 55 [cit. 2015-04-13]. Second edition. ISBN 0-201-39829-x..
- KINSHUK, JERATH, MA LIYE A PARK YOUNG-HOON. Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity. In: JMR, Journal of marketing research. American Marketing Association, © 2014, s. 480-486. ISSN 0022-2437..
- NĚMEC, R., Jak optimalizovat stránky pomocí Indexu USD. Robertnemec.com [online]. 2009 [cit. 2015-03-29]. Dostupné z: <http://robertnemec.com/optimalizace-index-usd/>.
- NĚMEC, R., Jak zjistíte, kolik vám která stránka vydělává - Index USD. Robertnemec.com[online]. 2009 [cit. 2015-03-29]. Dostupné z: <http://robertnemec.com/index-usd-kolik-stranky-vydelavaji/>.
- OBŠÍVAČ, T., PV219 Seminář webdesignu. ECKSTEINOVÁ, IVETA. E-shop [online]. 2010 [cit. 2015-10-14]. Dostupné z: <http://www.fi.muni.cz/xobsivac/PV219/prezentace10/e-shop.pdf>.
- PETERSON, E., The Big Book of Key Performance Indicators. First Edition. 2006. Dostupné z: http://design4interaction.com/wp-content/uploads/2012/09/The_Big-Book_of_Key_Performance_Indicators_by_Eric_Peterson.pdf.
- PROCHÁZKA, D., Poznáváme uživatele. In: Pokročilá uživatelská rozhraní [online]. 2011 [cit. 2015-12-13]..
- Rethink Web Analytics: Introducing Web Analytics 2.0. Occam's Razor by Avinash Kaushik [online]. Avinash Kaushik, Copyright©2015 [cit. 2015-12-06]. Dostupné z: <http://www.kaushik.net/avinash/rethink-web-analytics-introducing-web-analytics-20/>.
- ROUSE, M., Google Analytics. In: TechTarget [online]. 2011 [cit. 2015-03-29]. Dostupné z: <http://searchbusinessanalytics.techtarget.com/definition/Google-Analytics>.
- SHARIF, S., Where Should The Google Analytics Tracking Code Be Placed?. LunaMetrics [online]. 2012 [cit. 2015-03-29]. Dostupné z: <http://www.lunametrics.com/blog/2012/02/09/where-put-google-tracking-code/>.
- SOSTRE, P. a LECLAIRE, J., Web analytics for dummies. Hoboken, NJ: Wiley Pub., c2007, xviii, 362 p. ISBN 978-047-0098-240..
- The Difference Between Paid Search Results & Organic Search Results. In: Barcelona Creative Group [online]. © 2015 [cit. 2015-04-07]. Dostupné z: <http://info.barcelonacreative.com/blog/bid/268929/The-Difference-Between-Paid-Search-Results-Organic-Search-Results>.
- TIDWELL, J., Designing interfaces. 2nd ed. Sebastopol: O'Reilly, c2011, xxv, 547 s. ISBN 978-144-9379-704..
- UNGR, P., Webová analytika pro eshopy aneb co a jak vyhodnocovat. MůjPrvníEshop.cz [online]. [cit. 2015-03-30]. Dostupné z: <http://www.muoprvnieshop.cz/analytika-avyhodnocovani/>.

What is internal search and why should I track it?. In: Yahoo! Help [online]. 22.9.2010 [cit. 2015-04-07]. Dostupné z: <http://help.yahoo.com/l/us/yahoo/ywa/faqs/tracking/customcode/customcode-23.html>.

Prílohy

A Vyhľadávanie termínu Fantazie



Obr. 11: Vyhľadávanie termínu Fantazie

B Dátové CD

Priložené CD obsahuje:

- vstupné dáta;
- zdrojové kódy;
- dotazník.