

Tallinna Ülikool
Informaatika Instituut

Veebroomaja rakendus vastavalt meediaagentuuri vajadustele

Bakalaureusetöö

Autor: Uku Saarm

Juhendaja: Jaagup Kippar

Autor: „2015

Juhendaja: „2015

Instituudi direktor: „2015

Tallinn 2015

Autorideklaratsioon

Deklareerin, et käesolev bakalaureusetöö on minu töö tulemus ja seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

Autor: Uku Saarm

.....

(kuupäev)

.....

(autor)

Lihlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina _____ (sünnikuupäev: _____)
(*autori nimi*)

1. annan Tallinna Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

(*lõputöö pealkiri*)

mille juhendaja on _____,
(*juhendaja nimi*)

säilitamiseks ja üldsusele kättesaadavaks tegemiseks Tallinna Ülikooli Akadeemilise Raamatukogu repositooriumis.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tallinnas/Haapsalus/Rakveres/Helsingis, _____
(*digitaalne allkiri ja kuupäev*)

Lühendite ja mõistete sõnastik

Veebiroomaja - Robotprogramm, mis otsib veebis kindla ja korrapärase meetodiga uusi veebidokumente ja lisab leitud tulemused erinevatesse andmebaasidesse.

UGC – *User-Generated-Content* (kasutajate poolt loodud sisu)

ATL – *Above The Line* (peavoolu meediakanalite kasutamine)

BTL – *Below The Line* (alternatiivsete meediakanalite kasutamine)

PR – *Public Relations* (Rahvusvahelised suhted)

WOM - *Word-Of-Mouth* (suusõnaline turundus)

SEO - *Search engine optimization* (otsingumootori optimeerimine)

POS - *point of sale* (müügipunkt)

Affiliate-Marketing (vahendavad veebilehed)

Call to action – üleskutung

RSS – *Rich Site Summary* (vorming, milles edastatakse sagedasti uueneva veebi sisu)

Jooniste nimekiri

Joonis 1- Tüüpiline kõrghaseme arhitektuuriga veebiroomaja (C. Castillo. 2004)	17
Joonis 2- Roomaja töö võrdlemine tähistäevaga (C. Castillo, 2004).....	20
Joonis 3- Sainsbury's PR'i vastamine Twitteri Sainsbury's kaubandusketi kasutajakonto jälgijale. Tähelepanu juhtida asjaolule, et kliendi säuts ei ole suunatud otse Sainsbury kasutajatunnusele. (Twitter)	24
Joonis 4- Väljavõtte roomaja raportist (isiklik pildipank)	25
Joonis 5- Meediamonitooringu raporti esitamine graafiliselt (neilpatel).....	26
Joonis 6- Google Alerts'i seadistuspaneel (Google).....	28
Joonis 7- Twitteri otsingumootor (Twitter)	29
Joonis 9- Tulemus otsingusõnale "President" (Socialmention)	31

Tabelite nimekiri

Tabel 1 – Turukommunikatsiooni Offline ja Online võtete jaotus (Smith, 2001).....	10
Tabel 2 - Interneti kasutuse protsent Eestis (Internetworldstats).....	12

Sisukord

Sissejuhatus.....	7
Töös kasutatav ettevõtte.....	8
Eesmärk.....	8
Ülesande püstitus	8
Töö ülesehitus	9
1 Meediaagentuur.....	10
1.1 Muutused meediaagentuurides.....	10
1.2 Digitaalajastu.....	12
1.3 Firmade erinevus	14
2 Veebiprogramm.....	15
2.1 Ajalugu	15
2.2 Arhitektuur ja tööpõhimõtted.....	16
2.3 Kasutus	17
2.4 Esinevad probleemid	19
3 Veebiprogrammi rakendus meediaagentuuris	22
3.1 Meediamonitoring	22
3.2 Monitoringu protsess.....	24
3.3 Erinevad teenused	28
4 Sobivaimate teenuste kombineerimine	33
4.1 Monitoringu loomine.....	33
4.2 Teenused vastavalt klientide vajadustele	34
Kokkuvõte.....	36
Summary.....	37
Kasutatud kirjandus	38

Sissejuhatus

Eduka meediaagentuuri aluseks on inimeste loomuse ja käitumismallide mõistmine. Meediaagentuur peab mõistma oma tarbija käitumist; nad tagavad, et sõnumid suunatakse õigetele inimestele, õigel ajal, õiges kohas; nad veenduvad, et nende kliendid, ennast reklaamivate firmade näol, mõistavad viimase tehnoloogia kasutamise potentsiaali ja neile tagatakse ligipääs kõikidele olulistele meedia kanalitele (mediacom).

Tarbijad on tehnoloogia arenguga omaks võtnud hulgaliselt erinevaid meediakandjaid, informatsiooni ja meelelahutuse tarbimiseks. Meediaagentuur peab seega tagama, et klientide soovid ja sõnumid jõuaksid tarbijateni neid samu uusi kanaleid pidi, mida üha enam kasutusele võetakse. Meedias töötamise tööstandardid on tehnoloogiaajastul ilmselgelt muutunud ja see on iga endast lugupidava agentuuri kohustus, hoida ennast kursis kõikide muutuste ja tehnoloogiatega, tutvustades neid võimalusi nii oma klientidele, kui töötajatele.

Ühe sellise tehnoloogia valdkonnana näeb töö autor veebiroomajaid (ing.k *web crawler*). Veebiroomaja on programm, mille eesmärgiks on süstemaatiliselt läbi käia (roomata), spetsiifiline ala veebis, olgu selleks siis veebilehed, veebirakendused või mõni veebiteenus. Kui veebiroomajad on Interneti kasutusega seotud juba 1990-ndate keskpaigast (Thehistoryofseo, 2011), kuid kui neid esmakordselt hakati otsingumootorite jaoks ära kasutama, siis tänapäeval on lisaks kõikidele otsingumootoritele nende peale ehitatud ka hulgaliselt teise eesmärgiga programme. Suurem osa uutest roomamisrakendustest on tihedalt seotud meediaagentuuri tööga, eriti sotsiaalmeediaga, kuid ometigi näitavad agentuuri sisesed küsitlused, et teadlikkus sellest nii klientide kui ka isegi agentuurides töötajate seas äärmiselt madal. See tulemus ei pane imestama, sest näiteks informatsiooni saamiseks sotsiaalmeedia läbi roomamine on suhteliselt uus nähtus, suurt rolli mängib selle juures ka fakt, et enamus sotsiaalvõrgustikke on välja arenenud alles viimase aastakümneni jooksul. Kuid on oluline mõista ja omaks võtta veebis roomamise poolt kätte saadava kasutaja poolse sisu (ing.k *User-Generated-Content*, lühidalt UGC) hüvesid, millest firmad üha enam huvitatud on – nagu toote/teenuse hindamine, tagasiside, etteheited, brändi jälgimine, brändi analüüs, konkurentsi analüüs, üleüldine brändi maine jne (promptcloud).

Lisaks kasutajate poolse sisu hankimisele läbi veebiroomamise, analüüsitakse läbi ka teised valdkonnad, millega meediaagentuurid oma igapäeva töös kokku puutuvad. Seejärel vaadatakse läbi kõik veebiroomajate põhilised omadused ja iseärasused. Peale põhjalikku mõlema teema ülevaadet ja detailide välja toomist, proovitakse leida veebiroomajale praktilisi lahendusi, mis aitaks kaasa meediaagentuuri tööle kas mõne vana teenuse muutmisel, uue teenuse arendamisel.

Antud teema on ajendatud sellest, et autor puutub töö raames igapäevaselt kokku meediaagentuuri töös kasutatava tehnoloogiaga, tunneb selle vastu isiklikku huvi ning näeb vajadust anda põhjalik ülevaade, kuidas paljud roomajapõhised protsessid töötavad, et seeläbi leida kõige funktsionaalsemaid lahendusi agentuuri klientidele. Autori tööalane ülemus tunnistas informatsiooni päevakorrasust ning pidas töö läbiviimist asjakohaseks.

Töös kasutatav ettevõte

Antud töös käsitletav meediaagentuur on Eestis tunnustatud ettevõtte, mis tegeleb sotsiaalmeedia turunduse, veebirakenduste disaini, meedia ostu-müügiga, veebikampaniate analüüsiga, veebidisaini, otsingumootorite optimeerimise ja palju muuga. Konfidentsiaalsust ja konkurentsianalüüsi silmas pidades ei ole käesolevas töös kasutatud ettevõtte õiget nime.

Eesmärk

Käesoleva töö eesmärgiks on anda põhjalik ülevaade veebiroomajate olemusest ja kasutusest meediaagentuuri töös ning seeläbi leida lahendused ning funktsionaalsused, millest agentuuri kliendid kõige rohkem võidaksid.

Ülesande püstitus

Käesolevas töös analüüsitakse põhjalikult veebiroomaja kasutusvaldkondi, täpsemat kasutust meediaagentuuris ja erinevaid informatsiooni kaevandamis võimalusi. Töödeldakse läbi erinevad teemaga seostuvad teenused, tasuliste ja tasuta võimaluste erinevused ja kasutusolukorrad. Sealjuures proovitakse leida sobivaid veebiroomaja-põhiseid rakendusi agentuuri klientidele, mis oleksid ühtlasi nii uuenduslikud, kuid ka tulusad. Agentuuri vajadused ja nõuded on välja toodud ettevõtte töötajate ja autori isiklike kogemuste põhjal.

Töö eesmärgini jõudmiseks tuleb esmalt tutvustada meediaagentuuri tööd, nende poolt pakutavaid digitaalseid teenuseid ja klientide erinevuseid. Seejärel tutvustatakse veebiroomajat, andes põhjalik ülevaade selle taustast, arengust, tööpõhimõtetest, erinevatest kasutusmeetoditest ja potentsiaalidest. Seejärel vaadeldakse meediaagentuuri vajadusi sotsiaalmeedia ja veebi valdkondades. Kirjeldatakse rakendusi, mis on otseselt seotud info hankimise ja sotsiaalmeedia jälgimisega. Lõpus tuuakse välja veebiroomamisel põhinevaid lahendusi, mille rakendamine klientidele kasu toob.

Töö ülesehitus

Töö on üles ehitatud järgmisest osadest:

- Esimese peatükis tutvustatakse meediaagentuuri vajadust, muutuseid tehnoloogia arenguga, tööülesandeid, kasutatavaid meetmeid ja klientide erinevuseid.
- Teises peatükis tutvustatakse veebiroomajate vajadust, nende arengut, erinevaid tööpõhimõtteid, kasutusmeetodeid ja arhitektuure.
- Kolmandas peatükis tutvustatakse klientide vajadusi, kuidas meediaagentuurid kasutavad ära veebiroomamisel põhinevaid rakendusi, millised on nende puudused ja erinevused.
- Neljandas peatükis tuuakse välja veebiroomaja rakendamislahendused ja funktsionaalsused, millest erinevad meediaagentuuri kliendid kõige rohkem võidaksid.

1 Meediaagentuur

Ühe maailma juhtivama meediaagentuuri Mediacom'i defineerib meediaagentuuri järgmiselt: „*A media agency makes sure a marketing message appeals to consumers, appears in the right place, at the right time and that the advertiser pays the best possible price.*“ Eesti keelde tõlgituna: „Meediaagentuur tagab, et reklaami sõnum läheks tarbijatele korda, ilmuks õiges kohas, õigel ajal ja reklaami tellija maksaks võimalikult madalat hind“. Kokkuvõttes tähendab see seda, et nemad on inimesed, kes vastutavad selle eest, et õigetele inimestele saabuks õigel ajal õige sõnum, mis neile ka meeldib. Võidavad sellest kõik – sihtrühma kuuluvad inimesed saavad just neile suunatud ja läbimõeldud reklaame, sihtrühmast välja jäävaid inimesi puudutavad neid mitte huvitavad reklaamid aga minimaalselt.

Meediaagentuuri tegevusest rääkides ei saa mööda turunduskommunikatsioonist. Klassikalises käsitluses on turunduskommunikatsioon ostu/müügiga seotud sõnumi edastamine saatjalt vastuvõtjale (Kuusik et al 2010). Meediaagentuuri poolt pakutavad teenused ja turunduskommunikatsioon kattuvad väga laialt, kuid päris samaks neid siiski pidada ei saa. Meediaagentuurid pakuvad enamasti kõiki teenuseid, mida turunduskommunikatsiooni planide täideviimisel vaja on.

1.1 Muutused meediaagentuurides

Turunduskommunikatsiooni on siiani liigitatud kanalipõhiselt: ATL (*above the line*) mis kasutab nn peavoolu meedia kanaleid ja BTL (*below the line*), mis kasutab müügitoetus- ja alternatiivseid kanaleid. Viimasel ajal on mitmed tuntud autorid aga hakanud eristama ka lihtsalt *online* ja *offline* turunduskommunikatsiooni võtteid ja kanaleid (Smith, 2011). Ka käesolevas töös kasutatakse turunduskommunikatsiooni kanalitest ülevaate andmisel *offline* ja *online* liigitust, mis on välja toodud allolevas tabelis (Smith, 2001).

Tabel 1 – Turukommunikatsiooni Offline ja Online võtete jaotus (Smith, 2001).

Offline vormid	Online vormid		
Reklaam	Interaktiivne	TV	reklaam,

	reklaambännerid, klikipõhised reklaamid
PR (suhtekorraldus, ing.k public relations)	<i>Online</i> Optimeeritud PR, Viirusturundus
Sponsorlus	<i>Online</i> sponsorlus – kogukonnad, üritused jm
Müügitoetus / edutamine	Müügi- ja lisaväärtuskampaaniad kodulehtedel ja e-posti turundus
Otsepost	Uudiskirjad jm e-posti pakkumised
Müügipersonal / telemarketing	<i>Affiliate marketing</i> (vahendavad veebilehed), veebileht kui müügikanal
Pakendamine	Tootepildid ja –info kodulehel
POS (müügipunkt, ing.k point of sale)	Veebileht kui <i>online</i> POS, kampaaniad, <i>call to action</i> (üleskutsung), auhinnad jm
Müügikoha disain	Veebilehe disain, kasutusmugavus, SEO (otsingumootori optimeeritus, ing.k search engine optimization)
Näitused ja konverentsid	Virtuaalsed näitused, virtuaalkeskonnad
WOM (suusõnaline turundus, ing.k <i>word-of-mouth</i>)	Sotsiaalmeedia (Facebook), sotsiaalse sisu jagamise keskkond (Youtube), sotsiaalsed soovituskeskkonnad (Delicious)

Suurem osa tänapäevaseid *offline* kampaaniaid on siiski ühel või teisel viisil integreeritud *online* meediaga, toimib üksteise toetamine, kus mõlemad kanalid toovad ka vastastikkuselt tulu (Näiteks kutsutakse TV seriaalide lõpus inimesi üles külastama nende Facebooki lehekülge, seal leidub aga jällegi reklaami, mis suunab inimesed järgmist seriaali osa vaatama). Nii ei saagi või

pigem ei ole mõtet enam tugevat joont *online* ja *offline* turunduskommunikatsiooni võtete kasutamise vahele tõmmata.

Kanalite vaheliste piiride hägunemise ja tehnoloogia kiire arenguga on palju muutunud meediaagentuuride töös – kui varem põhines kogu töö suurel määral vaid reklaampindade ostumüügiga televisioonis, välireklaamis, ajakirjades ning raadios, siis nüüd on Interneti ja tehnika levikuga nende kõrvale tekkinud suurte võimalustega *Online* keskkond. Vanamoodsad reklaami- ja meediaagentuurid on sunnitud seetõttu ümber kohanema. Klassikalised toredad ja meeldejäädavad reklaamid ei ole küll kuhugi kadunud, kuid üha sagedamini ei piisa kliendile ainult sellest. Kui varem olid reklaamikanalite valik ja suunamismeetodid piiratud, ning klient sai põhimõtteliselt vaid öelda sihtrühma vanuse ja soo, siis tänapäeval on ka sihtrühma tutvustused ja tulemuste nõuded muutunud kordades võimalusterohkemaks.

1.2 Digitaalajastu

Aastal 2000 kasutas Interneti vaid 28,2% Eesti elanikkonnast. 12 aastaga kasvas see näitaja 78% peale (Tabel 2. Internetworldstats), tänaseks on see arv kindlasti veelgi tõusnud. Mis aga veelgi olulisem, 2014. aasta septembri seisuga kasutas PewresearchCenter'i andmete kohaselt 74% täiskasvanutest mõnda sotsiaalmeedia kanalit. Eelmiste aastatega võrreldes oli näha kasvu kõikides peamistes sotsiaalmeediakanalites. Vaid Facebooki puhul oli kasv minimaalne, mis tuleneb asjaolust, et enamik potentsiaalseid huvilisi juba kasutab seda keskkonda. See kõik tähendab, et inimesed ei kasuta Interneti vaid informatsiooni otsimiseks, vaid väga suurel määral just suhtlemiseks. Kogu virtuaalne sotsialiseerumine tähendab aga seda, et meediaagentuuride klientidele enam ei piisa, kui nende reklaami vaid telerist nähakse, vaja on midagi rohkemat, vaja on luua kahepoolne suhtlus ettevõtte ja tarbija vahel.

Tabel 2 - Interneti kasutuse protsent Eestis (Internetworldstats)

Aasta	Kasutajaid	Rahvaarv	Osakaal rahvastikust
2000	366 600	1 299 000	28,2 %
2004	621 000	1 344 840	46,2 %
2006	690 000	1 332 987	51,8 %

2007	760 000	1 332 987	57,0 %
2010	969 700	1 291 170	75,1 %
2012	993 785	1 274 709	78,0 %

Suhtlus kliendi ja firma vahel läbi sotsiaalmeedia on saanud uueks standardiks, raske on leida aktiivset firmat, mis ei omaks Facebooki kontot. Zephoria.com andmetel pidas juba aastal 2012 tervelt 42% turundajatest Facebooki kasutamist oma äri jaoks väga oluliseks või lausa kriitiliseks. Seal on küll tihe konkurents, kuid võrreldes teiste meediakanalitega on pakutavad võimalused liiga soodsad, et neid kasutamata jätta. Edu võtmeks on Facebooki turu õige ära kasutamine ja loomulikult teiste seast välja paistmine. Sotsiaalmeediat kasutatakse eelkõige meelelahutuslikul eesmärgil, seega õige väljundi loomine tagab klientide poolehoiu, muutudes nende jaoks korporatiivse brändi ja reklaami asemel pigem meelelahutuseks, mida nad vabast tahtest näha ja isegi sõpradega jagada soovivad. Virtuaalne suhtlus klientidega annab lisaks positiivse brändingu ka kvaliteetse ja vahetu tagasiside. Näiteks on Inc.com statistika andmetel on 20 parima firma Facebooki lehe hulgas Old Spice. Nende lehele minnes ei teki vähemutki tunnet, et seal külastajale midagi müüa või esitleda proovitaks, leht on puhtalt meelelahutusliku sisuga, kuid endiselt tugevalt brändiga seotud.

Olulise punktina tasub välja tuua meediamonitooringu, millest hiljem täpsemalt juttu tuleb. Lühidalt on tegemist veebiprogrammi põhineval teenusega, abil veebi läbi kammides on võimalik leida toodetele õiget sihtgruppi, teha taustauuringut, konkurentsianalüüsi ja hoida ennast kursis WOM-ga, ehk mida inimesed teavad toote või brändi kohta veebis räägivad.

Autori tööga seostuvas meediaagentuuris on varem olnud kasutusel meediamonitooringu võimalus, mis oli sisse ostetud kolmanda osapoole poolt, kuid klientide poolse skeptilisuse ja agentuuri poolse vähese kogemuse tulemusena ei kestnud monitooringupõhiste teenuste edastamine väga pikalt. Sisse ostetud roomaja teenust kasutati mõne brändi jälgimiseks, kasutades lisaks brändi nimele hulgaliselt sellega seostuvaid sõnu ja termineid. Iga päev edastas raport (ehk ülevaatlik e-mail) samal kellaajal kõik tulemused, mis vahepeal nende sõnade kohta Eesti meedias ilmunud olid.

1.3 Firmade erinevus

Et mõista paremini, milliseid teenuseid ja kuidas klientidele esitada, on oluline mõista nende põhilisi erinevuseid ja vajadusi. Mõni firma on tagasihoidlik ja diskreetne, mõni teine aga innovaatiline ja uudishimu täis, proovides leida uusi võimalusi oma brändi välja toomiseks. Nii saab jaotada brändid laias laastus kaheks – aktiivseteks ja passiivseteks.

Aktiivsed firmad – Firmad, mis proovivad turule tulla alati millegi uue ja innovaatilisega, kaasates kliente nii virtuaalselt, kui füüsiliselt. Sellised firmad proovivad igast kampaaniast maksimumi võtta, tehes enne kampaania algust eeluuringuid ja kuulates sotsiaalmeediat. Kampaaniate ajal seotakse välismeedia digitaalse meediaga, et tabada võimalikult laia sihtgruppi. Sellistel firmadel on eranditult olemas üks või rohkem sotsiaalmeedia kanali konto, kasutades seda ära reaalselt fännidega suhtlemiseks ja neilt informatsiooni saamiseks. Näiteks jälgida Coca-Cola tegevust ja kampaaniaid on ilmne, et tegemist on äärmiselt aktiivse brändiga, kes proovib igal võimalusel leida võimalusi oma tarbijatega suhtlemiseks ning neile millegi uue pakkumiseks. Aasta 2014. suvel toimunud „Share a Coke“ (vt. shareacoke) oli ideaalne näide hästi töötavast UGC kampaaniast, mis tõi endaga tarbijad ka virtuaalselt kaasa, pannes nad jagama enda tehtud pilte oma enda või sõbra nimega pudelist (Postano). Aktiivsete firmade puhul on meediaagentuuri jaoks üle poole töö tehtud, sest klient ise pakub välja oma ideed arvamused, agentuuri tööks jääb need ainult võimalukult leidlikult lahendada.

Passiivsed firmad - Firmad, kus tegevus toimub edukalt samal kohal püsimiseks, mitte edasiarenguks. Passiivsete firmade puhul ei olda tihti huvitatud kampaania eeltööst ega hilisemast analüüsist, kuidas kampaania läks. Kardetakse teistsuguseid lahendusi ja ollakse pessimistlikud uue tehnoloogia suhtes. Sellised firmad on enamasti täpselt sama strateegiat jälginud juba aastakümneid, ega soovi seda niipea muuta. Passiivsetele firmadele uusi ideid pakkudes lükatakse need kiirelt tagasi või näidatakse skeptilisust, ilma neid proovimata ega põhjalikumat uurimist. Sotsiaalmeedia kanalid on nüüd küll juba ka passiivsete firmadeni jõudnud, kuid enamasti postitatakse nende poolt vaid ühepoolset tugeva reklaami maitsega materjali või korraldatakse fännide juurde saamiseks tavalisi ja maitsetuid „jaga ja võida“ loosimisi.

2 Veebiroomaja

Veebiroomaja, ehk lihtsustatult roomaja, (inglise keeles *crawler*) on programm, mille eesmärgiks on süstemaatiliselt läbi käia (roomata), spetsiifiline ala veebis, olgu selleks siis veebilehed, veebirakendused või mõni veebiteenus. Veebiroomajad käituvad ringi roomates vastavalt oma eesmärgile - laevad alla veebilehti, töötlevad informatsiooni, salvestavad URL'e, või märgistavad veebi, ehk indekseerivad. Inglise keeles kasutatakse alternatiivina ka nimetusi nagu *web spider*, *ant*, *bot*, *automatic indexer*. Eestis ei ole roomajate eraldi kasutus veel suurte massideni jõudnud, seega on levinumate nimede seas lisaks roomajale ja veebiroomajale veel vaid veebiämblik, ehk ämblik.

2.1 Ajalugu

Roomajaid hakati aktiivselt kasutama 1990ndate alguses ja nende baasil loodi ka esimesed Interneti otsingumootorid. Aastal 1993 arendas Matthew Grey Wandex'i, esimese tekstipõhise otsingumootori, mille formaat on siiani standardiks tänapäeval levinud otsingumootoritele (Thehistoryofseo, 2011). Wandex'i tehnoloogia oli esimene, mis veebiroomaja abil indekseeritud veebilehti märgistas. Järgmine suurem areng toimus aastal 1994, kui Brian Pinkertoni poolt loodud WebCrawler'i otsingumootor hakkas indekseerima veebilehtede tekste täismahus, varasemate lehekülje pealkirja asemel. Kõik tänapäeval tuntumad otsingumootorid on alguse saanud vahemikus 1993 kuni 1998, sealhulgas ka Google, kuhu Matthew Grey ise edasi tööle läks.

Matthew Grey poolt loodud Windex'i põhjaks oli veebiroomaja World Wide Web Wanderer, ehk lühemalt Wanderer. Roomaja oli Perli programmeerimiskeele põhjal loodud programm, mille eesmärgiks oli mõõta World Wide Web'i suurust ja märgistada see Wandex'i jaoks. Wanderer mõõtis aktiivselt veebi suurust kuni 1995 aasta lõpuni.

Tänapäeval on lahutamatu veebiroomajatega seotud kõik otsingumootorid, ent struktuurid ja andmemahud on võrreldes algusaastatega muutunud kordades. Näiteks kui indekseerimine ja selle põhjal info leidmine ise töötab endisel viisil, siis tuhandetest tulemustest just õige tulemuse sorteerimine on läinud üha keerukamaks. Maailma vaieldamatult suurim otsingumootor Google

(reliablesoft) filtreerib näiteks otsijale tuhandete variantide seast kuvatavaid vastuseid, küsides tagataustal oma sorteerijalt üle 200 küsimuse, nagu „mitu korda esineb otsingusõna lehel?“, „kas otsingusõnad esinevad pealkirjas, URL-is?“, „kas lehel leidub sünonüüme sellele sõnale?“, kas lehekülje kvaliteet on kõrge, madal või isegi rämpsposti laadi?“, „mis on selle lehekülje *PageRank*?“ (Google asutajate poolt loodud valem, mis võtab arvesse teiste lehekülgede suunamist valitud lehele.), lõpuks pannakse neid kõiki faktoreid arvestades igale leheküljele oma skoor, mille alusel toimub lehekülgede kuvamine otsijale.

2.2 Arhitektuur ja tööpõhimõtted

Veebiroomajad on otsingumootorite tuumik, seega hoitakse nende algoritme ja täpset ehitust kõikide suuremate ettevõtete puhul ärisaladusena. V.Shkapenyuki ja T.Sueli 2002. aastal avaldatud teoses „Design and Implementation of a High-Performance Distributed Web Crawleron“, on välja toodud arvamus:

„Üsnagi lihtne on luua aeglast roomajat, mis laeb alla mõned leheküljed sekundis lühikese aja jooksul. Ehitades aga kõrgetulemuslikku süsteemi, mis suudab alla laadida sadu miljoneid lehti mitme nädala jooksul, ilmutab endas hulgaliselt väljakutseid süsteemi disainis, I/O ja veebi efektiivsuses ning jõudluses ja hallatavuses“.

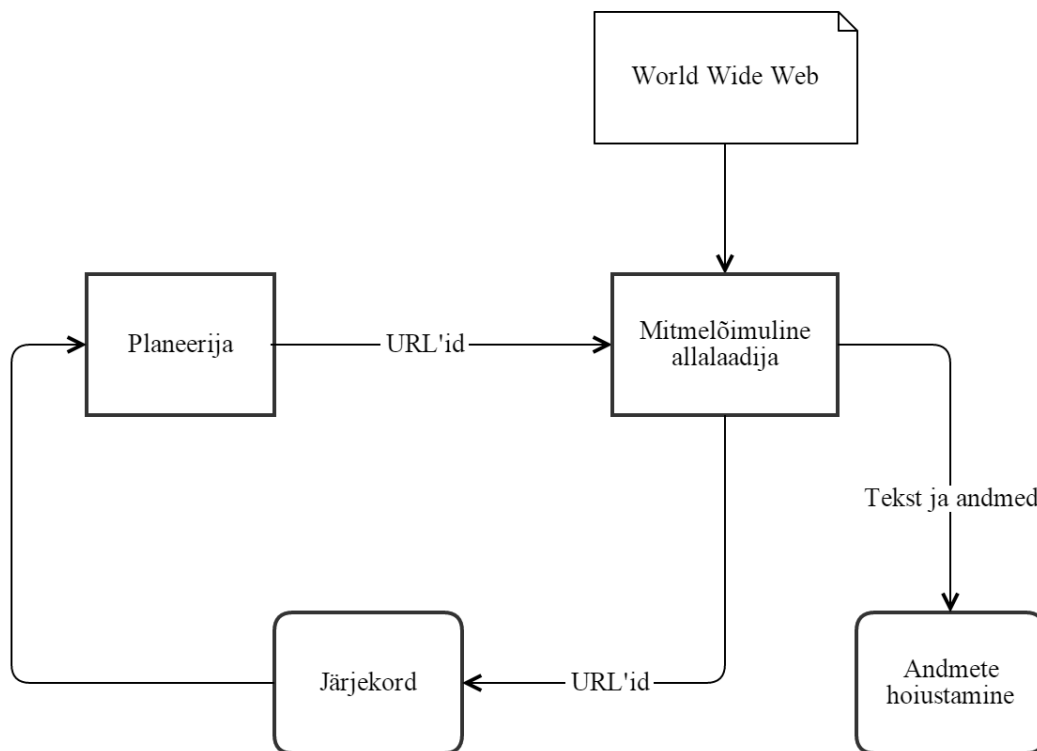
Kui mõne roomaja disain avaldatakse, puudub selle juurest tavaliselt suur hulk vajalikku informatsiooni, mis takistab teistel selle kopeerimist. Arhitektuuri puhul tuleb tänapäeval silmas pidada ka ebakvaliteetse sisuga arvestamist, mis takistab suurematel otsingumootoritel oma algoritmide avaldamist.

Vastavalt oma eesmärgile ehitatakse roomajaid väga erinevate põhimõtete, algoritmide ja arhitektuuridega. Laias laastus näeb kõige tavalisema veebiroomaja töö välja selline (C. Castillo, 2004):

- 1) Sisendiks antakse URL, mille sisu soovitakse alla laadida või töödelda.
- 2) Sisestatud URL'idest, mida külastama hakatakse luuakse järjekord, ehk front (inglise keeles *frontier*).
- 3) Roomaja töötleb URL-il olevat informatsiooni vastavalt oma eesmärgile.

- 4) Peale URL-i edukat töötlemist liigub roomaja edasi järgmise järjekorra elemendi juurde.
- 5) Juba ära külastatud URL lisatakse külastatud elementide järjekorda.
- 6) Roomaja töö võib kesta edasi nii kindla ettemääratud punktini kui ka teoreetiliselt lõpmatult.

Alljärgnevalt on tüüpilise veebiprogrammi arhitektuuri kujutatud graafiliselt (C. Castillo, 2004):



Joonis 1- Tüüpiline kõrgtaseme arhitektuuriga veebiprogrammi (C. Castillo, 2004) .

Joonisel on välja toodud tüüpiline kõrgtaseme arhitektuuriga veebiprogrammi, millele on sisse ehitatud planeerija ning mitmelõimuline allalaadija.

2.3 Kasutus

Roomajate ehitused võivad üksteisest paljuski erineda olenevalt sellest, mis eesmärgiga vastav programm on loodud. Tihti jagatakse veebiprogrammid kaheks, andes ülevaatliku kirjelduse kahest laiaast kasutusvaldkonnast:

Kõige levinumaks ja tähtsamaks veebiroomaja kasutusvaldkonnaks peetakse **otsingumootoreid**. Samal ajal, kui üks osa programmist veebi süstemaatiliselt läbi roomab, indekseerib teine osa programmist läbi käidud leheküljed.

Teine laiem kasutusvaldkond on **andmekaeve** (inglise keeles *web data mining*). Andmekaeve puhul on tegemist meetoditega, kus käiakse läbi erinevad veebilehekülgi, leidmaks varem ette määratud märksõnu. Pikaajaliselt sarnase informatsiooni kogumist nimetatakse ka monitooringuks.

Asjale sügavamalt lähenedes saab veebiroomajate kasutuse jagada aga viieks spetsiifiliseks valdkonnaks, millel kõigil on oma erinev eesmärk ja süsteemi ehitus (C. Castillo, 2004):

Breadth-First Crawler: Suure otsingumootori või veebiarhiivi (B.Kahle, 1997) rajamiseks kasutatavad kõrge suutlikkusega roomajad. Seda tüüpi roomajad alustavad väikese veebilehtede arvuga ning laiendavad sealt edasi oma otsingut, hargnedes mööda lehekülgedel olevaid linke. Iga lehekülge salvestatakse enne läbi lingi järgmisele liikumist.

Recrawling Pages for Updates: Peale lehekülgede algset alla laadimist, võib esineda vajadus nende sisu üle kontrollida või uuendada. Sellisel juhul võib vabalt kasutusele võtta *Breadth-First Crawler*'i või saates uue päringu kõikidele kogutud URL'idele. Kuid leidub ka eraldi tehnikaid, mille abil on värskena võimalik hoida tähtsamad leheküljed, veebisaidid või domeenid. Korralikud taasroomamisstrateegiad on hädavajalikud näiteks indekseeritud otsingute puhul, millel on piiratud roomamismaht (Roomajale on ette antud väikene ala, et seda saaks võimalikult kiiresti läbi töötada ja seega anda otsijale võimalikult värsket tulemust). Täpsemaid strateegiaid andmebaasi värskena hoidmise kohta leiab J.Cho ja H. Garcia-Molina 2000. aastal avaldatud tööst "Synchronizing a database to improve freshness".

Focused Crawling: Fokuseeritud roomamise eesmärgiks on leida võimalikult palju soovitud kriteeriumitega lehekülgi, sealjuures kasutades võimalikult vähe andmemahtu. Sellised kõrgemal tasemel otsingumootorid omavad võimalust roomata kindla suunaga, sihtides vaid ette määratud lehekülje tüüpe, aruteluteemasid, keeli, pilte, mp3 failiformaate, teadustöid vms. Ka selle tegutsemise viisi puhul on optimaalsel struktuuril äärmiselt tähtis roll, läbi aastate on välja arendatud erinevaid lähenemisi, mida on käsitletud S. Chakrabarti, M. van den Bergi, ja B. Domi 1999. aastal avaldatud töös "Focused crawling: A new approach to topic-specific web resource

discovery” ning J. Rennie ja A. McCallumi 1999. aastal avaldatud töös “Using reinforcement learning to spider the web efficiently”.

Random Walking and Sampling: See meetod on mõeldud olemasolevate lehekülgede testimiseks ja otsingumootorite suuruse ning kvaliteedi hindamiseks. Muudes valdkondades on selle metoodika kasutus minimaalne. Põhiliste suvaliselt veebigraafe läbi „jalutavate“ tehnikate kohta on kirjutatud näiteks Z. Bar-Yossefi, A. Bergi, S. Chieni, J. Fakcharoenpholi ja D. Weitz 2000. aastal ilmunud töö “Approximating aggregate queries about web pages via random walks” ning M. R. Henzingeri, A. Heydoni, M. Mitzenmacheri ja M. Najorki poolt 1999. aastal ilmunud töö “Measuring index quality using random walks on the web”.

Crawling the „Hidden Web“: Palju veebis leiduvast informatsioonist paikneb andmebaasides ja on kättesaadav vaid õigete päringute esitamisel või andmete sisestamisel. Hiljuti on suurenenud huvi sellisele informatsioonile automaatselt ligi pääsemisele, ehk „Hidden Webile“, „Deep Webile“ või „Federated Facts and Figuresile“. S. Raghavani ja H. Garcia-Molina poolt 2001. aastal avaldatud töö „Crawling the hidden web“, heidab põhjaliku pilgu erinevatele tehnikatele, mida selliste andmete hankimiseks kasutatakse.

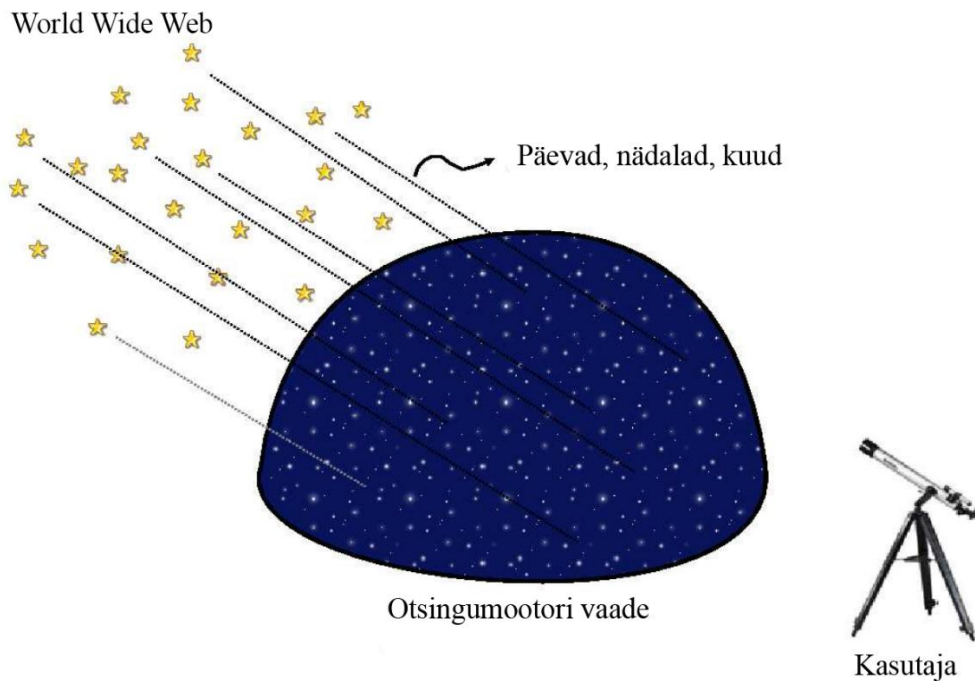
2.4 Esinevad probleemid

Veebiroomamise muudavad keeruliseks peamiselt kaks elementi: suur maht ja uuenemise kiirus, kui iga päevaga lisatakse ja eemaldatakse tohutul hulgal veebilehti. Ühtlasi tuleb arvestada, et pakutavad Interneti kiirused on ajaga vähem edasi arenenud kui protsessorid ja salvestusvõimalused. Suuremad andmemahud tähendavad aga seda, et roomaja suudab korrigeerida salvestada vaid osa veebilehest teatud aja jooksul, nii et ta peab suutma ise prioritseerida olulisema informatsiooni mitteolulisest. Kiire lehekülgede uuenemine võib endaga kaasa tuua olukorra, kus roomaja viimase leheküljeni jõudmise ajaks on esimesele lehele lisatud juba uut informatsiooni või siis see kustutatud.

Veebi läbi roomamist on võrreldud tähistava vaatamisega (C. Castillo, 2004) (joonis 2): pilt mida me näeme on vaid peegeldus tähtede olekust teises ajas, kuni valgus erinevaid vahemaid

läbib. Samamoodi ei saa veebiroomaja abil otsest hetkeolukorra pilti veebist, sest seda täpset ajahetke on võimatu tabada (R.Baeza-Yates ja B.Ribeiro-Neto, 1999).

Roomamine võtab aega ja veeb on äärmiselt dünaamiline, otsingumootori poolt kuvatav tulemus veebist esindab hetkeolukorrast erinevaid ajalisi seise. See on sarnane tähistaeva vaatamisele, kus meie poolt nähtavad tähed ei eksisteeri tegelikult sellel sama ajahetkel (C.Castillo)



Joonis 2- Roomaja töö võrdlemine tähistaevaga (C. Castillo, 2004).

Arvutiteadlane J.Edward on märkinud, „Võttes arvesse, et veebiroomamiste läbiviimiste mahud ei ole piiritletud ega tasuta, on muutunud hädavajalikuks roomata veebi mitte ainult võimalikult laiaulatuslikult, vaid sealjuures ka efektiivselt, kui soovitakse säilitada veidigi kvaliteeti ja informatsiooni värskust (J. Edwards, K. S. McCurley, ja J. A. Tomlin. 2001). Veebiroomaja tulemus ongi kombinatsioon järgnevatest eeskirjadest (inglise keeles *policy*) (C. Castillo, 2004):

- **Valimiseeskiri** (*A selection policy*), määrab ära milliseid lehti alla laetakse.
- **Taaskülastamiseeskiri** (*re-visit policy*), määrab ära millal külastatud lehekülgi uuesti kontrollitakse.

- **Viisakuseeskiri** (*Politeness policy*), määrab ära kuidas vältida külastatavate lehekülgede ülekoormust.
- **Paralleelsuseeskiri** (*parallelization policy*), määrab ära kuidas toimub paralleelne roomamine ja selle töö koordineerimine.

3 Veebiroomaja rakendus meediaagentuuris

Kui eelnevalt tutvustati erinevaid tegevusi, mida tänapäeval meediaagentuurilt oodatakse ja veebiroomajate põhimõtteid, siis käesolevas peatükis pannakse need kaks valdkonda kokku ja vaadeldakse täpselt, kuidas veebiroomajad ja meediaagentuurid omavahel seotud on. Tutvustatakse meediaagentuuride infootsinguga seostuvaid vajadusi ja nõudeid. Vajaduste selgitamisel põhineb autor oma enda isiklikel tööalastel kogemustel antud vallas ning läbi viidud vestlustel oma ala tippspetsialistidega.

3.1 Meediamonitooring

Iga digitaalselt aktiivne firma on huvitatud, milline info nende kohta Internetis liigub ja mida inimesed omavahel nende brändist räägivad. Suhtlemist võimaldavatel lehekülgedel roomates on ettevõtetel võimalus koguda enda jaoks väärtuslikku informatsiooni, mis on tänapäeval äärmiselt hinnatud ressurss. Meediamonitooring ehk mediaseire, on veebiroomajat kasutav protsess, mis on mõeldud sotsiaalmeedia ja teiste RSS (ing.k *Rich Site Summary* või *Really Simple Syndication*) lehekülgede jälgimiseks. See tähendab, et roomaja saadetakse informatsiooni otsima regulaarselt, kindlate otsingusõnadega, piiritletud veebikeskkonda, kust ta peale igat otsinguprotsessi lõppu kõik otsingusõnadega seostuvad leheküljed välja toob. Selliste regulaarsete otsingutega on võimalik silma peal hoida peaaegu kõigel meedias toimival, seega monitooringu erinevaid kasutusviise leidub palju (promptcloud).

Toodete hindamine ja arvustamine – Võimalik on enda poolt valitud toote kohta saada kõik sotsiaalmeedias ja RSS'is kirjutatud informatsioon, olgu see positiivne või negatiivne. See annab võimaluse saada ilma kõrvaliste mõjutajateta teada, mida inimesed tootest arvavad.

Tagasiside – Peale ürituste või kampaaniate korraldamist on võimalik kuulata inimeste tagasisidet, saades teada, mida nad arvasid ja tehes sealt edasi järeldusi, mida teine kord muuta. Sama kehtib ka uute toodete välja laskmise puhul, mille kohta roomates äärmiselt kiirelt ja palju arvamusi saab.

Etteheited – Pidev monitooring annab võimaluse saada kiirelt teada, kui klientidel on etteheited mõne toote või teenuse kohta. See annab võimaluse kiireks reageerimiseks, olukorra

selgitamiseks või vabandamiseks. Lisaks saab nii informatsiooni minevikus esinenud kaebuste üle, mis annab teadmiseid edasiseks arenguks.

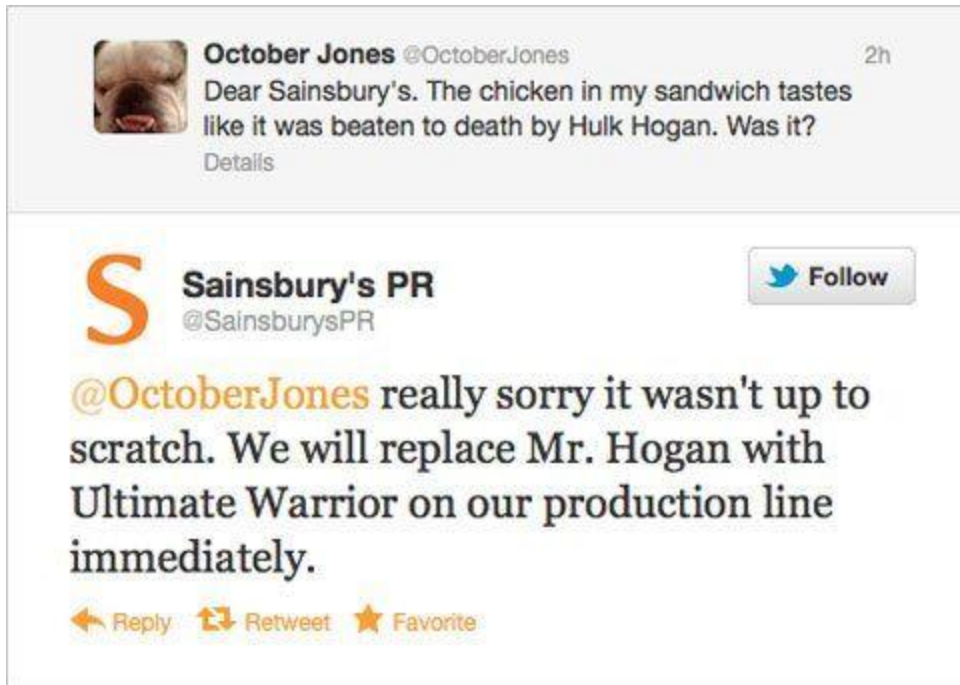
Konkurentide jälgimine – Täpselt nagu enda brändi, on võimalik jälgida ka oma konkurentide kohta liikuvat informatsiooni ja nende tegemisi. Nii on võimalik saada hea ettekujutus nende põhilistest sotsiaalmeedia eesmärkidest või kasvõi anda kiire vastulööök mõne nende kampaania algusfaasis.

Maine jälgimine – Spetsiaalselt seadistatud veebiroomajatega on võimalik automaatselt tuvastada otsingu sõna kohta käiva teksti sentimentaalsust, ehk kas tegemist on negatiivse või positiivse sõnumiga. Kui selline automaatne analüüs ei pruugi alati 100% täpne olla, siis ise aeg-ajalt brändi kohta käivatel sõnumitel silma peal hoidmine annab üsna hea ettekujutuse, milliseks üldist mainet peetakse.

Sihtrühma leidmine – Sisestades tunnussõnadeks oma brändiga tugevalt seostuvaid sõnu, leiab roomaja kerge vaevaga sihtgrupid ja veebikeskkonnad, kus just nendel teemadel kõige aktiivsemalt räägitakse.

Sihtrühma huvialade leidmine – Tehes roomaja abil kindlaks, millistes erinevates veebikeskkondades sihtrühma esindajad peamiselt asuvad, on nii kasutajaid kui teemasid analüüsides teha järeldusi, millised teemad neid veel huvitada võiksid. See informatsioon annab võimaluse pakkuda klientidele neile meelepärase sisuga reklaami või sõnumeid.

Kogu selle informatsiooni teeb eriti väärtuslikuks fakt, et inimesed räägivad veebis brändidest ja toodetest vabatahtlikult ja täiesti ausalt. Nende samade inimestega on võimalik ühendust võtta ja suhelda, muutes brändi mainet ja tuntust. Hangitud informatsiooni üks brändile tulusamaid kasutusvõimalusi on kasutajate poolse sisu (UGC - *User-Generated-Content*) rakendamine. (vt. joonis 3).



Joonis 3- Sainsbury's PR'i vastamine Twitteri Sainsbury's kaubandusketi kasutajakonto jälgijale. Tähelepanu juhtida asjaolule, et kliendi säuts ei ole suunatud otse Sainsbury kasutajatunnusele. (Twitter)

(Est.k – October Jones: “ Kallis Sainsbury’s. Kanavõileib maitstes nagu seda oleks Hulk Hogan klobinud. Oli nii?” Sainsbury’s PR: „Mul on kahju, et su võileib ei olnud viimasepeal. Me vahetame oma tooteliinil oleva Mr.Hogani koheselt Ultimate Warriori vastu“).

See on hea näide UGC korrektsest kasutamisest, Sainsbury’s ei pidanud ise midagi looma ega planeerima, vajalik oli vaid saada kätte vajalik informatsioon ja sellele õigesti vastata.

3.2 Monitooringu protsess

Kogu tehniline protsess, mis monitooringu alla läheb on keerulisem kui peale vaadates arvata võiks. Kvaliteetsete tulemuste saamiseks tuleb läbi mõelda kõik detailid, muidu võib tekkida väga kergelt olukord, kus kokku roomatud informatsiooni ladestub nii suurtes mahtudes, et sealt vajaliku leidmine osutub äärmiselt keerukaks. Hea monitooringu protsess seisnebki väga suurel määral täpsel roomaja seadistamisel, mis tagab, et kogu informatsioon, mis jälgimisele saadakse,

on vajalik ja kasutatav. Näiteks saates roomaja ilma igasuguste piiranguteta vaid märgusõna „Kalev“ otsima, oleks pool või rohkem kogutud informatsioonist kommivabriku jaoks automaatselt kasutu, seda ei teataks aga enne, kui läbi oleks vaadatud kõik roomamisest saadud tulemused. Erinevatel Roomamismootoritel ja teenustel on väga erinevad võimalused, mida kõiki tuleb oma eesmärgi seades silmas pidada. Lihtsa monitooringu seadistusprotsess Eesti otsingu puhul on näitlikuna järgnev.

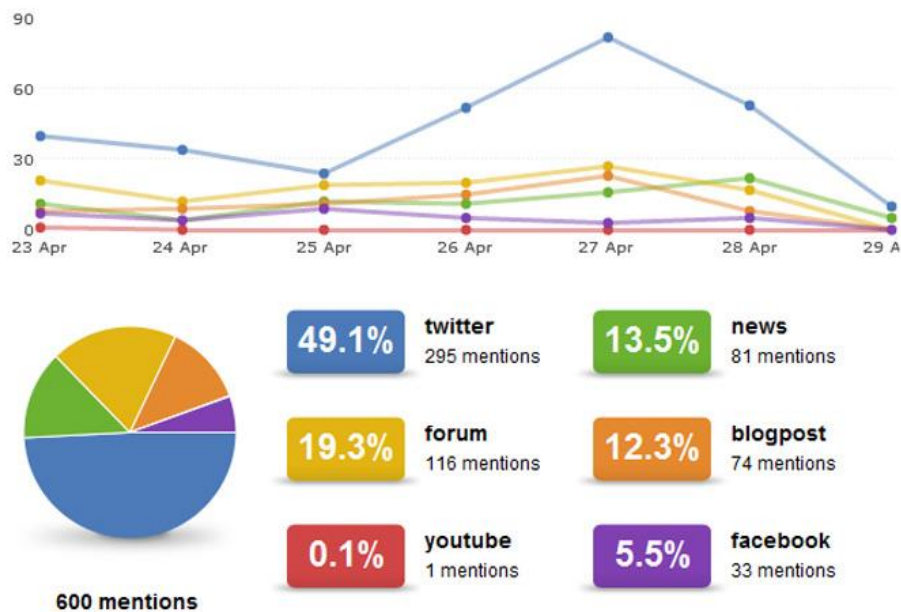
Esmalt mõeldakse läbi bränd ja sellega seostuvad tunnussõnad, tüüpiliselt valitakse ühe teema raames kuni 10 tunnussõna. Seejärel vähendatakse otsitavat ala, kiirema tulemuse saamiseks. Tüüpiliselt valitakse otsitavateks veebikeskkondadeks näiteks Facebook, Twitter, Postimees, Delfi, jms. Veebikeskkondi lisades tuleb kindlasti arvestada oma firma olemusega. (Näiteks viies läbi otsingut Eesti arvutimängurite kohta ning pannes veebikeskkonnaks Buduaari foorumi, on tegemist suure tõenäosusega maha visatud ressursiga.) Peale seda määratakse ära otsingute tihedus, ehk mitu korda ja millal otsinguid tehakse. Viimasena märgitakse ära, kuhu informatsioon saadetakse, mis on tihti mõne kliendi või töötaja e-maili aadress. Valitud aadressile või aadressidele hakkavad varem märgitud tihedusega saabuma ülevaated soovitud märksõnade ilmumise kohta (ehk raportid), vastavalt määratud seadistustele. Väljavõtte mõnest tulemusest mitme tunnussõnalise teema raportis võib välja näha näiteks selline (Joonis 4).



Joonis 4- Väljavõtte roomaja raportist (isiklik pildipank).

Raportis on välja toodud varem lisatud otsingusõnad, mille põhjal tulemused leiti, leheküljed, kust informatsioon on leitud, kellaaeg, millal otsing on läbi viidud ning lühike ülevaade kontekstist. Samamoodi, nagu on erinevad monitooringuprogrammide seadistusvõimalused, on

erinevad ka raportid, mille kaudu infot otsijale edastatakse. Otsingutulemuste saatmine e-mailile on enamasti ainult veebipõhiste jälgimisteenuste puhul kohustuslik. Suuremate ja keerukamate programmide puhul salvestatakse kogu leitud informatsioon ka arvutisse või serverisse, kus seda otse jälgida ja töödelda saab. Kui eelnevalt oli välja toodud näide raportist ühe kindla teenuse puhul (Joonis 4), siis mõne teise teenuse esitus võib sellest täielikult erineda (Joonist.5.).



Joonis 5- Meediamonitooringu raporti esitamise graafiliselt (neilpatel).

Selles raportis on välja toodud ühe tunnussõna esinemine veebis, kuid seda on tehtud ülevaatlilikumana, pikemas perspektiivis. Välja on toodud tunnussõna kõikumine kuupäevade kaupa, läbi erinevate meediakeskkondade. Ära on märgitud, kui populaarne tunnussõna igas keskkonnas oli ja kogu ajaperioodil neis kanalites mainimiste arv. Selline kuju annab küll kiire ja ülevaatliku pildi tunnussõna levikust, kuid täpsemat sihtrühma või nende teisi huvialasid on läbi sellise meetodi võimatu leida.

Kui kergema monitooringu saab korda saata vaid üksiku tunnussõnaga, siis maksimaalse tulemuse saamiseks on vaja paljusid võimalusi toetavat roomajat. Kui just toodi välja lihtsa monitooringu seadistus otsingu jaoks, siis järgnevas lõigus käsitletakse juba keerulisemat monitooringut. Eesti juhtiva meediaagentuuri digitaalmeedia spetsialistiga suheldes selgusid tema arvates põhilised tegurid, mis on vajalikud kvaliteetse info hanke puhul.

Protsessi kiirus - Elades kiir uudiste ajastul, kus sageli on uudis enne üleval kui kõik faktid ja asjaolud on selgunud, on info kättesaamisele kuluval ajal kriitiline väärtus. Meedia valdkond on armutu ja isegi pooletunnine edu konkurentide ees võib osutada mõne uudise või kampaania puhul määravaks.

Kindlate allikate valik – Eelnevalt väljatooduna on veeb liiga sügav, et seda kõike võimalik jälgida oleks. Et saavutada suurim kiirus ja värskus informatsiooni hankimise puhul, on vajalik ära märkida vaid vajalikud allikad, ehk URL-id, kust informatsiooni otsima hakatakse.

Keele filtreerimine – Programmiga peab olema valiku korral võimalik kuvada vaid tulemusi, mis esinevad eesti keeles või eesti keelses keskkonnas.

Otsingusõnade väljafiltreerimise võimalus – Mõnikord võib ühe otsingusõnaga seostuda erinevaid teemasid, kuid nendest vajalik on vaid üks teema. Sellisel juhul on vajalik võimalus lisada mittesoovitud teema kohta väljafiltreerimise jaoks sõnu, mille koosinemisel otsingusõnaga tulemust ei registreerita. Näiteks otsides Yamaha muusikainstrumentide kohta informatsiooni, oleks mõttekas välja filtreerida kõik mototehnikaga seostuv.

Automaatne raporti edasi saatmine – Nii on võimalik saada kätte tulemused ükskõik kus viibides ning samamoodi saata tulemusi piiramatule hulgale seotud inimestele ilma igasuguse vaevata.

Raporti saatmine kellaajaliselt – Kasulik oleks automaatselt roomamistulemusi saata iga päev või vähemalt tööpäevadel täpselt samal ajal. Selline süsteem säilitab paremat korda, teades, et e-mailide vaheline periood katab täpselt ühe ööpäeva tulemused. Loomulikult oleks endiselt võimalus sooritada lisaotsinguid ka graafikuväliselt.

Mitme otsingusõnast koosneva teema loomine – Iga otsitav teema peaks koosnema rohkematest sõnadest, kui põhiline tunnussõna. Teema alla võiks kuuluda umbes kuni 10 märksõna, mille kõigi tulemused kajastatakse koos ühes raportis, sama teema all (vt. joonis 4).

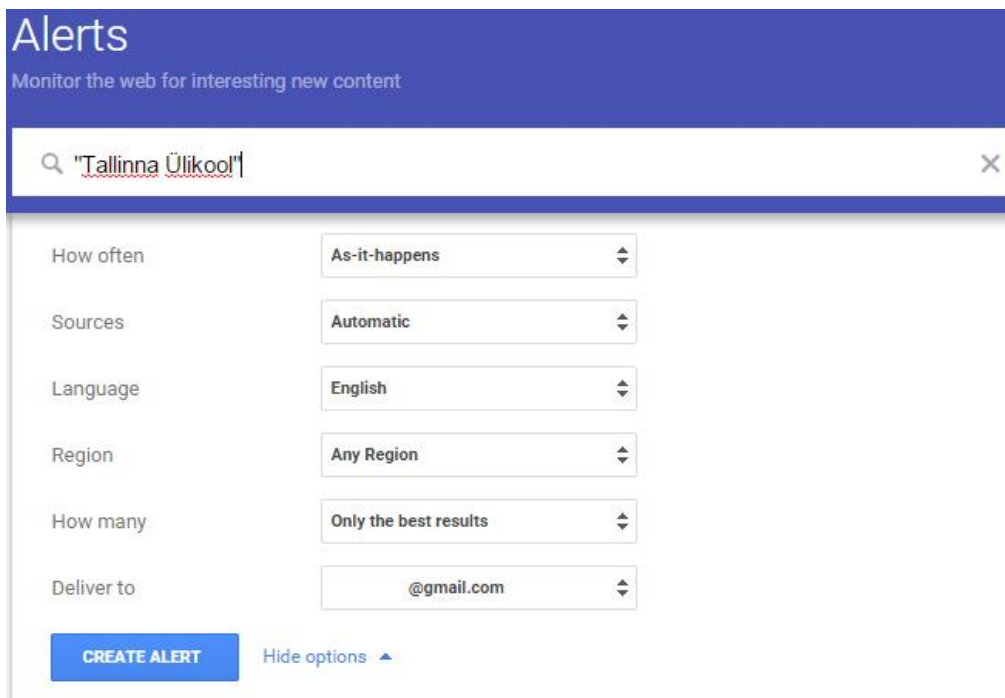
Alarmsõnade määramine – Kui saatmisgraafiku alusel saadetakse automaatselt raport näiteks 20 erineva teema kohta, oleks vajalik määrata ka nn alarmsõnad, mille roomatavatelt lehtedelt leidmise korral teavitatakse otsingu sooritajat näiteks e-maili teel. Alarmsõnu peaks saama määrata ka hulgana, andes märku vaid kõikide alarmsõnade esinemisel. Näiteks seades ühe

teema alla alarmsõnaks vaid „Pariis“, ei ole sellest palju abi, seatuna valve alla aga koos sõnaga „Rahutused“, tuleb tulemus suure tõenäosusega ootuspärasem.

3.3 Erinevad teenused

Eelnevalt mainitud erinevad meediamonitooringu seaded ja hulgaliselt varieeruvaid raporteerimis viise panevad meediaagentuurid raskesse olukorda, sest nende seast tuleb valida just õiged. Paljude klientide puhul ei ole tasuliste teenuste poolt pakutavad kõik võimalikud variandid aga sugugi olulised. Leidub hulgaliselt tasuta meetodeid, mille asjakohasel koos ära kasutamisel on võimalik kätte saada piisav tulemus põhilistele olukordadele. Tasuta leiduvatest meetoditest tasuks kaaluda alljärgnevat võimalusi.

Korralikku tasuta monitooringu teenust pakub näites Google, Google Alerts'i näol (Joonis 6).



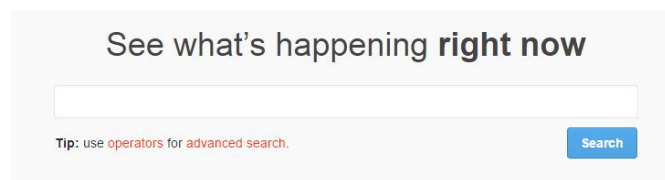
The image shows the Google Alerts configuration interface. At the top, it says 'Alerts' and 'Monitor the web for interesting new content'. Below that is a search bar containing the text '"Tallinna Ülikool"'. Underneath the search bar are several dropdown menus for configuration: 'How often' is set to 'As-it-happens', 'Sources' is 'Automatic', 'Language' is 'English', 'Region' is 'Any Region', 'How many' is 'Only the best results', and 'Deliver to' is '@gmail.com'. At the bottom left is a blue 'CREATE ALERT' button, and at the bottom right is a 'Hide options' link with a small upward-pointing triangle.

Joonis 6- Google Alerts'i seadistuspaneel (Google).

Google Alerts on teenus, mis annab otsijale märku, kui valitud märksõna on indekseeritud Google'i otsingumootori poolt. Seadistades on võimalik valida, kas teateid toimunust antakse edasi kord päevas, nädalas või koheselt, kui tulemus leitakse, enda poolt täpsemat aega määrata ei saa. *Sources*, ehk allikate alt saab määrata kust informatsiooni otsitakse, näiteks blogidest

(*Blogs*), uudistest (*News*), veebist (*Web*), videodest (*Videos*) või veel mõnest teisest allikatest, oma valikuid sisestada ei saa, samamoodi ei saa ka allikaid välja filtreerida, soovides näiteks informatsiooni saada kõikjalt peale Facebooki. Järgmisena saab valida soovitud keele (*Language*), mida arvestatakse tulemuste otsimisel. Peale seda saab kindlaks määrata otsingu Regiooni (*Region*). Järgmisena saab seadistaja paika panna tulemuste arvu (*How many*), mille alt saab valida vaid parimate vastuste kuvamise (*Only the best results*) või kõikide tulemuste (*All results*) vahel. Viimasena tuleb otsijal sisestada e-maili aadress, kuhu tulemusi saatma hakatakse. Autori arvates peaks tasuta teenuste kasutamise puhul Google Alerts moodustama kindla põhja, mille ümber vastavalt vajadusele rakendada ka teisi võimalusi.

Ka **Twitter** pakub kasutajatele enda poolt monitooringu võimalus, mis sisaldab ka täpsema otsingu võimalust (*advanced search*) (vt. Joonis7).



Joonis 7- Twitteri otsingumootor (Twitter).

Kuid Twitteri otsing pole veel Google'i omaga võrreldes nii arenenud, seega tuleb näiteks manuaalselt ära märkida erinevad kääneldõpud, kasutades „OR“ sõna. Keelte alt ei ole võimalik valida eesti keelt, paika ei saa panna automaatset tulemuste edastamist, otsingutulemusi ei ole võimalik filtreerida ning kogu tulemuste otsimine toimub vaid läbi Twitteri kanali. Autori arvates tuleb Twitteri kasutamine abiliseks põhiliselt globaalsete brändide puhul, sest kogemused on näidanud, et eestlaste seas ei ole see kanal veel endiselt eriti populaarne.

Facebooki monitooringuks kasutamise puhul tuleb arvestada, et võrreldes näiteks Twitteriga, on nad tunduvalt kinnisemad oma informatsiooni levitamise suhtes (näiteks ei saa tellida otsingutulemusi RSS'iga). Lisaks tuleb arvestada, et siiani on enamus inimeste profiilid võõrastele ligipääsematud. Otsingutulemusi ei ole võimalik tellida e-maili teel, seega tuleb neid lihtsalt manuaalselt teatud aja tagant kontrollimas käima (Okia). Autori arvates tasub Facebooki abi kindlasti rakendada sihtgruppide parema ülevaate saamiseks, nii on võimalik kerge vaevaga teada saada, millised leheküljed, filmid, muusika või tooted veel brändi fännidele meeldivad.

Socialmention on viimane veebikeskkond, mis seoses tasuta meediamonitooringu võimalustega vaatlemist vajab. Socialmention annab statistilise ülevaate sisestatud otsingusõna esinemise kohta. Numbriliselt tuuakse välja otsingusõna „tugevus“ (ing.k *strenght*), mis näitab protsentuaalset võimalust, seda sotsiaalmeedias arutatakse, mida suurem protsent, seda aktiivsemalt on sõna kasutuses. *Sentiment*, ehk sentimentaalsus on näitaja, mis toob välja otsingusõna positiivsetes ja negatiivsetes situatsioonides mainimise vahekorra (Joonisel 9, on iga negatiivse mainimise kohta 3 positiivset mainimist, ehk 3:1). Kiindumus (ing.k *passion*) on näitaja, mille protsent määrab ära tõenäosust, et sellel teemal juba sõna võtnud isikud seda uuesti teevad. Viimane peamine näitaja on *reach*, ehk haardeulatus, mis näitab kui laia erineva publikuni antud tunnussõnaga seostuvad teemad levivad. Lisaks sellele on tulemuste lehel välja toodud keskmine mainimis kordade kiirus, viimane mainimine, erinevate autorite arv, Twitteril põhinevad vastamised, positiivse, negatiivse ja neutraalse alatooniga mainimiste arv ja peamised otsingusõnad, mis antud märksõnaga seoses populaarsed on. Lisaks sellele on võimalus saata kogu informatsiooni otse RSSi või tellida märguandeid oma e-mailile. Leheküljel näiteks asukoha valik, keele valik ei toimi korrektselt, sisestada ei saa ka soovimatuid allikaid ja olemasolevate allikate valik on täpseks kasutamiseks liiga üldine. Lehekülge külastades ei olnud töökorras ei “*RSS Feed*” ega “*Email Alert*” võimalused. Eesti keelsetele otsingutele tuli äärmisel vähe tulemusi, mis tõstatab ka küsimuse, millistest allikates nende roomaja oma infot kogub. Autori arvates võib seda rakendust kasutades leida küll huvitavat informatsiooni, kuid Eesti turul põhjaliku ja usaldusväärse monitooringu läbi viimiseks ei tundu see ainsa vahendina piisavalt kindel.



Mentions about President

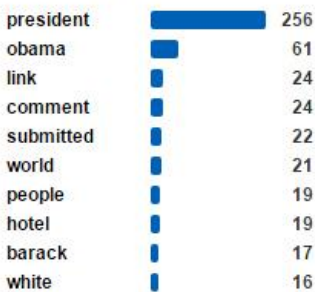
Sort By: Results:

Results 1 - 15 of 170 mentions

Sentiment



Top Keywords



- [President](#)
 President - Alanland2020.jpg President Looking
s690.photobucket.com/albums/vv264/republicofalanland/?action=view&at=Alanland2020.jpg
 2 minutes ago - by republicofalanland on [photobucket](#)
- [chester-a-arthur-ole-peter-hansen-balling](#)
 chester-a-arthur-ole-peter-hansen-balling - chester-a-arthur-ole-peter-hansen-b.jpg US President
[s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US President/?ac](https://s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US%20President/?ac)
 2 minutes ago - by elanglikaytanjua on [photobucket](#)
- [bush_1-1.jpg](#)
 bush_1-1.jpg US President
[s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US President/?ac](https://s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US%20President/?ac)
 2 minutes ago - by elanglikaytanjua on [photobucket](#)
- [warren_g_harding](#)
 warren_g_harding - warren_g_harding.jpg US President
[s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US President/?ac](https://s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US%20President/?ac)
 2 minutes ago - by elanglikaytanjua on [photobucket](#)
- [grant](#)
 grant - grant.jpg US President
[s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US President/?ac](https://s1006.photobucket.com/albums/af190/elanglikaytanjua/elanglikaytanjua/US%20President/?ac)
 2 minutes ago - by elanglikaytanjua on [photobucket](#)
- [tjefferson](#)
 tjefferson - tjefferson.jpg US President

Joonis 8- Tulemus otsingusõnale "President" (Socialmention)

Teenuste alla tuleb arvestada ka eraldi info töötlus meetodid, mille abil roomamistest kätte saadud informatsiooni praktiliselt vaadelda ja analüüsida. Üks sellistest tasuta võimalustest oli pikka aega Google Reader, mille abil on võimalik kõik oma Google Alertsist, Twitteri otsingtest ja RSS'i voogudest tulev informatsioon tõsta kaustadesse, ülevaatlikumaks jälgimiseks, kuid nüüdseks on see suletud (googleblog).

Suur hulk veebilehti (nt. Gizmodo, Lifehacker) soovivad sobivaks alternatiiviks rakendust nimega **Feedly**. Feedly't peetakse üheks parimaks variandiks neile, kellel on vaja ühte andmete hoidjat kõikidele oma seadmetele. Lisaks Android'i ja iOS'i toetamisele on neil olemas ka oma veebipõhine lehekülg. Nende rakenduse abil saab luua iga oma jälgitava teema jaoks eraldi

kaustad, kuhu informatsioon laekub, andes nii selge ja konkreetse ülevaate. Selline kasutus väldib inforägistikku ja kuhjumist näiteks e-maili kontole, kuhu, kuhu kõik roomajad oma info edastavad. Autori poolt üks eelistatumaid programme selles valdkonnas, kasutajaliides on selge ja puhas ning meenutab stiili poolest Gmail'i konto visuaalset poolt.

Teise alternatiivina on levinud **Digg Reader**, mis töötab sarnasel põhimõttel. Rakendus on võimalik siduda oma personaalse või töö Gmail'i kontoga, mis muudab kasutamise äärmiselt kergeks. Digg'ile on võimalik lisada ka otsene ühendus Facebooki või Twitteriga, postitades või vastates soovi korral otse läbi selle platformi. Kui Feedly puhul võis välimuse poolest paralleele tuua Gmail'iga, sarnaneb Digg readeri välimus iOS kasutajaliidesega, samamoodi on rõhk puhtusel ja elegantsusel.

Viimasena sobib ülesande jaoks **NewsBlur**, mis pakub lisaks oma veebilehele ka nutiseadmete kasutajatele nii Android'i kui iOS'i rakendust. Tasuta kontod saavad sisse tuua informatsiooni kuni 64'st allikast, millest peaks piisama igale kergemale kasutajale, kuid asjaga tõsisemalt tegeledes maksab saab piiramatute võimalustega kasutada rakendust 1\$ kuutasu eest (gizmodo).

Reaalsuses jõuab nende kolme rakenduse vahel valimine lõpuks puhtalt maitsele ja harjumusele. Praktiliselt töötavad kõik eespool mainitud programmid sarnaselt, üksikute kasutusharjumuste erinevustega. Kõik rakendused on kasutatavad täiesti tasuta, seega parima tulemuse saamiseks on kasutajal soovitatav läbi proovides leida enda jaoks sobiv variant. Autori maitse jaoks on NewsBlur liiga robustne, kõik vajalikud funktsioonid on küll olemas, kuid kasutajaliides jätab aegunud ja kohmaka mulje.

4 Sobivaimate teenuste kombineerimine

Kogunenud informatsiooni põhjal on selge, et kiiresti areneval digitaalajastul on meediamonitooringul oluline roll. Olgu selle eesmärgiks sihtgrupi paremini tundma õppimine, tagasiside leidmine, üldise maine jälgimine, UGC leidmine või midagi muud, kasu on sellest igale firmale, olgu tegemist suurema, väiksema, aktiivse või passiivsema ettevõttega. Erinevad veebiroomamis võimalused jagunevad kiiresti laiali, pakkudes erinevaid võimalusi neile kõigile. Seega on oluline pakkuda hästi valitud meediamonitooringu teenuseid õigetele klientidele.

4.1 Monitooringu loomine

Õige roomaja leidmise puhul tuleb ennekõike kliendi eesmärkide ja olemusega. Tegeledes suurte kavatsuste ja plaanidega aktiivse firmaga, tuleb arvestades peaaegu kõikide 3. peatükis välja toodud vajadustega, mis on esitatud roomamis meetodil info hankimise puhul veebiroomajale. Neid vajadusi arvestades filtreeritakse juba välja suurema osa kergemaid ja tasuta pakutavaid monitooringu teenuseid. Paljud sellised teenused töötavad vaid ühe otsingusõna põhiselt ja ilma mingisuguse tsükli planeerimisvõimaluseta (näiteks Socialmention ja Twitter'i otsing). Ammugi ei saa sarnastel lehekülgedel valida täpseid otsingukanaleid, välja ei filtreerida sõnu ega allikaid. Kui kõrvale jätta kõik tasuta ja veebipõhised pakkumised, jääb järele umbes kümme tõeselt võetavat teenusepakkujat (näiteks Mediavantage, Lithium ja Exacttarget), kes kõik pakuvad põhimõtteliselt sama teenust, kuid natuke erinevate nüanssidega. Autori agentuuri poolt kasutatavat teenusepakkujat siinkohal välja ei tooda.

Tegeledes klientidega kellel on väiksemad soovid, piisab nende vajaduste rahuldamiseks ka vaid vabavaralistest teenustest, pakkudes neile ülevaadet nende brändi maine kohta, analüüsides nende sihtrühma huvialasid, mida kampaaniatega siduda ning tuues välja uusi potentsiaalseid sihtrühmi. Eelmises peatükis mainitud Google Alerts, koos mõne teise 3. peatükis mainitud tasuta monitooringuteenusega peaksid olema piisavad kõikide vähenõudlike klientide jaoks. Suurim osa tööst jääb kõige laiemate võimalustega Google Alerts'ile, kuid see töötab vaid Google'i enda indekseeringute tuvastamise põhjal, mis tähendab, et süstemaatiliselt maksab aktiivselt kasutada ka teisi mainitud teenuseid, tagades nii täieliku info kättesaamise. Lisaks on

tungivalt soovituslik võtta kasutusele mõni 3. peatükis välja toodud monitooringu abiprogramm, näiteks Feedly, mille abil erinevatest tasuta teenustest sisse jooksvad informatsiooni paremini hallata.

Suuremate klientide puhul tuleks veebiroomamist teenust alguses veel edasi sisse osta, kuni nõudlus kõrgemal tasemel monitooringu vastu muutub suuremaks. Sisse ostetud teenuse hind on küll odav, kuid pikas perspektiivis tekib rohkem kliente, kes soovivad peale standart lahendustele ka põhjalikumat analüüsi, võib see märkamatult kuhjuda, ületades tulususe piiri. Kui monitooringu laiem ja läbimõeldum tutvustamine osutub edukaks, on põhjust mõelda ka oma veebiroomaja arendamisele, mis tagaks kõikide vajaduste täitmise, ning kujutaks endast vaid ühekordset suuremat investeringut. Kuna töös käsitletava agentuuri poolt seatud standardi on kõrged, ei piisaks siinkohal lihtsa fokuseeritud roomaja loomisest. Muidugi oleks projekti käiku minekuga vaja läbi viia roomaja haldamis koolitused teatud töötajate seas ja neid minimaalse koormusega alati asja juures hoida, kuid kui teenus osutub tulevikus veelgi edukamaks, on alati võimalus suurendada monitooringuga tegelevate inimeste arvu ja osakoormust.

4.2 Teenused vastavalt klientide vajadustele

Arvestades, et meediamonitooringut saab pakkuda väga erinevatel viisidel, olgu selleks siis pigem passiivne jälgimine, mille kõrvalt arvestusi teha ja informatsiooni koguda või aktiivne jälgimine, kus leitud informatsiooni põhjal reageeritakse koheselt, võttes sõna viimaste sündmuste osas või vastates oma brändiga seostuvatel teemadel sõna võtvatele inimestele, tuleb igale brändile läheneda individuaalselt, luues just nende jaoks sobiv jälgimisstrateegia. Näiteks firmade puhul, kelle peamiseks sihtgrupiks on noored, võib sõna võtta ka kõige kummalisemates olukordades, tutvustades neile oma brändi vaba ja avatud poolt. Samasugune lähenemine ei pruugi töötada aga vanematele naisterahvastele suunatud brändi puhul, kus võiks fookuses olla pigem konkreetsete vajaduste analüüs ja teemaga seotud küsimustele vastamine.

Kuna terve meediamonitooringu valdkond on suhteliselt uus ja tundmatu, ei pruugi selle aktiivne kasutamine sobida kõigile, kuid minimaalsel tasandil selle ära kasutamine toob sisse kasulikku informatsiooni iga firma puhul. Peale lühiajalist kasutamist saab roomaja kaudu teada põhilise

sihtrühm, potentsiaalne sihtrühm, aktiivseimad ajad, huvi pakkuvad teemad, kui aktiivne on kasutajaskond ja palju muud. See tähendab, et tasuta rakendustel põhinevad monitooringu alused võiks sisse panna iga kampaania ja kliendi strateegiate puhul, saades nii hea ja kindlalt värsket ülevaade viimasest veebi ning sihtrühma olukorrast.

Roomajate abil monitooringut tehes ei saa mööda vaadata juba varem mainitud faktist, et tegemist ei ole konkreetsetes numbrites mõõdetava väärtusega. See muudab terve teenuse mahamüümist keerulisemaks. Kui muidu on igal teenusel oma hind, mille eest lubatakse kindel arv vaatamisi, klikke, külastusi või midagi muud, siis monitooringu puhul ei saa midagi konkreetset lubada. Nii võib klientidele tunduda, et nende raha läheks selle teenusega tuulde. Neid punkte silmas pidades saab teha järelduse, et eelistatav sihtgrupp terviklikuma monitooringu pakkumiseks on julged ja aktiivsed firmad, kes ei karda proovida uusi meetmeid, samas omades juba eelnevalt piisavalt kogemusi, et tunnustada terve protsessi tulutegurit ka ilma kindlate numbriliste näitajateta.

Ülejäänud firmadele tuleks pakkuda teiste teenustega koos ühekordset analüüsi teenust tasuta rakenduste põhjal, mille tulemuste hilisemal tutvustamisel on veel võimalik tekitada huvi ka pikaajalisema meediajälgimise vastu.

Kokkuvõte

Käesoleva töö eesmärgiks on anda põhjalik ülevaade veebiprogrammide olemusest ja kasutusest meediaagentuuri töös ning seeläbi leida lahendused ning funktsionaalsused, millest agentuuri kliendid kõige rohkem võidaksid. Lahenduste valik põhines nii suurimal funktsionaalsusel kui ka erinevate klientide reaalsetel vajadustel. Töö eesmärk sai saavutatud, leides lahenduse, et kõige mõttekam on kasutada sisse ostetud monitooringu teenust kõige aktiivsemate klientide jaoks, kes soovivad kasutusele võtta kõik võimalused, et tagada oma brändi või kampaaniate edu. Kõikidele teistele klientidele peaks agentuur koos teiste teenustega pakkuma ühekordset standard meedia monitooringut. Standard monitooring põhineks peamiselt tasuta Google Alerts'i rakendamisel koos mõne teise vaba teenusega, vastavalt kliendi olemusele, tagades nii piisavalt tugeva monitooringu põhja iga keskmise vajadusega firma jaoks. Nii on võimalik pakkuda parimaid võimalikke lahendusi aktiivsetele ja nõudlikele klientidele, kuid samas tutvustada uusi võimalusi ja vahendeid passiivsetele klientidele, kelles esimesed tulemused võivad teenuse vastu suuremat huvi äratada.

Tulemuse saamiseks tutvustati töös esmalt meediaagentuuri plusse, teenuseid ja erinevaid kliente koos nende vajadustega. Seejärel selgitati veebiprogrammide olemust, põhimõtteid, kasutusmeetodeid ja erinevuseid. Kolmandaks pandi kokku meediaagentuuri töö ja veebiprogrammid, tutvustades erinevaid viise, kuidas agentuurid kasutavad meediamonitooringut, vahendeid selle läbiviimiseks ja tulemusi. Viimases osas tuuakse välja soovitused, kuidas veebiprogrammidele põhinevaid teenuseid klientide erinevuseid arvestades kõige paremini ära kasutada.

Tulevikus on siinset tööd võimalik rakendada meediaagentuuride töötajatel, kes ei ole täiesti teadlikud kõikidest võimalikest võimalustest ja hüvedest, mida pakuvad veebiprogrammid. Nii saavad nad parema ülevaate erinevatest võimalustest ning oskavad neid teenuseid paremini ka klientidele soovitada ja tutvustada. Tööd on võimalik kasutada tulevikus tugeva põhjana meediaagentuuri oma programmi arendamisel, millel on olemas kõik agentuuri jaoks vajalikud funktsioonid. Seda aga juhul, kui töös välja toodud muudatused osutuvad pikas perspektiivis piisavalt tulemusrikkasteks, et oma programmi loomine ennast õigustaks.

Summary

Web Crawler Adaption According to the Needs of a Media Agency

The aim of this thesis was to find the best web-crawling based solutions for a media agency that would benefit their clients the most. The choice had to be based on functional properties as well as the actual needs of the clients. As a result, the decision was made, that the best way to approach this manner, would be to use a fee-based web crawling service with the most active clients who are want to use all the options to insure their brand's and campaign's success. For all the other clients, the agency should strongly recommend a one time, standard media monitoring service as a part of any campaign. The standard monitoring will be mainly based on a free Google Alerts tool plus a few others tools depending on the exact client, to ensure the best result. This way, we get to offer the best possible solutions for the active clients, but also introduce the different possibilities to more passive clients, that might spark an interest after the first results.

To get to this result, the paper introduces the main perks and functions of a media agency, plus the different types of clients, and their needs. Then the web crawlers are introduced, explaining exactly what they are, how do they work, how are they used and what are the main differences between them. Thirdly, the media agency and web crawler are put together, introducing ways how the media agencies use web crawlers for their benefit. In the last part, the suggestions are made, how to most effectively use crawler based services, to benefit different clients.

In the future, this thesis can be applied for media agency workers that are not yet fully aware of all the possible functions and benefits of web crawling, so they get a better understanding of all the possibilities and can also introduce these services to the agencies clients. This paper can also serve as a strong starting point for creating the agencies own crawler with all the needed functions, should the new strategy turn out to be successful enough, that it would be beneficial.

Kasutatud kirjandus

1. C, Castillo. (2014). *Effective Web Crawling, University of Chile*. Viimati vaadatud 03.05.2015, aadressil: http://chato.cl/papers/crawling_thesis/effective_web_crawling.pdf
2. V.Shkapenyuk & T.Suel (n.d). *Design and Implementation of a High-Performance Distribute Web Crawler**, Polytechnic University, Brooklyn, NY, 202. Viimati vaadatud 03.05.2015, aadressil: <http://cis.poly.edu/suel/papers/crawl.pdf>
3. B. Kahle. (1997). *Archiving the internet. Scientific American*. Viimati vaadatud 03.05.2015, aadressil: <http://www.uibk.ac.at/voeb/texte/kahle.html>
4. J. Cho and H. Garcia-Molina. (1999). *Synchronizing a database to improve freshness*. Viimati vaadatud 03.04.2015 aadressil: <http://oak.cs.ucla.edu/~cho/papers/cho-synch.pdf>
5. S. Chakrabarti, M. van den Berg, and B. Dom. (1999). *Focused crawling: A new approach to topic-specific web resource discovery*. Viimati vaadatud 03.05.2015, aadressil: <http://www.math.unipd.it/~dulli/corso04/www1999f.pdf>
6. Z. Bar-Yossef, A. Berg, S. Chien, J. Fakcharoenphol, and D.Weitz (2000). *Approximating aggregate queries about web pages via random walks*. Viimati vaadatud 03.05.2015, aadressil: <http://www.vldb.org/conf/2000/P535.pdf>
7. M. R. Henzinger, A. Heydon, M. Mitzenmacher, and M. Najork. (n.d). *Measuring index quality using random walks on the web*. Viimati vaadatud 03.05.2015, aadressil: <http://www.ra.ethz.ch/cdstore/www8/data/2142/pdf/pd1.pdf>
8. S. Raghavan and H. Garcia-Molina. (2001). *Crawling the hidden web*. In *Proc. of 27th Int. Conf. on Very Large Data Bases*. Viimati vaadatud 03.05.2015, aadressil: <http://ilpubs.stanford.edu:8090/725/1/2001-19.pdf>
9. Ricardo Baeza-Yates and Berthier Ribeiro-Neto. (1999). *Modern Information Retrieval*. ACM Press / Addison-Wesley. Viimati vaadatud 03.05.2015, aadressil: <http://people.ischool.berkeley.edu/~hearst/irbook/print/chap10.pdf>
10. Sergei Brin and Lawrence Page. (1998). *The anatomy of a large-scale hypertextual Web search engine*. *Computer Networks and ISDN Systems*. Viimati vaadatud 03.05.2015, aadressil: <http://snap.stanford.edu/class/cs224w-readings/Brin98Anatomy.pdf>
11. D. Eichmann. (1994). *The RBSE spider: balancing effective search against web load*. In *Proceedings of the first World Wide Web Conference, Geneva, Switzerland*. Viimati

- vaadatud 03.05.2015, aadressil:
http://www.researchgate.net/publication/200111196_The_RBSE_spider_balancing_effective_search_against_web_load
12. Brian Pinkerton. (1994). *Finding what people want: Experiences with the WebCrawler*. In Proceedings of the first World Wide Web Conference, Geneva, Switzerland. Viimati vaadatud 03.05.2015, aadressil:
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.22.890>
13. Smith, P.R. & Zook Z. (2011). *Marketing Communications: integrating offline and online with social media*. London: Kogan Page. Viimati vaadatud 03.05.2015, aadressil:
http://www.digitalmarketingassoc.com/wp-content/uploads/2012/06/MarComms_5th-ed-sample_chapter-7-MAY-1.pdf
14. Kuusik, A & K. Aarna & T. Mehine & I. Prinsthal & L. Sepp & M. Seppo & K. Virk (2010). Teadlik turundus. Tartu: Tartu Ülikooli Kirjastus.
15. Internetworldstats – Viimati vaadatud 03.05.2015, aadressil:
<http://www.internetworldstats.com/eu/ee.htm>
16. Marketingproff – Viimati vaadatud 03.05.2015, aadressil <http://marketingproff.eu/mis-seo-kokkuvotte/>
17. Searchenginewatch - Viimati vaadatud 03.05.2015, aadressil:
<http://searchenginewatch.com/sew/how-to/2396193/11-seo-tactics-you-need-to-know-in-2015>
18. Comscore – Viimati vaadatud 03.05.2015, aadressil:
<http://www.comscore.com/Insights/Blog/Major-Mobile-Milestones-in-May-Apps-Now-Drive-Half-of-All-Time-Spent-on-Digital>
19. Engadget – Viimati vaadatud 03.05.2015, aadressil:
<http://www.engadget.com/2015/04/21/google-mobile-search/>
20. Searchengineland – Viimati vaadatud 03.05.2015, aadressil:
<http://searchengineland.com/report-google-apple-safari-search-deal-expiring-yahoo-bing-want-209711>
21. Searchmarketingstandard – Viimati vaadatud 03.05.2015, aadressil:
<http://www.searchmarketingstandard.com/6-tips-to-improve-crawling-and-boost-seo>

22. Mediacom – Viimati vaadatud 03.05.2015, aadressil: <http://www.mediacom.com/en/what-we-do/our-industry/media-agencies/what-benefits-does-a-media-agency-deliver.aspx>
23. Promptcloud – Viimati vaadatud 03.05.2015, aadressil: <https://www.promptcloud.com/social-media-networking-sites-crawling-service/>
24. Wordtracker – Viimati vaadatud 03.05.2015, aadressil: <http://www.wordtracker.com/academy/learn-seo/getting-started/seo-mistakes>
25. Gliffy- Viimati vaadatud 03.05.2015, aadressil: <https://www.gliffy.com>
26. Thehistoryofseo – Viimati vaadatud 03.05.2015, aadressil: http://www.thehistoryofseo.com/TheIndustry/Short_History_of_Early_Search_Engines.aspx
27. Socialmention – Viimati vaadatud 03.05.2015, aadressil: <http://www.socialmention.com/>
28. Mediavantage – Viimati vaadatud 03.05.2015, aadressil: <http://www.mediavantage.com/monitor.html>
29. Lithium – Viimati vaadatud 03.05.2015, aadressil: <http://www.lithium.com/>
30. Exacttarget – Viimati vaadatud 03.05.2015, aadressil: <http://www.exacttarget.com/products/social-media-marketing/social-studio>
31. Neilpatel – Viimati vaadatud 03.05.2015, aadressil: <http://neilpatel.com/2015/02/24/the-definitive-guide-to-creating-high-converting-landing-pages/>
32. Shareacoke – Viimati vaadatud 03.05.2015, aadressil: <http://www.shareacoke.com/>
33. Postano – Viimati vaadatud 03.05.2015, aadressil: <http://www.postano.com/blog/14-best-social-media-campaigns-of-2014>
34. Okia – Viimati vaadatud 03.05.2015, aadressil: <http://www.okia.ee/blog/2010/02/sotsiaalmeedia-monitooring-ehk-mida-sinu-brandist-sotsiaalmeedias-raagitakse/>
35. Silverhage – Viimati vaadatud 03.05.2015, aadressil: <http://silverhage.com/google-reader-ninja/>
36. Facebook – Viimati vaadatud 03.05.2015, aadressil: <https://www.facebook.com/>
37. Twitter – Viimati vaadatud 03.05.2015, aadressil: <https://twitter.com/search-home>
38. Googleblog – Viimati vaadatud 03.05.2015, aadressil: <http://googleblog.blogspot.com.au/2013/03/a-second-spring-of-cleaning.html>