

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

DIPLOMSKO DELO

**ANALIZA PODATKOV,
ZAJETIH V SPLETNIH PROGRAMSKIH REŠITVAH**

Ljubljana, avgust 2008

ZORAN VIDMAR

IZJAVA

Študent Zoran Vidmar izjavljam, da sem avtor tega diplomskega dela, ki sem ga napisal pod mentorstvom prof. dr. Jurija Jakliča, in dovolim njegovo objavo na fakultetnih spletnih straneh.

V Ljubljani, dne 28.08.2008

Podpis_____

KAZALO

UVOD.....	1
1 ANALIZE SPLETNIH REŠITEV.....	2
1.1 Primeri uporabe analiz spletnih rešitev	4
1.1.1 Potrebe uporabnikov.....	5
1.1.2 Podpora trženju.....	5
1.1.3 Izboljševanje spletišča	6
1.1.4 Povečevanje prihodkov.....	6
1.2 Analitika podatkov poteka povezav	6
1.2.1 Analiza strežniških dnevniških datotek	8
1.2.2 Analiza dnevniških datotek spletnih programskih rešitev.....	9
2 VIDIKI ANALIZE PODATKOV SISTEMA ZA SPLETNO PRODAJO	10
2.1 Pravni vidik	11
2.2 Vsebinski vidik.....	13
2.2.1 Vidik spremljanja razvoja kupcev	14
2.2.2 Pomen merjenja, testiranja in učenja.....	16
2.3 Tehnični vidik.....	17
2.4 Celosten pogled na reševanje problema	17
3 NAČRTOVANJE USTREZNEGA MODELA ZA ANALIZE PODATKOV	19
3.1 Kratek pregled programskih rešitev spletne trgovine.....	20
3.2 Arhitektura sistema za analize podatkov spletnih rešitev.....	23
3.2.1 Opredelitev podatkov poslovnega procesa	24
3.2.2 Objava podatkov.....	25
3.2.3 Komponenta za interakcijo spletnih obiskovalcev	25
3.2.4 Izdelava podatkovnega skladišča	28
3.2.5 Komponenta za analize.....	28
3.2.6 Posredovanje rezultatov.....	28
4 PREGLED POSAMEZNIH REŠITEV	29
4.1 Sistem za zajem podatkov spletnih rešitev	29
4.2 Priprava poročil	31
4.3 Možnosti uporabe OLAP orodij	32
4.4 Možnosti uporabe podatkovnega rudarjenja.....	34
SKLEP	36
LITERATURA IN VIRI.....	39

KAZALO SLIK

Slika 1: Področja uporabe analiz spletnih rešitev	5
Slika 2: Pogled iz vidika razvoja posameznika - od spletnega obiskovalca do spletnega kupca.	15
Slika 3: Poenostavljena visoko-nivojska arhitektura sistema za integracijo analiz podatkov spletnih programskih rešitev	19
Slika 4: Primer arhitekture spletnih programskih rešitev	21
Slika 5: Podrobnejša visoko-nivojska arhitektura sistema za integracijo analiz podatkov spletnih programskih rešitev	24
Slika 6: ER diagram – zajem podatkov v komponenti za interakcijo	30

KAZALO PRILOG

Priloga 1: Pomeni nekaterih najpomembnejših izrazov, ki jih ZVOP-1 (2005) uporablja ...	1
Priloga 2: Nekateri najpomembnejši zakonski členi	2

UVOD

V zadnjih letih smo priča bliskovitemu razvoju spletnih tehnologij. Na začetku so le najpogumnejša podjetja posegala po prednostih in priložnostih, ki jih ponuja svetovni splet. Zaradi široke uveljavljenosti uporabe interneta, je v zadnjem času čedalje več podjetij, ki si želijo izboriti svoj tržni delež na spletu. Če vzamemo za primer neko spletno trgovino, lahko to storimo na več načinov, v osnovi pa lahko te aktivnosti delimo na pridobivanje novih strank ter aktivnosti za obdržanje obstoječih strank; kar je za podjetje načeloma tudi cenejša alternativa. V obeh primerih bo uspešnost in učinkovitost naših aktivnosti še najbolj odvisna od poglobljenega poznavanja strank (Berry & Linoff, 2001, str. 1-19).

Nekateri spletni trgovci za analiziranje svojega spletnega mesta še vedno uporabljajo zelo preproste kazalce, ki izhajajo iz merjenja obiskanosti na podlagi preučevanja strežniških dnevniških datotek. Te beležijo predvsem zahteve spletnih obiskovalcev po ogledu posameznih strani, kar pa še zdaleč ni dovolj za podrobnejši vpogled v navade, interesna področja in posamezne odločitve spletnih obiskovalcev, ki bi jih želeli izluščiti iz teh podatkov. Od tu izhajajo potrebe po zajetju podatkov v spletnih programskih rešitvah, ki nastajajo neposredno ob njihovi uporabi, oziroma nastajajo ob interakciji spletnih obiskovalcev (Ansari, Kohavi, Mason & Zheng, 2001, str. 4-5).

Namen našega diplomskega dela je celovit način predstavitve problematike analiz podatkov spletnih programskih rešitev. Poskušali bomo ugotoviti kako izkoristiti prednosti, ki nam jih ponujajo spletne tehnologije, na ustrezen način pristopiti k zagotavljanju dovolj kakovostnih, zanesljivih ter bogatih podatkov, ki bodo na najbolj celovit način odražali odločitve ali lastnosti strank ter njihovo uporabo spletnega sistema. Tako bomo preučili tudi primer arhitekture sistema, katerega osnovna ideja je celostni zajem in skupna raba podatkov o poslovnem procesu, uporabnikih in njihovi uporabi spletnih programskih rešitev. Podrobneje bomo preučili zajem podatkov, predstavili pa bomo tudi možnosti uporabe poročil in orodij OLAP in podatkovnega rudarjenja, za namene analiz podatkov spletnih programskih rešitev. V zadnjem poglavju bomo, poleg sklepnih misli in ugotovitev, omenili še možnosti ter napotke za nadaljnje delo in razvoj.

Pri našem preučevanju bomo uporabili različne metodološke pristope. V okviru splošne preučevalne metode bomo obravnavali problematiko analize podatkov spletnih programskih rešitev, in sicer z uporabo opisne metode, primerjalne metode ter metode kompilacije. Pri analizah podatkov spletnih programskih rešitev gre v bistvu za nekakšen presek različnih znanj, ki skupaj tvorijo dokaj specifično in ozko področje. Tako je tudi literature, ki bi neposredno definirala obravnavano problematiko, precej malo. Zato bomo v naši diplomski nalogi uporabili večinoma tujo strokovno literaturo ter različne internetne vire. Za obravnavanje posameznih problemskih področij se bomo opirali predvsem na dela avtorjev, ki se ukvarjajo s posameznimi tematikami, ter jih poskušali, s pomočjo sinteze združiti v smiselno celoto. Poleg celostne predstavitve značilnosti, posebnosti ter problemov analiz podatkov spletnih programskih rešitev, bomo obravnavali tudi možnosti njihovega reševanja. Za razlage iz vsebinskega vidika se bomo sklicevali predvsem na avtorja Berry in Linoff (2001), za podrobnejšo razlago primera možne arhitekture, sistema za analize podatkov spletnih rešitev, pa na avtorje Ansari, Kohavi, Mason in Zheng (2001). Primer spletne trgovine, ki ga bomo uporabili za lažje razumevanje, izhaja iz znanj pridobljenih med študijem ter praktičnih izkušenj s področja razvoja in izdelave spletnih rešitev.

1 ANALIZE SPLETNIH REŠITEV

Ena poglavitnih nalog informatike je zniževanje stroškov ter skrajševanje časa potrebnega za izvajanje poslovnih procesov oziroma aktivnosti znotraj le-teh. Na strateškem področju podjetju informatika predstavlja orodje za ugotavljanje poslovnih priložnosti, lahko pa jo uporabi tudi kot orožje za zagotavljanje svojih konkurenčnih prednosti (Kovačič, 1998, str. 47-48).

Uporaba spletnih tehnologij še vedno narašča, tako po številu uporabnikov kot tudi po dodani vrednosti, ki jo le-te ustvarjajo. Prav tako smo priča hitremu razvoju tehnologij, s katerimi spletnim obisovalcem ponujamo svoje storitve, povečujeta se tako obseg vsebin kot tudi sama kompleksnost spletnih programskih rešitev.

Vsako trgovsko podjetje mora, še pred pričetkom načrtovanja trženjskih aktivnosti, spoznati ciljne porabnike in njihovo nakupno vedenje, da bi tako oblikovalo najboljšo ponudbo izdelkov (Potočnik, 2001, str. 119).

Pomen analiz uporabe spletnih rešitev se povečuje predvsem zaradi različnih **poslovno-analitičnih potreb**, oziroma natančnejše, potreb **poslovne inteligence** (angl. *Business Intelligence*). S poslovno inteligenco (poslovnim obveščanjem) tako označujemo sistem, ki omogoča analizo podatkov o poslovanju organizacije in posledicah sprejetih odločitev ([Islovar] Poslovno obveščanje, 2008b). Razvoj poslovne inteligence je tesno povezan z **informacijsko demokratizacijo**, kar omogoča možnost dostopa do podatkov in njihovo analizo čedalje večjemu številu poslovnih uporabnikov (Jaklič, 2002, str. 178).

Poslovna inteligenca je z vidika spletnih rešitev zelo aktualna tema. Na področju spletnih rešitev se srečujemo s tako imenovano **spletno inteligenco** (angl. *Web Intelligence*), katero razumemo kot kombinacijo uporabe **spletne analitike**, ki preučuje podatke o obnašanju in interakciji naših spletnih obiskovalcev, ter **poslovne inteligence**, s katero skušamo na podlagi zbranih podatkov, z uporabo odločitvenih sistemov, podpreti ali izboljšati določene odločitve poslovnega procesa (Web Intelligence [University California, Irvine Extension], 2008).

Z vidika trženja velikokrat govorimo o **managementu odnosov z odjemalci** (angl. *customer relationship management*, krat. CRM). Sam koncept temelji na ideji: bolj kot poznamo svoje stranke, boljše lahko z njimi vzdržujemo ali izboljšujemo poslovni odnos. Cilj je nekakšna dolgoročna maksimizacija odnosov s kupci. Pri tem se predvsem osredotočamo na obdržanje donosnejših kupcev in na to, kako stranke, ki nam ne prinašajo zaslužkov, spremeniti v donosne kupce (Kimball & Ross, 2002, str. 142-143).

Glede na obravnavo problemskega področja, v okviru upravljanja odnosov s kupci, govorimo o (Jaklič, 2002, str. 185-186):

- **analitičnem CRM**, kjer se organizacija ukvarja s postopki za doseganje temeljitega vpogleda v želje in potrebe strank, razumevanje njihovega vedenja ter predvidevanje njihovih namer,

- **operativnem CRM**, ki se nanaša na izvajanje procesov povezovanja organizacije s strankami, za katere je značilna tudi prenova in informatizacija,
- **organizacijskem CRM**, ki se nanaša predvsem na komuniciranje med organizacijo in strankami, z uporabo različnih oblik obvečanja in komunikacije.

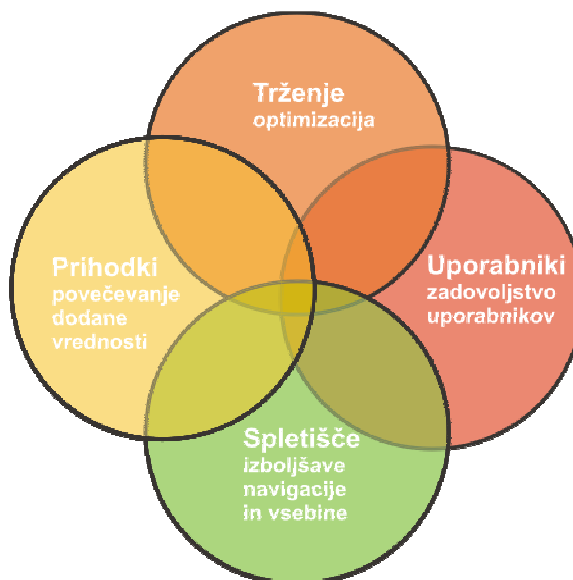
Kot lastnike spletne trgovine nas tako še posebej zanima kdo so naši obiskovalci, kaj jih najbolj interesira in kako se jim čimbolj približati ter tako povečati spletni obisk oziroma spletno prodajo izdelkov. Najpogosteje uporabljena orodja za spremljanje obiska spletnega mesta so tako imenovane statistike obiska spletnih strani. V praksi se nam dogaja, da nam podatki, ki jih imamo na razpolago, ne zagotavljajo ključnih oziroma vsaj zadostnih informacij, ki bi jih v določenem trenutku potrebovali za rešitev nekega kompleksnejšega vprašanja.

V nadaljevanju si bomo na kratko ogledali področja uporabe analiz spletnih rešitev in poskušali ugotoviti načine pristopa k razvoju sistema, ki nam bi omogočil celostni zajem in analizo podatkov o uporabnikih in njihovi uporabi naših spletnih programskih rešitev.

1.1 Primeri uporabe analiz spletnih rešitev

Analitika spletnih rešitev temelji na raziskovanju, preučevanju in statistični obdelavi podatkov o uporabnikih in njihovi uporabi spletnega mesta. Analize vzorcev uporabe spletnih rešitev postajajo ključnega pomena za obstoj in uspeh vsakega sistema za spletno poslovanje. Z ustreznimi tehnikami in izbiro primernih kazalcev, lahko merimo, spremljamo in preučujemo obnašanje spletnih obiskovalcev ter tako ugotavljamo trende uporabe naših spletnih rešitev. S tem načinom si zagotovimo dobro podporno okolje tudi z vidika ocenjevanja uspešnosti posameznih aktivnosti, ki jih izvajamo v okviru tekočega poslovnega procesa.

Slika 1: Področja uporabe analiz spletnih rešitev



Vir: Web Analytics [Softlutions], 2008

Različne poslovne aktivnosti, ki jih moramo kontinuirano izvajati za optimalno delovanje našega spletnega sistema, nam iz različnih vidikov narekujejo potrebe po analizah spletnih rešitev. Z ustreznimi analizami podatkov o obisku in uporabi spletnega mesta, lahko poslovni proces spletne prodaje podpremo na številne načine.

1.1.1 Potrebe uporabnikov

Kot smo že omenili so osrednja tema preučevanja spletni uporabniki. S preučevanjem uporabe spletnih rešitev lahko poskušamo na različnih nivojih ugotavljati in primerjati vzorce obnašanja, interesna področja in nakupne navade uporabnikov. Ti podatki lahko postanejo zelo uporabni, še zlasti če izhajamo iz vidika identificiranja in zadovoljevanja potreb ter želja spletnih obiskovalcev. S poosebitvami spletnega mesta želimo olajšati uporabo in na nek način povečati zadovoljstvo uporabnikov ter s tem doseči večji spletni obisk in posredno večjo prodajo (Berry & Linoff, 2001, str. 34).

1.1.2 Podpora trženju

Tako podjetja, ki se ukvarjajo s spletnim poslovanjem, uporabljajo raznovrstne marketinške prijeme, s katerimi poskušajo pritegniti čimveč novih spletnih obiskovalcev ali ponovni obisk že obstoječih strank. Za spletno oglaševanje lahko tako porabimo precej finančnih sredstev, ki pa jih moramo nekako upravičiti. Ponavadi to storimo s

spremljanjem izbranih kazalcev odziva na posamezno trženjsko akcijo, s katerim želimo izboljšati učinkovitost pozicioniranja oglasnih pasic. Z ustreznimi analizami podatkov, o obisku in uporabi spletnega mesta, lahko trženjski vidik spletne prodaje podpremo še na številne druge načine (Berry & Linoff, 2001, 10-14).

1.1.3 Izboljševanje spletišča

Vsebina spletnih strani se praviloma na določen časovni rok spreminja in prav tako želje in potrebe naših uporabnikov. Te spremembe lahko sčasoma povzročijo nepreglednost in neurejenost spletnega mesta. Prav tako se včasih pojavijo napake v obliki pretrganih povezav ali manjkajočih strani. Zato moramo skrbeti za kontinuiran razvoj in optimizacijo spletnega mesta. Na podlagi analiz podatkov o obisku in uporabi spletnega mesta, lahko napake in potrebe po izboljšavah hitreje prepoznamo in pravočasno poskrbimo za implementacijo.

1.1.4 Povečevanje prihodkov

V kolikor govorimo o sistemih spletne prodaje, je zagotovo naš osnovni cilj generiranje prihodkov oziroma še boljše, dodane vrednosti iz naslova spletnega poslovanja. Tudi vsa že ostala področja se konec koncev odražajo v uspešnosti poslovanja spletnega mesta ali morda celo podjetja kot celote.

V nadaljevanju diplomskega dela bomo videli, da multidisciplinarnost analiz spletnih rešitev izhaja tako iz poslovnega kot tudi iz tehnološkega poznavanja problema.

1.2 Analitika podatkov poteka povezav

Potek povezav (angl. *clickstream*) lahko definiramo kot zapis podatkov o straneh, ki si jih je posamezni spletni obiskovalec med obiskom ogledoval, ter o zaporedju ogledov.

Podatki o obisku spletne strani (angl. *clickstream data*) so lahko shranjeni v različnih dnevniških datotekah, ki jih uporabniki spletnih strani ustvarjajo sami, z navigiranjem in klikanjem elementov spletnih strani ([Islovar] Clickstream, 2008a).

Analiza podatkov o poteku povezav (angl. *clickstream data analysis*), ki jo včasih imenujemo kar **analitika podatkov poteka povezav** (angl. *clickstream analytics*), je proces zbiranja, analiziranja in poročanja agregatnih podatkov o obisku spletne strani. Analiza podatkov poteka povezav se izkaže za najbolj učinkovito v uporabi z ostalimi, bolj tradicionalnimi, tržnimi pristopi, ki jih ponavadi uporabljamo za evaluacijo rezultatov posameznih aktivnosti poslovnega procesa ([SearchCRM.com] Clickstream analysis, 2008).

Iz osnovnih podatkov o straneh, ki so si jih posamezni obiskovalci ogledali, lahko zelo hitro ugotovimo, kdaj in kje so vstopili na naše spletišče, koliko časa so si posamezne strani ogledovali in v katerem trenutku, oziroma kje, so naše spletišče zapustili. Z ustrezno agregacijo podatkov lahko dokaj preprosto ugotovimo tudi najpogosteje obiskane strani, koliko časa traja v povprečju en ogled ali posamezni spletni obisk itd.

Analizo poteka povezav ponavadi uporabljamo za ugotavljanje vzorcev uporabe našega spletišča, s tem poskušamo dobiti tudi nekakšen vpogled v preference, obnašanje in interesna področja naših obiskovalcev. To analizo lahko uporabimo v pomoč pri reševanju večine problemov, ki smo jih navedli že v področjih uporabe. Rezultate analize poteka povezav lahko uporabimo kot del podsistema za avtomatizirano priporočanje spletnih vsebin, ter tako neposredno poskrbimo za večjo uporabnost naših spletnih rešitev, posredno pa za zadovoljstvo naših spletnih uporabnikov.

Iz vidika zajemanja podatkov, oziroma katere podake analiziramo, lahko ločimo dva nivoja (stopnji) **analiz o poteku povezav**:

- **analize spletnega prometa** (angl. *traffic analysis*), ki temeljijo na preučevanju podatkov v strežniških dnevniških datotekah
- in **analize uporabe spletnih rešitev** (angl. *e-commerce analysis*), ki temeljijo na preučevanju podatkov dnevniških datotek programskih rešitev, oziroma na nivoju programskega strežnika (angl. *application server*).

Tovrstna analiza se pogosto prične s preučevanjem strežniških dnevniških datotek, kar ima svoje prednosti in slabosti. Kot bomo videli v nadaljevanju, ni smiselno zapostaviti informacij, ki se skrivajo v strežniških dnevniških datotekah.

1.2.1 Analiza strežniških dnevniških datotek

Strežniška dnevniška datoteka je datoteka (ali več datotek), v katero spletni strežnik beleži vse zahteve po ogledu strani (angl. *page requests*), in sicer v obliki IP naslova uporabnika, datuma in ure zahtevka, zahtevane strani v obliki http naslova, količine prenesenih podatkov, napotitelja uporabnika in tipa spletnega brskalnika, ki ga le-ta uporablja.

Prednost dnevniških datotek spletnega strežnika je, da podatke za analize pridobivamo na dokaj preprost in za spletne uporabnike še najmanj obremenjujoč način in tako še najmanj posegamo v njihovo »informacijsko zasebnost«. Po drugi strani se v teh datotekah ne beležijo visoko-nivojski dogodki in metapodatki, ki so ključnega pomena za naše analize. Posebej težko si pomagamo z analizo strežniških dnevniških datotek tudi v primerih, ko imamo opravka z dinamičnimi spletnimi vsebinami.

Glavna pomanjkljivost strežniških dnevniških datotek se kaže prav v tem, da moramo vložiti ogromno truda preden lahko iz njih dobimo primerne podatke za našo analizo. Te datoteke vsebujejo številne, za našo analizo včasih nepomembne podatke. Ob ogledu ene strani se evidentirajo tudi vsi zahtevki po ogledu vseh slik, ki se avtomatsko naložijo v brskalnik, kot sestavni del te posamezne spletne strani. Zato moramo podatke najprej **prefiltrirati, izločiti vplive spletnih pajkov** (uporabljajo jih razni spletni iskalniki za osveževanje metapodatkov), **poidentificirati uporabnike, razvrstiti podatke** o ogledih strani na posamezne spletne seje in nekako ugotoviti **kdaj so se posamezni spletni obiski zaključili** (Berry & Linoff, 2001, str. 35-41).

Za analizo dnevniških datotek ponavadi uporabljamo programske rešitve, ki z uporabo statističnih metod analizirajo podatke. Ta orodja nam tako z dokaj preprostimi pa tudi kompleksnimi uporabniškimi vmesniki, nudijo podatke o uporabi našega spletnega mesta. Statistike obiska nam omogočajo analizo obiska spletnih strani obogateno z grafičnimi pregledi, urejeno po posameznih indikatorjih in posameznih časovnih obdobjih (urah, dnevih, mesecih itd.). Nekateri najpogostejši in **najpomembnejši kazalci**, ki jih statistike obiska ponavadi prikazujejo, so ([Wikipedia] Web log analysis software, 2008):

- število obiskov in število posameznih oz. enoličnih obiskovalcev v izbranem časovnem obdobju,

- strani na katerih so obiskovalci največkrat vstopili na našo spletno stran in strani kjer so jo največkrat zapustili,
- trajanje obiskov in zadnji obiski,
- prometne konice po dnevih v tednu, mesecu ali letu,
- domene oz. države gostiteljev od koder prihajajo naši obiskovalci,
- datoteke, ki so si jih obiskovalci ogledovali,
- informacije o uporabljenem operacijskem sistemu obiskovalcev oz. katere spletne brskalnike uporabljajo (angl. *user agent*),
- informacije o spletnih robotkih,
- kdo so najpogostejše naši HTTP napotitelji oz. od kje prihajajo naši obiskovalci,
- iskalniki in ključne besede, ki so jih uporabili pri iskanju naše spletne strani,
- HTTP napake.

V strežniških dnevniških datotekah lahko najdemo dragocene informacije o obiskanosti našega spletnega mesta, ne nudijo pa nam podrobnejših podatkov o uporabnikih, uporabi spletnih programskih rešitev in ključnih poslovnih dogodkov.

1.2.2 Analiza dnevniških datotek spletnih programskih rešitev

Za razliko od analize strežniških dnevniških datotek, kjer se moramo vedno ukvarjati s filtracijo in čiščenjem podatkov, ugotavljanjem spletnih sej in identifikacijo posameznih uporabnikov, je prednost **analize dnevniških datotek spletnih programskih rešitev** (angl. *application log analysis*) v tem, da podatke o uporabnikih ter njihovi uporabi našega spletišča zajamemo že na nivoju posameznih spletnih programskih rešitev. Da bi prišli do podatkov, ki bi nam omogočili poglobljeno razumevanje spletnih obiskovalcev, moramo v naš sistem integrirati tudi podatke o obisku strani (angl. *clickstream data*). Tako dokaj preprosto dobimo podatke o posameznih spletnih sejah, uporabnikih in celo o njihovi uporabi posameznih rešitev našega spletnega sistema (npr. prijava uporabnika v sistem, dodajanje artikla v košarico, potrditev nakupa).

Sodobne spletne programske rešitve temeljijo predvsem na uporabi dinamičnih spletnih vsebin. Prednost analize dnevniških datotek spletnih programskih rešitev je tudi v tem, da v njih lahko zajamemo podrobnejše podatke o uporabi dinamičnih spletnih vsebin. V strežniški dnevniški datoteki se beležijo zahtevki po ogledih strani v obliki spletnega

naslova, kar pa nam, v primeru dinamičnih spletnih strani, ne pove kaj dosti, saj se lahko zgodi, da se na posameznem naslovu, v različnih časovnih obdobjih, prikazujejo različne spletne vsebine (npr. različni izdelki, različne skupine). V dnevniških datotekah spletnih programskih rešitev tako lahko beležimo podatke o posameznih zaslonskih prikazih ali, drugače povedano, o dogodkih in procesih, ki se na osnovi interakcije obiskovalcev izvajajo na nivoju programskega strežnika (Linoff & Berry, 2001, 41-42).

Za lažje razumevanje si oglejmo še malce poenostavljen primer. Spletno podjetje izvede oglaševalsko akcijo s pozicioniranjem spletnega oglasa v obliki spletne pasice na dveh spletnih portalih. Na podlagi dnevniških datotek spletnega strežnika primerjamo podatke o obisku, ki so ga bili deležni iz naslova spletnega oglaševanja, in ugotovimo, da podatki o številu napotenih obiskovalcev govorijo v prid prvega spletnega portala. Kljub temu pa še vedno ne vemo, koliko od teh obiskovalcev je postopek nakupa dejansko dokončalo in koliko so ti obiskovalci dejansko zapravili. Te podatke lahko dobimo le z analizo dnevniških datotek programskega strežnika. Tako lahko dobimo čisto drugačno, včasih malce paradoksalno, sliko iste situacije. Lahko se nam izkaže, da je večja učinkovitost oglaševalske akcije na drugem spletnem portalu, ker so ti obiskovalci opravili večji obseg nakupov (Linoff & Berry, 2001, str. 35-43).

Ta primer nam jasno kaže, da nam brez celostnega (tako vsebinskega kot tudi tehničnega) razumevanja ni kaj početi. Kot smo že omenili, moramo v naš sistem najprej integrirati ustrezen sistem, ki bo nudil dovolj kakovosten in širok zajem podatkov. V nadaljevanju si bomo, za lažje razumevanje, najprej ogledali primer spletne trgovine, omejitve in značilnosti, ki jih moramo upoštevati. Potem pa si bomo na tem primeru pomagati s primerom arhitekture, s katero bomo poskušali reševati probleme in izzive analiz podatkov spletnih rešitev.

2 VIDIKI ANALIZE PODATKOV SISTEMA ZA SPLETNO PRODAJO

Če se osredotočimo zgolj na analizo uporabe spletnih programskih rešitev, kaj hitro ugotovimo, da gre za kompleksen in multidisciplinaren problem. V nadaljevanju se bomo, za lažjo predstavitev in reševanje obravnavane problematike, oprli na vzorčni primer.

Predstavljajmo si podjetje, ki prek svojih spletnih strani, s pomočjo spletne trgovine, prodaja računalniško opremo.

Za panogo prodaje računalniške opreme lahko na kratko rečemo, da se nahaja v fazi zrelosti. Spletne trgovine, ki prodajajo večinoma standardizirane produkte, se soočajo s številnimi konkurenti, ki po konkurenčnih cenah ponujajo istovrstne, če ne celo enake, produkte istih blagovnih znamk. Zaenkrat pustimo ob strani cenovno konkurenčnost prodajnega programa, varnost, kakovost in zanesljivost storitev ter ugled in zaupanje, ki ga podjetje uživa v očeh kupcev, kar so tudi bistvene komponente pri odločanju posameznika za nakup preko spleta. Osredotočili se bomo na učinkovitost posameznega spletnega mesta in doseganje konkurenčnih prednosti s pomočjo analiz podatkov zajetih v spletnih programskih rešitvah.

Na namišljenem primeru bomo predvideli uporabo nekaterih klasičnih programskih rešitev. Na podlagi vzorčnega sistema spletne prodaje bomo skušali opisati in nakazati osnovne značilnosti, pristope in omejitve pri reševanju problematike analiz podatkov spletnih rešitev. Ker pa so kupci naše spletne trgovine tako pravne kot tudi fizične osebe, moramo zaradi občutljivosti tematike najprej omeniti pogled iz vidika varstva osebnih podatkov posameznika.

2.1 Pravni vidik

Poznavanje zakonske osnove je pomembno dejstvo, predvsem zaradi omejitev in pravil, ki jih moramo upoštevati tako pri načrtovanju, razvoju in uporabi programskih rešitev. V našem primeru, ko se ukvarjamo z zajemom in analizo podatkov spletnih programskih rešitev, se te omejitve nanašajo predvsem na pravice posameznika.

Pravica do varstva osebnih podatkov je temeljna človekova pravica. Varuje jo tudi slovenska ustava in sistemski zakon, Zakon o varstvu podakov, ki je v veljavi od 1.1.2005. Iz 38. člena slovenske ustave izhaja, da je na področju obdelave osebnih podatkov prepovedano vse kar ni z zakonom izrecno dovoljeno. To pomeni, da lahko obdelujemo samo tiste osebne podatke, ki so z zakonom opredeljeni, jasno pa morata biti razvidna tudi namen obdelave in ustrezno varovanje tajnosti podatkov.

V nadaljevanju si bomo na kratko ogledali zakonsko podlago (Bien, Bogataj, Pirc, Prelesnik & Žaucer, 2006), najpomembnejše izraze (za podrobnosti glej Prilogo 1) in glavne zakonske člene (za podrobnosti glej Prilogo 2), ki so relevantni za razumevanje in ustrezno reševanje našega primera.

Zakon o varstvu osebnih podatkov (v nadaljevanju ZVOP-1) je krovni zakon, ki natančneje določa podatkovno zasebnost, ter jo podrobneje opredeljuje tudi z vidika avtomatizirane obdelave s sredstvi informacijske tehnologije. Zakon se nanaša na osebne podatke posameznikov oz. fizičnih oseb, kar v našem primeru pomeni, da se nanaša predvsem na določen del naših spletnih uporabnikov. Na kratko bomo poskusili povzeti glavna določila ZVOP-1 (2005), ki se nanašajo na obdelavo osebnih podatkov našega vzorčnega primera.

V zasebnem sektorju se lahko obdelujejo osebni podatki, ki jih določa zakon, ali če je za to podana osebna privolitev. Za področje neposrednega trženja, natančneje za namene ponujanja blaga, kamor sodi tudi primer naše spletne prodajalne, zakon dovoljuje uporabo osebnih podatkov, ki smo jih zbrali iz javno dostopnih virov ali v okviru opravljanja dejavnosti. Po zakonu lahko uporabljamo naslednje osebne podatke: ime in priimek, naslov prebivališča, telefonska številka, številka telefaksa in naslov elektronske pošte. Na podlagi osebne privolitve posameznika lahko upravljalec obdeluje tudi druge osebne podatke. Pravni vidik se tako še najbolj odraža v modulu za registracijo uporabnikov, ki ga bomo omenili v nadaljevanju. Spletni obiskovalci so kot uporabniki sistema za nas anonimni, dokler ne sprejmejo splošnih pogojev poslovanja in uporabe osebnih podatkov, opravijo prvo registracijo in opravijo prijavo v naš spletni sistem.

ZVOP-1 (2005) tudi natančneje določa varstvo osebnih podatkov, vodenje, povezovanje in iznos posameznih zbirk, pravice in obveščanje uporabnikov, poročanje ter ostale dolžnosti in aktivnosti v poslovno-organizacijskem in tehničnem smislu, ki pa se jim, zaradi obsežnosti, v diplomskem delu ne bomo podrobneje posvečali.

Za pravno podlago našega primera bi bila lahko relevantna še dva posebna zakona, in sicer **Zakon o varstvu potrošnikov** in **Zakon o elektronskih komunikacijah**, ki možnosti neposrednega trženja s pomočjo elektronskih komunikacij še dodatno omejita. Zakon o

elektronskih komunikacijah ureja predvsem vidike elektronske komunikacije, vse ostale oblike pa so podrejene ureditvi določeni v ZVOP-1 (2005).

2.2 Vsebinski vidik

Zasnova vsakega spletnega mesta praviloma izhaja iz strateških usmeritev in ciljev, ki jih podjetja s svojo dejavnostjo na spletu želijo doseči. Celotna zasnova spletne trgovine je podrejena osnovnemu namenu - prodaji izdelkov končnim kupcem. Tako z vsebinsko, menujsko in grafično strukturo se želimo približati našim obiskovalcem.

Vsebinska zasnova spletnega mesta temelji na uporabi tehnologij dinamičnih spletnih vsebin, ki preko uporabniškega vmesnika odražajo že smiselno urejene strukture vsebin, ki jih ponujamo našim spletnim obiskovalcem. Spletne vsebine se načeloma konstantno spreminjajo, saj moramo kot upravljalci spletne trgovine skrbeti za aktualnost novic, prodajnega programa in ostalih novosti, ki se nanašajo na ažurnost spletnih vsebin. Tako iz spletnih programskih rešitev izhaja potreba po beleženju sprememb - v smislu hierarhičnega in časovnega okvira posameznih spletnih vsebin, ki si jih naši spletni obiskovalci ogledujejo.

Iz vidika analiz tako izhaja potreba po beleženju podatkov oziroma metapodatkov o posameznih spletnih vsebinah, s katerimi skušamo zajeti vsebinsko plat spletnega mesta na način, ki ga razumejo naši spletni obiskovalci. Tako obogateni podatki nam po eni strani omogočajo lažje razumevanje rezultatov analiz, po drugi strani pa nam v kombinaciji s podatki o posameznih izbirah uporabnikov odpirajo čisto nove možnosti uporabe.

Če gledamo iz vsebinskega vidika, lahko prikaze posameznih spletnih vsebin razdelimo tudi na **navigacijske strani**, ki so namenjene hitremu pregledu podatkov različnih spletnih vsebin in na **ciljne strani**, kjer so posamezne spletne vsebine predstavljene bolj podrobno. Ob analiziranju tako imenovane lepljivosti (angl. *stickiness*) oziroma podatkov o času, ki ga posamezni obiskovalec posveti ogledu posameznega zaslonskega prikaza, pridemo do zanimivih ugotovitev, ki pa jih je smiselno interpretirati iz vsebinskega vidika. Če obiskovalci porabijo relativno veliko časa pri podrobnih ogledih izdelkov, lahko sklepamo, da jih te vsebine zanimajo in obratno. Če pa obiskovalci porabijo relativno veliko časa na

navigacijskih straneh, lahko očitno sklepamo, da te navigacijske strani najbrž niso dovolj pregledno urejene (Berry & Linoff, 2001, str. 32-34).

2.2.1 Vidik spremljanja razvoja kupcev

Pri večini nakupnih odočitev lahko na različne načine sodeluje več udeležencev, ki imajo pri tem različne vloge (npr. pobudnik, vplivnež, odločevalec, kupec in porabnik). Pomembnost in število nakupnih udeležencev naraščata sorazmerno z zapletenostjo nakupne situacije. Za zapletene nakupe so značilne naslednje stopnje: opredelitev potrebe, iskanje informacij, primerjava različnih možnosti, nakup in ponakupno vedenje (Potočnik, 2001, str.119).

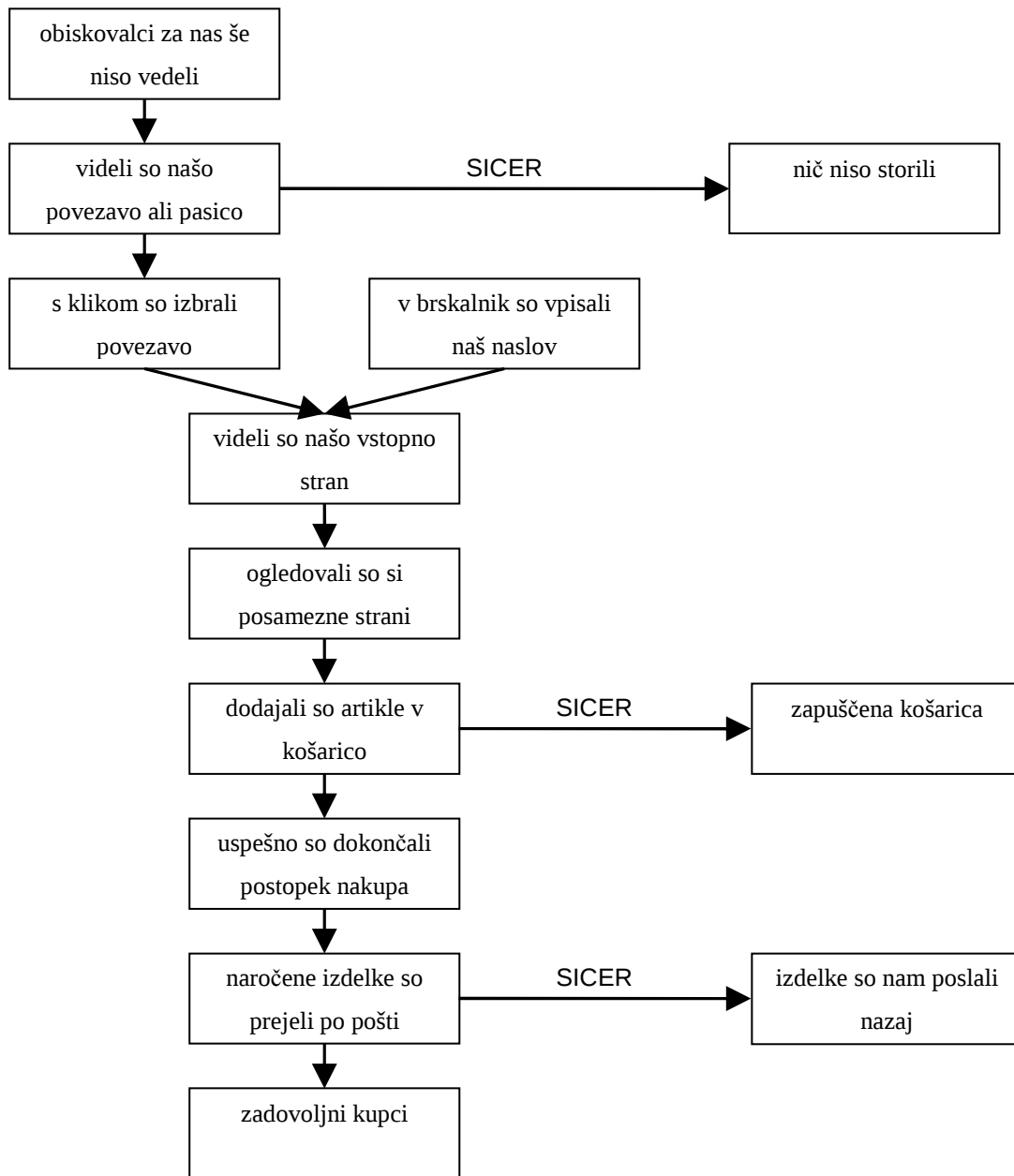
Ker so naši obiskovalci in njihova uporaba spletnih rešitev osrednja tema našega preučevanja, omenimo še pogled z vidika razvoja posameznika od spletnega obiskovalca do spletnega kupca (Slika 2, str. 15).

Razlikovanje med različnimi segmenti uporabnikov je pomembno predvsem iz vidika upravljanja odnosov s strankami, katerega pogost cilj je pridobiti čimveč rednih, ponovnih kupcev. V te namene se ukvarjajo tudi s privabljanjem novih spletnih obiskovalcev, zagotavljanjem kupcem, da iskane izdelke lažje najdejo in izvajanjem aktivnostimi, s katerimi želijo pospešiti sam nakup (Berry & Linoff, 2001, str. 71-73).

Za spletne obiskovalce lahko rečemo, da jih imenujemo kupce šele takrat, ko so preko našega sistema opravili nakup. To bi bila zelo preprosta delitev. Če pa si postopek posameznega nakupa podrobneje ogledamo, ugotovimo, da lahko uporabnike še podrobneje razdelimo na (Berry & Linoff, 2001, str.73):

- tiste, ki nas niso nikoli obiskali ali videli oglasa,
- tiste, ki so videli naš oglas, ampak nas niso obiskali,
- tiste, ki so obiskali našo vstopno stran, ampak z ogledom niso nadaljevali,
- tiste, ki so pregledovali izdelke, ampak jih niso dodajali v košarico,
- tiste, ki so opustili nakup vsebine nakupovalne košarice,
- tiste kupce, ki so opravili en nakup,
- tiste kupce, ki so opravili več zaporednih nakupov.

Slika 2: Pogled iz vidika razvoja posameznika - od spletnega obiskovalca do spletnega kupca.



Vir: M., J., A., Berry & G., S., Linoff, *Mining the Web: Transforming Customer Data into Customer Value*, 2001, str.72.

Če se še osredotočimo na anatomijo nakupa je za nas, kot upravljalca spletnega mesta, zelo pomemben podatek o artiklih, ki so jih obiskovalci dodali v košarico in katere od teh so dejansko kupili. Zelo zanimivo je v katerem koraku so kupci prekinili postopek nakupa in pa katere artikle so imeli v **vsebini zapuščene košarice**. Pri tem nas lahko tovrstni podatki tudi zavedejo, saj spletni obiskovalci velikokrat dodajajo različne izdelke v nakupovalno košarico z namenom, da bi jih nato primerjali med seboj in na koncu izbrali najbolj

primerne. Zato je prepoznavanje posameznih odločitev posameznikov in razumevanje njihovih vzrokov zelo pomembno, tako za potrebe trženja kot tudi za potrebe izboljšav spletišča.

Omenimo še nekaj uporabnih kazalcev, ki so pomembni predvsem iz vidika analiz spletnih nakupov (Berry & Linoff, 2001, str. 75):

- pogostnost nakupov,
- povprečna vrednost nakupovalne košarice,
- skupno število različnih kupljenih izdelkov,
- skupno število kupljenih izdelkov različnih kategorij,
- dan v tednu in čas nakupa,
- odziv na priporočila in posebne ponudbe,
- ter primerjava spletnih nakupov in ostalih nakupov v primeru klasične prodaje.

2.2.2 Pomen merjenja, testiranja in učenja

Glavni cilj analiziranja podatkov spletnih rešitev je, da z njimi poskušamo podpreti ali nekako izboljšati poslovni proces. Zelo pomemben faktor za uspešno analizo podatkov je pripravljenost za učenje in prilagajanje. Spoznanja, ki jih ugotovimo pri analizi podatkov, nas lahko vodijo do novih in nepričakovanih odkritij, ki nas lahko zanesejo v preučevanje čisto novih razsežnosti. Dognanja in rezultate, ki jih dobimo na podlagi analiz podatkov, lahko preverimo in podkrepimo tudi z nadaljnjo izvedbo različnih trženjskih akcij, izboljšavami spletnega mesta in podobno (Berry & Linoff, 2001, str. 316-317).

Po znanstveni metodi za povečevanje učenja, moramo k reševanju pblemov pristopiti z uravnoteženo mešanico kreativnosti, skeptičnosti in samozavesti. Kreativnosti pri postavljanju novih hipotez, skeptičnosti pri njihovem potrjevanju, zavračanju ali razvijanju alternativ in samozavesti, da na podlagi realnih rezultatov preučevanje nadaljujemo (Berry & Linoff, 2001, str. 281).

Informacije iščemo tako v absolutnih vrednostih podatkov ali pa te informacije skušamo izluščiti iz relativnih vrednosti oziroma njihovih primerjav. Za razumevanje, interpretacijo in primerjave podatkov je bistvenega pomena tudi zgodovinska komponenta oziroma čas, kar je tudi ena od glavnih značilnosti podatkovnih skladišč.

Za konec omenimo še učenje, ki je praviloma kontinuiran proces. Za učenje lahko rečemo, da temelji na nekakšnem zgodovinskem spominu. Pomnjenje podatkov ni zadosten pogoj za učenje, je pa lahko dobra osnova za analize. Za ustvarjanje optimalnega okolja za učenje moramo s podatki zadostiti določenim zahtevam, tako iz vidika razpoložljivosti kot tudi struktur podatkov (Berry & Linoff, 2001, str. 337).

2.3 Tehnični vidik

Spletne trgovine so ponavadi zgrajene modularno in sestojijo iz različnih, namenskih rešitev, ki navzven tvorijo celovit sistem. Posamezne programske rešitve so lahko povezane ali pa ločene in medsebojno neodvisne. Modularna zasnova sistema, izhaja neposredno iz samega koncepta spletne prodaje oziroma dodatnih storitev, ki jih uporabnikom omogočamo z uporabo raznovrstnih spletnih programskih rešitev. Tako lahko zelo hitro ugotovimo, da imamo opravka z različnimi podatki, ki se lahko hranijo v različnih podatkovnih virih. Pogosto se srečujemo z mešanjem tehnologij oziroma uporabo različnih tehnoloških platform in pristopov.

Ker naše spletišče temelji na uporabi dinamičnih spletnih vsebin, si samo s pomočjo analize strežniških dnevniških datotek (pri preučevanju vedenja obiskovalcev ter ugotavljanju njihovih vzorcev uporabe spletnih programskih rešitev) nimamo kaj dosti pomagati. Eden glavnih problemov ostaja identifikacija posameznih spletnih obiskovalcev, ki jih moramo za potrebe analiz povezati s posameznimi ogledi strani, nakupi izdelkov in podobno.

Spletne kupce lahko identificiramo na več načinov (Berry & Linoff, 2001, str. 75):

- s pomočjo spletnega sistema za registracijo uporabnika,
- na podlagi naslova za dostavo,
- z uporabo spletnega piškotka, ki se zapiše obiskovalcu ob obisku,
- na podlagi podatkov kreditne kartice.

2.4 Celosten pogled na reševanje problema

Kot smo že omenili, so spletni obiskovalci, njihove preference in vzorci uporabe spletnih programskih rešitev, glavna tema našega preučevanja. Naš glavni problem se kaže v tem, da za naše analize potrebujemo kakovostne, zanesljive in dovolj bogate podatke, ki bodo

na najbolj celovit način odražali delovanje in uporabo našega sistema. Za razumevanje podatkov je predpogoj dobro poznavanje celotnega procesa, tako iz poslovnega kot tudi iz tehnološkega vidika.

V smislu ustvarjanja optimalnega okolja za učenje, na podlagi analiz vedenja strank, moramo zagotoviti vsaj pet osnovnih zahtev, tako iz vidika širine in globine podatkov. Širina se nanaša na povezavo različnih podatkov iz različnih podatkovnih virov, globina pa se nanaša na časovno dimenzijo, ki mora biti integrirana v podatkih. Zahteve, ki jim moramo zadostiti, so tako (Berry & Linoff, 2001, str. 337-341):

- možnost, da lahko podatke oziroma informacije, s katero posamezno stranko opisujemo v nekem časovnem obdobju, združimo v en sam zapis, ki ga imenujemo podpis stranke (angl. *customer signature*),
- možnost, da lahko na podlagi podatkov ustvarimo podpise strank, ki jih opisujejo kadarkoli v preteklosti, kar je še posebej pomembno, če se lotimo preučevanja iz vidika prej in po nekem dogodku,
- da bi lahko razumeli posamezne podpise strank, moramo imeti možnost vrtanja v globino, v podrobne podatke o posameznih odločitvah strank,
- potrebno je spremljati počasne spremembe podatkov o strankah, ki se dogajajo v daljšem časovnem obdobju,
- moramo razumeti trženjske akcije in ostale aktivnosti, ki so se zgodile v preteklosti, še posebej kaj smo od njih pričakovali in kaj je bil dejanski rezultat.

Zaradi uporabe različnih namenskih programskih rešitev in različnih operativnih podatkovnih virov, lahko govorimo o porazdeljenem sistemu. Z vidika analiz podatkov moramo zato najprej zagotoviti primeren sistem, ki nam bo omogočal ustrezne integracije in transformacije podatkov.

V nadaljevanju si bomo ogledali primer sistema za analize podatkov spletnih rešitev, katerega osnovna ideja je celosten zajem podatkov o poslovnem procesu, uporabnikih in njihovi uporabi spletnih programskih rešitev.

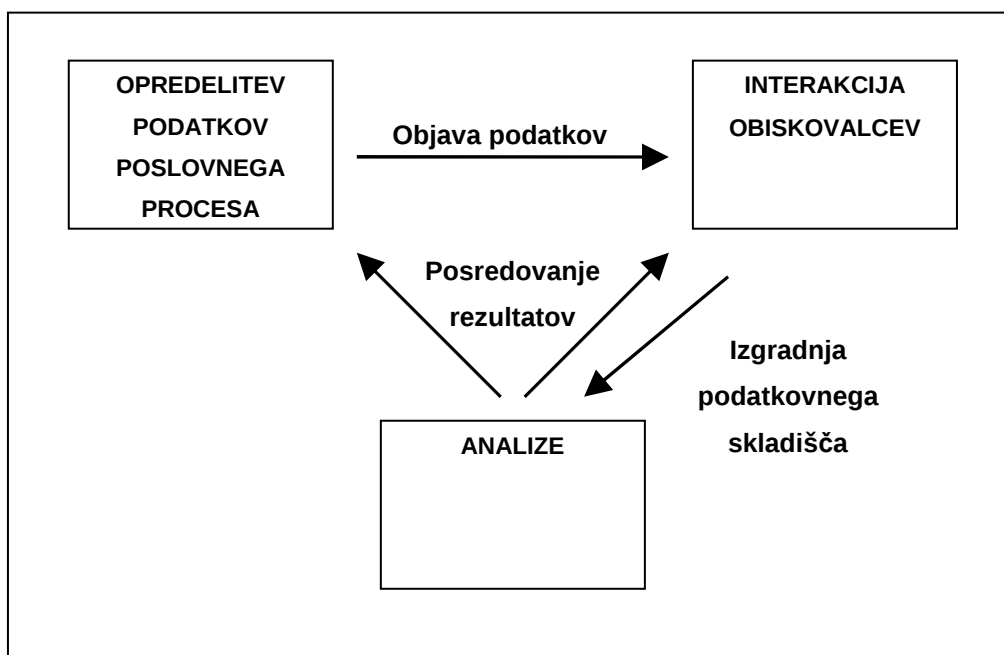
3 NAČRTOVANJE USTREZNEGA MODELA ZA ANALIZE PODATKOV

Na področju spletnih rešitev ni neke standardne ali predpisane arhitekture, lahko pa rečemo, da obstajajo podobne potrebe in značilnosti podprtja različnih poslovnih aktivnosti, ki tako narekujejo različne performančne zahteve, različno organiziranost podatkov, različne načine interakcije z uporabniki ter različne potrebe po predstavitvi informacij (Berry & Linoff, 2001, str. 83-85).

Področje spletnega poslovanja, je idealno tudi za izvajanje rudarjenja podatkov v širšem smislu ali tako imenovanega procesa odkrivanja znanja (angl. *knowledge discovery process*). Sestoji lahko iz preprocesiranja, poročanja, raziskovalne analize, vizualizacije in modeliranja (Ansari et al., 2001, str. 1).

Vse to nas vodi do arhitekture, kjer imamo lahko različne naloge podprte z različnimi komponentami. Za njihove medsebojne povezave, oziroma transformacije podatkov, skrbijo posamezni podatkovni mostovi. Ti podatkovni mostovi omogočajo komponentam tudi dostop do podatkov in metapodatkov, ki jih lahko dodelimo v skupno uporabo.

Slika 3: Poenostavljena visoko-nivojska arhitektura sistema za analize podatkov spletnih programskih rešitev



Vir: S., Ansari et al., *Integrating E-Commerce and Data Mining: Architecture and Challenges*, 2001, str. 2

Komponenta za opredelitev podatkov poslovnega procesa je povezana s komponento za interakcijo kupcev. Ta komponenta je tako po eni strani vir določenih podatkov o naših uporabnikih, po drugi strani pa predstavlja tudi ciljno mesto poslovnih pravil in ocen, ki so rezultat komponente za analize ali za opredelitev podatkov poslovnega procesa. Preko komponente za interakcijo se v komponento za analize, ki je namenjena predvsem analitični obdelavi podatkov, pošljejo tudi podatki, ki izhajajo iz komponente poslovnega procesa. Rezultati komponente za analize, ki so ponavadi v obliki nekakšnih modelov in ocen, se nato s pomočjo povratne povezave posredujejo nazaj v komponento za interakcijo ali pa v komponento za opredelitev poslovnega procesa.

Kot navajajo Ansari in sodelavci (2001, str. 2-4), se lahko arhitektura sistema od primera do primera razlikuje, poudarjajajo predvsem pomen integracije sistema za analize podatkov spletnih rešitev v sistem spletne prodaje. Na ta način si lahko prihranimo precej časa, celo do 80 odstotkov, ki ga moramo sicer vložiti v procese predprocesiranja, čiščenja in razumevanje podatkov, v okviru projektov odkrivanja znanja.

V nadaljevanju si bomo najprej ogledali primer osnovnih programskih rešitev spletne trgovine nato podrobneje priporočljivo arhitekturo sistema za integracijo analiz podatkov spletnih rešitev, ki jo bomo poskusili uporabiti na našem primeru.

3.1 Kratek pregled programskih rešitev spletne trgovine

Spletne trgovine ponavadi sestojijo iz posameznih, namenskih spletnih programskih rešitev, ki navzven tvorijo celovit sistem. Pri spletnih rešitvah se tako pogosto srečujemo z mešanjem tehnologij oziroma uporabo različnih tehnoloških platform in pristopov. Za lažje razumevanje kje in kako lahko posamezni podatki nastajajo, je na Sliki 4 prikazan primer uporabe rešitev za spletno prodajo.

Slika 4: Primer arhitekture spletnih programskih rešitev



Vir: Lastno delo.

Spletno mesto je v osnovi sestavljeno iz spletne programske rešitve za upravljanje spletnih vsebin in spletnega kataloga, ki služi predvsem predstavitvi posameznih vsebin in prodajnih artiklov. Na nivoju višje se nahaja nakupovalna košarica, ki tekom posameznega spletnega obiska omogoča obiskovalcem tako dodajanje kot tudi vpogled v njeno vsebino. Na najvišjem nivoju se nahaja še sistem za registracijo uporabnikov, ki skrbi predvsem za identifikacijo posameznih uporabnikov. V nadaljevanju si bomo še podrobneje ogledali posamezne programske rešitve, ki jih bomo uporabili v našem primeru spletne trgovine.

Poslovni programski paket

Najprej bomo predpostavili uporabo neke poslovne programske rešitve za poslovanje podjetij. Predpostavimo, da program poleg osnovnih funkcij omogoča tudi nadgradnjo z modulom, ki poslovnim uporabnikom omogoča urejanje in upravljanje spletne trgovine. Poslovni programski paket sicer ni neposredno del spletnega mesta, je pa v našem primeru zanimiv, ker se neposredno nanaša na modul za registracijo uporabnikov in katalog izdelkov.

V programski rešitvi lahko vsakemu posameznemu artiklu določimo osnovne podatke (naziv, kratek opis, proizvajalec, cene, količine itd.) in podrobnejše podatke (podroben opis, tehnične specifikacije, slike itd.). Izdelku lahko določimo prikaz v e-trgovini ali pa ga umaknemo iz spletne ponudbe. Predpostavimo, da se vsak artikel pojavlja samo v eni skupini izdelkov. Artiklom lahko po potrebi določimo tudi druge artikle in sicer kot

dodatno opremo in sorodne izdelke. Izdelki so smiselno urejeni po uporabnikom razumljivih skupinah, ki odražajo neke sorodne lastnosti izdelkov. Skupine so hierarhično urejene in lahko vsebujejo posamezne izdelke ali skupine. V našem primeru imamo skupine izdelkov razčlenjene na tri nivoje v globino (družina, skupina in podskupina izdelkov). V šifrantu kupcev imamo osnovne podatke o strankah, ločimo pa med pravnimi in fizičnimi osebami. Programska rešitev omogoča tako izvoz izbranih podatkov v elektronsko trgovino (npr. podatki o izdelkih, uporabnikih) kot tudi uvoz podatkov iz elektronske trgovine (npr. podatki o novoregistriranih uporabnikih, prejeta naročila).

Registracija uporabnikov

Gledano z vidika pravic uporabnikov, spletno mesto v osnovi sestoji iz dveh delov. Javni del, ki je prosto dostopen vsem spletnim obiskovalcem ter uporabniški del, ki od spletnih uporabnikov za ogled zahteva predhodno registracijo in uporabniško prijavo v spletni sistem. V našem primeru je uporabnik primoran opraviti registriracijo le, če hoče dokončati nakup in tako naročiti izbrane izdelke, ki se nahajajo v nakupovalni košarici. Z registracijo uporabnik sprejme splošne pogoje poslovanja in soglaša z obdelavo njegovih osebnih podatkov. Ob registraciji si izbere uporabniško ime in geslo, ter vpiše svoje osebne podatke, ki so ob morebitnem nakupu nujni za uspešno dostavo kupljenega blaga. Modul za registracijo tako uporabnikom omogoča, da jim ob ponovnem nakupu ni potrebno ponovno vnašati vseh podatkov, temveč le predizbrano uporabniško ime in geslo za prijavo v sistem.

Spletni katalog izdelkov

Izdelki so razvrščeni po prednastavljenih, hierarhično urejenih skupinah. Razvrstitev izdelkov določimo že v poslovnem programskem paketu, ki smo ga že omenili. Katalog izdelkov lahko iz navigacijskega vidika razčlenimo na tri osnovne zaslone: seznam skupin, seznam izdelkov in podrobni ogled posameznega izdelka.

Nakupovalna košarica

Spletno trgovino sestavlja tudi nakupovalna košarica, ki uporabnikom omogoča dodajanje izdelkov, spreminjanje količine ali odstranjevanje posameznih artiklov. Po končanem ogledu spletnih strani mora obiskovalec za dokončanje nakupa izbrati potrditev naročila, s

katerim dokončno potrdi nakup vsebine nakupovalne košarice. Za uspešen nakup mora biti uporabnik prijavljen v sistem.

Sistem za upravljanje ostalih spletnih vsebin

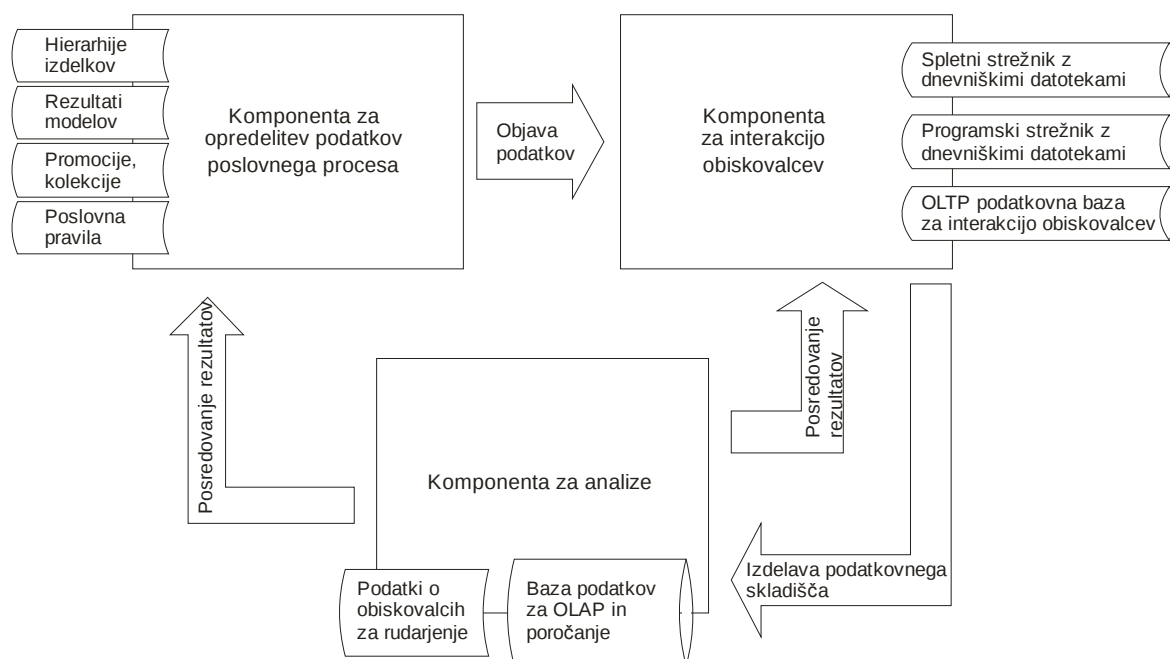
Da bi pritegnili spletne obiskovalce poskušajo ponudniki obogatiti vsebinsko plat spletišča z aktualnimi novicami iz sveta računalništva in primerjalnimi ocenami posameznih izdelkov. Spletna rešitev je preprost sistem za urejanje spletnih vsebin (angl. *Content Management System*, krat. *CMS*). Z geslom zaščiten uredniški del omogoča urednikom urejanje vsebine spletne strani neposredno prek spletnega brskalnika. Novice imamo razdeljene na štiri posamezne kategorije (akcije, novosti iz sveta računalništva, primerjalne ocene izdelkov in razno, kamor spadajo še vse ostale vsebine kot so opis podjetja, splošni pogoji uporabe, kontakt itd.).

3.2 Arhitektura sistema za analize podatkov spletnih rešitev

Za optimalno učinkovitost procesa odkrivanja znanja, moramo sistem za analize podatkov integrirati v naš obstoječi sistem za spletno prodajo. Z vidika integracije sistema za analize podatkov se idealna, visoko-nivojska arhitektura, sestoji iz treh glavnih komponent (angl. *main components*) in treh podatkovnih mostov (angl. *data transfer bridges*), ki te komponente povezujejo med seboj.

Na Sliki 5 (str. 24) vidimo primer podrobnejše arhitekture, s katero lahko rešujemo posamezne probleme integracije analiz podatkov spletnih programskih rešitev, ki smo jo v poenostavljeni obliki obravnavali že na začetku tega poglavja (Slika 3 na strani 19).

Slika 5: Podrobnejša visoko-nivojska arhitektura sistema za analize podatkov spletnih programskih rešitev



Vir: M., J., A., Berry & G., S., Linoff, *Mining the Web: Transforming Customer Data into Customer Value*, 2001, str.86

Arhitekturo sistema tako gradijo posebna področna podatkovna skladišča, ki temeljijo na različnih podatkovnih shemah, včasih na enakih podatkovnih poljih in skupni rabi podatkov ter metapodatkov. Vsaka komponenta ima svoje področno podatkovno skladišče, kjer se vodijo posamezni podatki o (Berry & Linoff, 2001, str. 87):

- pravih poslovnega procesa,
- o posameznih transakcijah, oziroma zahtevkih obiskovalcev,
- in podatki o obnašanju naših spletnih obiskovalcev.

V nadaljevanju si bomo še podrobneje ogledali značilnosti posameznih komponent ter ostalih sestavin te priporočljive arhitekture.

3.2.1 Opredelitev podatkov poslovnega procesa

V komponenti za opredelitev podatkov poslovnega procesa (angl. *business data definition module*) morajo poslovni uporabniki natančno definirati podatke in metapodatke, podatke o podatkih, ki so relevantni za spletno poslovanje.

To so lahko podatki o posameznih izdelkih (hierarhična skupina, vrsta uporabe, lastnosti in opis izdelka, proizvajalec, cena itd.), podatki o spletnih vsebinah (vrsta vsebine, grafične predloge, tematika, fotografije in ostalo gradivo, priponke, povezave itd.) ali pa podatki o različnih oblikovnih predlogah, ki jih ponavadi uporabljamo za posamezne prikaze in podatke ki jih prikazujejo (vstopna stran, pregled skupine izdelkov, podroben ogled posameznega izdelka, novice itd.).

Lahko rečemo, da to področno podatkovno skladišče vsebuje podatke o prodajnih artiklih, spletnih vsebinah in poslovnih pravilih, ki izhajajo iz poslovnega procesa podjetja. Izbira dovolj bogatega nabora podatkov je ključnega pomena, tako za nadaljnje procese raziskovanja podatkov kot tudi za poosebljanje spletnih vsebin obiskovalcem.

3.2.2 Objava podatkov

Podatkovni most za objavo podatkov (angl. *stage data bridge*) povezuje komponento za opredelitev podatkov poslovnega procesa s komponento za interakcijo obiskovalcev. Tu gre v bistvu za prenos oziroma posredovanje podatkov in metapodatkov, ki izhajajo iz prejšnje komponente, v komponento za interakcijo obiskovalcev.

3.2.3 Komponenta za interakcijo spletnih obiskovalcev

Komponenta za interakcijo (angl. *customer interaction module*) je, poenostavljeno povedano, spletni uporabniški vmesnik, ki povezuje podjetje in kupce direktno s procesiranjem njihovih transakcij, ki jih izvajajo na našem spletnem portalu. Komponento za interakcijo obiskovalcev lahko opišemo tudi kot vmesnik za zajem podatkov o posameznih odločitvah uporabnikov, ki lahko istočasno deluje kot sistem za ponujanje spletnih vsebin uporabnikom.

Sestavni del komponente za interakcijo je praviloma tudi **sistem za zajem spletnih podatkov**, ki je ponavadi del sistema za sprotno transakcijsko obdelavo podatkov (angl. *on-line transaction processing*, krat. *OLTP*). Zahteve po sprotni transakcijski obdelavi podatkov se tako neposredno odražajo v normalizirani, relacijski podatkovni bazi, ki je kot taka optimalna za uporabo pri hitrem procesiranju številnih zahtevkov, ki jih sprožajo spletni obiskovalci (Ansari et al., 2001, str. 3).

3.2.3.1 Sistem za zajem podatkov o uporabi spletnega mesta

Zelo pomembno je, da za zajem podatkov poskrbimo že na aplikacijskem nivoju spletnega strežniika (ang. *application server layer*) in ne na nivoju spletnega strežnika kot celote. Na ta način si prihranimo veliko truda, ki ga lahko kasneje vložimo v samo pripravo podatkov za nadaljnjo obdelavo.

Za čim širšo uporabnost zajetih podatkov poskrbimo tako, da v dnevniške datoteke spletnih programskih rešitev beležimo dovolj širok in raznolik spekter podatkov o uporabnikih, njihovih odločitvah in uporabi spletnih programskih rešitev. Le tako lahko dobimo ustrezne podatke, ki jih potrebujemo za uporabo v nadaljnjih analizah naših spletnih uporabnikov. Te podatke lahko porabimo tudi za morebitne nadaljnje prilagoditve oziroma poosebitve spletnega mesta uporabnikom, ki jih ponavadi rešujemo s pomočjo dinamičnih spletnih vsebin (Berry & Linoff, 2001, str. 85).

3.2.3.2 Beleženje dogodkov

Spletno mesto, kot ga vidijo obiskovalci, gradijo številne spletne povezave, ki se skrivajo za posameznimi elementi (teksti, slike, animacije...) in tako povezujejo posamezne spletne strani v neko logično celoto. Bistvena prednost, ki nam jo ponuja navedena arhitektura, je zajem podatkov programskih rešitev o posameznih dogodkih, ki so jih sprožili posamezni spletni obiskovalci s klikom na posamezno spletno povezavo. Tako poskušamo zajeti predvsem podatke, ki odgovarjajo na vprašanja o obiskovalcih in njihovi uporabi spletnih rešitev, v smislu kaj, kdo in kdaj. Predvsem je za nadaljnje analize zanimivo zaporedje posameznih dogodkov, na nivoju posameznega obiska spletnega obiskovalca. Na podlagi teh podatkov lahko nato k analizam pristopamo iz različnih vidikov. Oglejmo si še nekatere glavne dogodke, ki so relevantni za analize podatkov našega primera spletne trgovine.

Ogled posamezne strani

Ogled posamezne spletne strani je v našem primeru **osnovni dogodek**, ki je lahko zanimiv tako iz vidika števila ogledov kot tudi časa trajanja posameznega ogleda. Ob beleženju uporabnikovega ogleda posamezne strani je zelo smiselno razmisliti tudi o potrebah analize po nujnosti beleženja podrobnejših podatkov, kot so: katere posameze vsebine so

bile prikazane, v katero vsebinsko področje spada, katere grafične predloge smo uporabili za prikaz itd.

Status spletnih obiskovalcev

Z vidika poslovnih dogodkov nam status spletnega obiskovalca pove marsikaj. Spletne obiskovalce lahko delimo na nove (**neregistrirane**) in obstoječe (**registrirane**) uporabnike spletnega sistema. Prav tako je za nas zelo pomemben podatek o **prijavi in odjavi** obstoječega, že registriranega uporabnika, v naš spletni sistem, saj nam omogoča tudi njegovo identifikacijo. Kot upravljalci spletnega mesta poskušamo stimulirati tako registracijo kot tudi prijave uporabnikov v spletni sistem, saj jih lahko le tako na zanesljiv način identificiramo in povežemo njihove ponovne obiske tudi v nekem daljšem časovnem obdobju.

Nakupovalna košarica in koraki nakupa

Kot smo že omenili, je prodaja izdelkov osnovni namen spletne prodaje. **Dodajanje, urejanje ali brisanje** prodajnega artikla v nakupovalni košarici odraža posamezne nakupne odločitve spletnega obiskovalca. Sam postopek nakupa lahko razčlenimo na posamezne korake, in sicer **izbor** z dodajanjem artiklov v košarico, **potrditev** naročila z vnosom podatkov o dostavi in **dokončna potrditev** naročila izdelkov vključno z informativnim izračunom poštnine. Še posebej so zanimivi podatki o vsebinah tako imenovanih zavrženih nakupovalnih košaric, za katere obiskovalci niso dokončali postopka nakupa. Ključnega pomena za izboljšave spletnega mesta so tudi podatki o tem, v katerem koraku so obiskovalci spletni nakup opustili.

Izbor dogodkov je predvsem odvisen od namena našega preučevanja. V našem primeru bi bili zanimivi tudi razni dogodki, kot so kliki na oglasne pasice in ostale spletne povezave, ki jih ponujamo v okviru spletnega mesta.

Beleženje eksternih dogodkov

Včasih je smiselna tudi vpeljava sistematičnega beleženja drugih zunanjih (eksternih) dogodkov (npr. posameznih promocijskih akcij), ki so relevantni za našo analizo oziroma lahko vplivajo na rezultate naše analize. Take dogodke je težko zajeti v okviru obstoječega

informacijskega sistema, zato jih za namene analize ponavadi beležimo ročno, opisno in po nekem kronološkem in hierarhičnem zaporedju.

3.2.4 Izdelava podatkovnega skladišča

Most za izdelavo podatkovnega skladišča (angl. *build data warehouse bridge*) je namenjen prenosu podatkov, ki smo jih zajeli v komponenti za interakcijo obiskovalcev ali definirali v komponenti za opredelitev poslovnega procesa, v komponento za analize. Sistem za zajem podatkov je ponavadi sestavni del sistema za sprotno transakcijsko obdelavo podatkov. Ker pa je relacijska shema OLTP sistema nadzorovana s strani arhitekture, jo lahko na preprost način pretvorimo v večdimezijsko, zvezdasto shemo, ki je najbolj primerna za analize (Ansari et al., 2001, str. 3).

3.2.5 Komponenta za analize

Komponento za analize (angl. *analysis module*) ponavadi predstavljajo razna orodja, ki tvorijo integrirano okolje za podporo odločanju. S temi orodji lahko pridobljene podatke poljubno pretvorimo, na njih izvedemo algoritme za rudarjenje podatkov, pripravimo poročila in primerne grafične preglede ali pa jih analiziramo s pomočjo orodij za sprotno analitično obdelavo podatkov (angl. *On-line Application Processing*, krat. *OLAP*). Bogat izbor podatkov in metapodatkov daje komponenti za analize številne prednosti (Ansari et al., 2001, str. 5-7).

3.2.6 Posredovanje rezultatov

Most za posredovanje rezultatov (angl. *deploy results bridge*) je ključnega pomena za zaokrožitev cikličnega procesa. Ta podatkovni most nam, z ustrezno transformacijo podatkov, omogoča prenos rezultatov (podatkov, modelov, ocen, rezultatov in novih atributov) tako v komponento za opredelitev podatkov poslovnega procesa kot tudi nazaj v komponento za interakcijo obiskovalcev, za morebitne potrebe poosebitve spletnega mesta.

Osnovni ideji arhitekture sistema sta omogočanje **skupne rabe metapodatkov**, ki jih tako lahko uporabljamo v vseh treh komponentah, in **možnost uporabe rezultatov analize**, ki

jih lahko direktno uporabimo v komponentah za opredelitev poslovnega procesa in interakcijo obiskovalcev.

4 PREGLED POSAMEZNIH REŠITEV

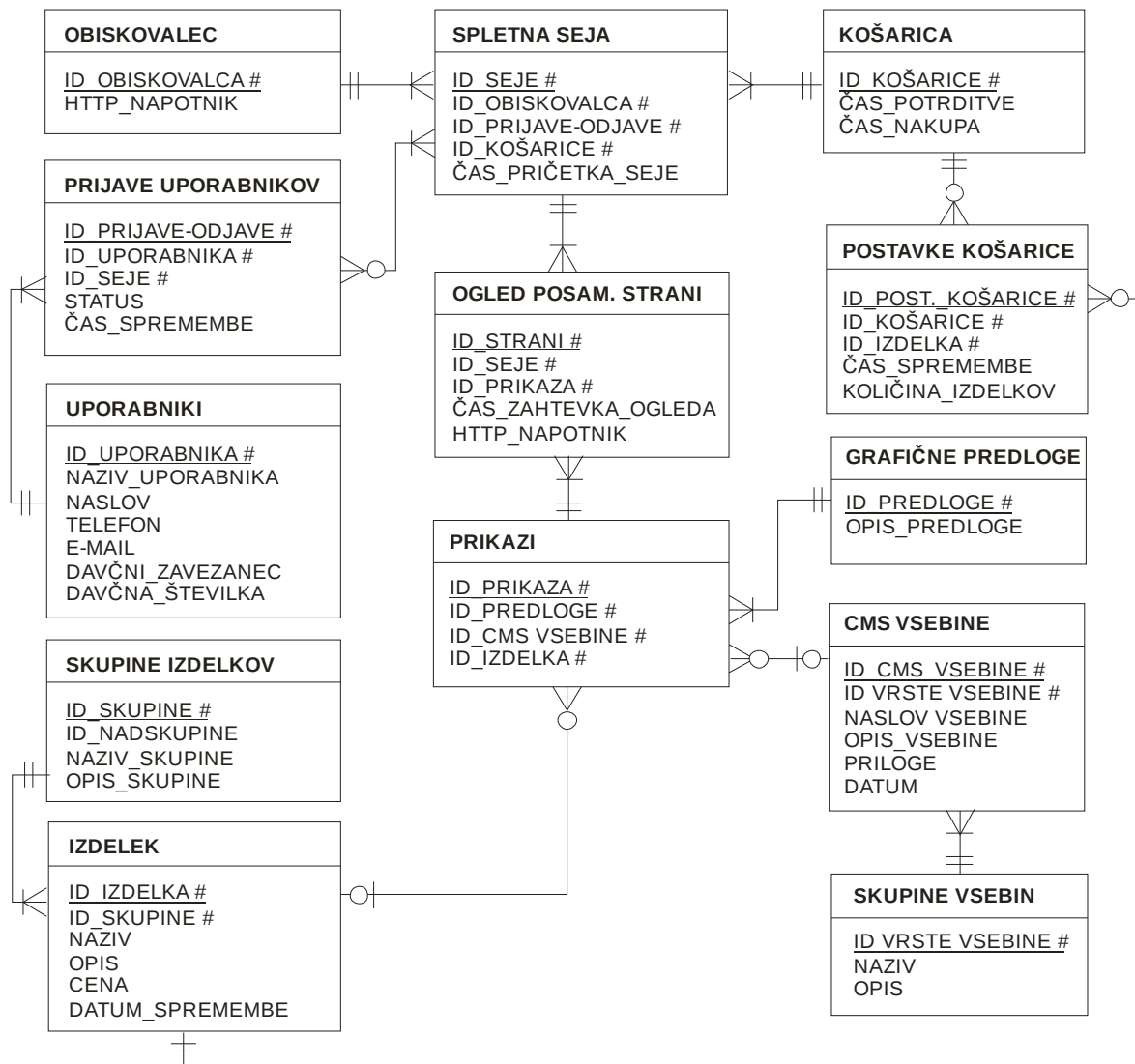
Na koncu bomo predstavili še glavne značilnosti, prednosti, slabosti in možnosti uporabe poročil, sprotne analitične obdelave podatkov ter podatkovnega rudarjenja, ki se nahajajo v naši arhitekturi kot sestavni del komponente za analize. Še prej pa si podrobneje ogledimo primer rešitve za zajem podatkov, ki se v arhitekturi nahaja v komponenti za interakcijo spletnih obiskovalcev.

4.1 Sistem za zajem podatkov spletnih rešitev

Kot smo že omenili je sistem za zajem podatkov ponavadi del OLTP (angl. *On-line Transaction Processing*), ki je sestavni del komponente za interakcijo spletnih obiskovalcev.

V nadaljevanju si pogledjmo kako bi lahko v našem primeru deloval celostni zajem podatkov s katerim bomo poskušali, že na nivoju spletnih programskih rešitev, zajeti ključne podatke o posameznih obiskovalcih ter njihovi uporabi spletnih programskih rešitev.

Slika 6: ER diagram – zajem podatkov v komponenti za interakcijo



Vir: Lastno delo.

Spletna seja

Ob prihodu na spletno stran se obiskovalcu avtomatsko zapiše spletni piškotek in dodeli spletna seja, kateri lahko istočasno pripada samo en spletni obiskovalec. Ob tem beležimo tudi čas pričetka seje. Na tem mestu moramo opozoriti tudi na problem ugotavljanja oziroma prepoznavanja zaključka posamezne seje, ki smo ga sicer že omenili.

Obiskovalec

Vsak spletni obiskovalec lahko znotraj posameznega spletnega obiska hkrati sodeluje le v eni seji sicer pa velja, da v nekem daljšem časovnem obdobju, če nas ponovno obiskuje, sodeluje v večih spletnih sejah.

Ogled posamezne strani

Ogled posamezne strani je v našem primeru osnovni dogodek na katerega vežemo več različnih spremenljivk. Tako za vsak posamezen ogled strani beležimo tudi čas ogleda in napotitelja, kar nam pove kako je obiskovalec na to stran prišel. Ogledi posameznih strani, oziroma zaporedje ki ga ti znotraj posamezne seje tvorijo, je še posebej zanimivo iz vidika analize poteka spletnih povezav. Če pa želimo vedenje obiskovalcev spremljati tudi iz vidika razvoja kupcev, moramo vpeljati v sistem tudi beleženje posameznih odločitev, ki jih je obiskovalec med ogledom opravil.

Prikaz

Spletnim obiskovalcem se lahko, v okviru ogleda posamezne strani, prikažejo različne vsebine (izdelki ali novice), ki za prikaz uporabljajo tudi različne grafične predloge.

Vsebina nakupovalne košarice

Beleženje dodajanja, odstranjevanja in popravljanja vsebine nakupovalne košarice zagotovo sodi med osnovne dogodke, ki lahko odražajo vedenje uporabnikov. V naše analize jih moramo vključiti, če želimo dobiti podrobnejši vpogled v njihove nakupne odločitve.

Nakupovalna košarica

Poleg vsebine moramo beležili tudi status spletne košarice, je bila njena vsebina naročena ali je končala v zavrženi košarici.

4.2 Priprava poročil

S poročanjem v obliki poročil poskušamo poslovnim uporabnikom omogočiti nekakšen kontinuiran pregled nad tekočim delovanjem in uporabo sistema za spletno prodajo. S pomočjo predpripravljenih vprašanj poskušamo iz različnih vidikov zajeti glavne indikatorje, ki na svoj način definirajo naše spletno poslovanje. Na podlagi teh kazalcev lahko poslovni uporabniki skozi različne poslovne potrebe, kot smo jih navedli že v začetku tega diplomskega dela, spremljajo oziroma ocenjujejo rezultate ter jih primerjajo med seboj v nekem časovnem obdobju.

Tako lahko s poročanjem zadovoljujemo različne informacijske potrebe poslovnih uporabnikov in tako odgovarjamo na različna vprašanja, kot so:

- Koliko obiskovalcev je v tekočem mesecu opravilo registracijo?
- Kateri so naši količinsko najboljše in najslabše prodajani artikli v tekočem mesecu?
- Kateri so naši najboljše in najslabše prodajani artikli po obsegu prihodkov, v tekočem mesecu?
- Katere posamezne izdelke si je v zadnjem mesecu ogledalo največ spletnih obiskovalcev?
- Kateri izdelki se največkrat pojavljajo v tako imenovanih zavrženih košaricah?

Prednost in smiselnost generiranja poročil se kaže predvsem v tem, da jih prilagodimo potrebam posameznih poslovnih uporabnikov. Urejenost, preglednost in enostavnost uporabe poročil bistveno vplivajo na razumevanje poslovnih uporabnikov in njihovo uporabo novopridobljenih informacij.

4.3 Možnosti uporabe OLAP orodij

Bistvo sprotne analitične obdelave podatkov je, da nam omogoča neposreden dostop do podatkovnih virov in izdelavo različnih pogledov na podatke. Z uporabo OLAP orodij lahko poslovnim uporabnikom omogočimo, da si sami izbirajo poglede na podatke, ki jih za neko odočitev potrebujejo. Prav tako lahko z enostavnim spreminjanjem pogledov odkrivajo, kateri podatki so relevantni za sprejemanje poslovnih odločitev. Prednosti, ki ju zagotavljajo ta orodja, sta predvsem velika prilagodljivost in samostojnost poslovnih uporabnikov pri dostopu do podatkov. Predpogoj sta seveda ustrezno predpripravljen podatkovni vir in preprostost uporabe orodij.

Poleg grafičnih prikazov, primerjav, prikaza izjem, izračunov in agregacije podatkov, orodja OLAP omogočajo tudi izvajanje tipičnih operacij nad pregledi podatkov:

- **Zvijanje** (angl. *roll-up*). Podatke prikažemo manj podrobno.
- **Vrtanje v globino** (angl. *drill-down*). Podatke prikažemo bolj podrobno.
- **Rezanje** (angl. *slice and dice*). Naredimo izbor podatkov ter prikažemo podkocko.
- **Vrtenje** (angl. *pivot*). Obračamo pogled na podatke.

Sistemi, ki so namenjeni sprotni transakcijski obdelavi podatkov, praviloma temeljijo na relacijski shemi podatkovne baze, ki je kot taka optimalna za delovanje sistema na operativnem nivoju in manj primerna za nadaljnjo analitično obdelavo. Najprimernejši podatkovni vir za uporabo OLAP orodij je področno podatkovno skladišče, ki vsebuje integrirane podatke, podatkovni model pa je praviloma večdimenzionalen, preprost in prilagojen potrebam analitika (Jaklič, 2002, str. 180-182).

Glede na večdimenzionalni pogled na podatke, ki ga posamezna OLAP orodja omogočajo, ločujemo **relacijski OLAP** (angl. *Relational OLAP* krat. *ROLAP*) in **večdimenzionalni OLAP** (angl. *Multidimensional OLAP* krat. *MOLAP*). Pri prvem gre za sisteme, ki so prilagojeni upravljanju relacijskih podatkovnih baz ter kot taki omogočajo učinkovito delo z velikimi količinami podatkov. S posebnimi orodji lahko podatke tudi preslikamo v večdimenzionalno kocko ter s tem skrijemo zapletenost relacijskega modela. Pri večdimenzionalnem OLAP-u gre za sisteme, ki temeljijo na uporabi večdimenzionalnih podatkovnih baz.

Sisteme, ki temeljijo na uporabi večdimenzionalnih podatkovnih baz, odlikujejo boljše zmogljivosti in tesnejše integrirana orodja. Večdimenzionalne podatkovne baze se lahko izkažejo za še posebej uporabne pri preučevanjih, kadar je relevantno oziroma želimo vedeti kaj se je in kaj se ni zgodilo (Kimball & Ross, 2002, str. 253).

Če želimo na našem primeru uporabiti OLAP orodja, moramo tako najprej poskrbeti za ustrezen prenos podatkov iz komponente za opredelitev podatkov poslovnega procesa in komponente za interakcijo kupcev v komponento za analize (Berry & Linoff, 2001). To se v naši arhitekturi zgodi s pomočjo podatkovnega mostu za izgradnjo podatkovnega skladišča, ki poleg uporabe postopkov ETL (angl. *Extract Transform Load*, krat. *ETL*) omogoča tudi uvoz in integracijo podatkov iz obeh komponent (Ansari et al., 2001, str. 3).

Tako za pripravo poročil, ki smo jih že omenili, kot tudi za nadaljnjo uporabo OLAP orodij, bomo potrebovali večdimenzionalno podatkovno bazo, ki nam bo omogočala retrospektivne poizvedbe iz različnih vidikov (Berry & Linoff, 2001, str 88-89).

OLAP poizvedbe nam lahko služijo za odgovore na vprašanja kot so:

- Kateri so najboljše in najslabše prodajani produkti?
- Katere skupine izdelkov nam prinašajo največ zaslužka?
- Katere izdelke si je ogledalo največ obiskovalcev?
- Od kje prihajajo obiskovalci na naše strani?

4.4 Možnosti uporabe podatkovnega rudarjenja

Pri analizi podatkov spletnih rešitev se soočamo z veliko količino podatkov, elektronski zajem podatkov nam zagotavlja zanesljive podatke, za pridobljeno znanje ali ugotovitve pa velja, da jih lahko hitro uporabimo ali implementiramo v sistem ter poskrbimo za merjenje njihovih koristi in učinkov. Zato je področje analiz spletnih rešitev idealno za uporabo podatkovnega rudarjenja (Ansari et al., 2001, str. 1).

Rudarjenje v podatkih (angl. *Data Mining*) lahko definiramo kot proces odkrivanja vzorcev in povezav v množici podatkov, z namenom boljših poslovnih odločitev. Za rudarjenje je značilno, da iskanje pravil in vzorcev poteka avtomatično. Pomembna zahteva pri uporabi orodij za rudarjenje je njihova preprostost in razumljivost. To je še posebej pomembno, ker so njihovi uporabniki praviloma (poslovni) analitiki, ki ponavadi ne obvladujejo poglobljenih znanj tehnik podatkovnega rudarjenja.

Rudarjenje v podatkih pa ne pomeni samo uporabe orodij za rudarjenje, temveč se to poimenovanje nanaša tudi na celoten proces, od identifikacije ter jasne definicije poslovnega problema, zbiranja in priprave potrebnih podatkov, uporabe orodij za rudarjenje, verifikacije in uvajanja rezultatov nazaj v poslovanje. Iz tega je razvidno, da gre za interdisciplinarno področje, zato naj bi v procesu sodelovali informatik, podatkovni analitik, poznavalec poslovnega področja in projektni vodja (Jaklič, 2002, str. 194-196).

Naloge, ki jih najpogosteje poskušamo opravljati z uporabo podatkovnega rudarjenja, so (Larose, 2005, str. 11-17):

- **Opisovanje** (angl. *Description*). Ko nekako poskušamo najti način za opisovanje vzorcev in trendov, ki se skrivajo v podatkih. Kakovostno opisovanje lahko dosežemo

tudi z raziskovalno analizo podatkov, ki je grafična metoda za ugotavljanje vzorcev in trendov v podatkih.

- **Ocenjevanje** (angl. *Estimation*) je podobno kot opisovanje, le da so ciljne spremenljivke pri ocenjevanju v obliki numeričnih vrednosti in ne kategoričnih.
- **Uvrščanje v skupine glede na podobnosti** (angl. *Classification*) ciljnih spremenljivk.
- **Predvidevanje** (angl. *Prediction*) je podobno kot uvrščanje v skupine in ocenjevanje, z razliko, da se rezultati pri predvidevanju nahajajo v prihodnosti. Katerekoli tehnike ali algoritmi, ki se jih poslužujemo za uvrščanje v skupine glede na podobnosti, lahko pod določenimi pogoji uporabljamo tudi za predvidevanje. To obsega tako uporabo tradicionalnih statističnih metod kot tudi tehnik uporabe podatkovnega rudarjenja kot so nevronske mreže (angl. *neural network*), odločitvena drevesa (angl. *decision tree*) in tehnika najbližjega sosedu (angl. *k-nearest neighbor*).
- **Razvrščanje v skupine** (angl. *Clustering*) poteka na način, da je znotraj skupine dosežena kar največja podobnost podatkovnih zapisov, med posameznimi skupinami pa karseda največja različnost.
- **Asociacija** (angl. *Association*) je namenjena ugotavljanju povezav med posameznimi atributi in sicer v smislu asociacijskih pravil ČE-POTEM, ki so v obliki povezave med vzrokom in posledico.

Podatkovno rudarjenje se je na področju trženja uveljavilo predvsem zaradi poudarjanja poslovne usmeritve k posameznemu kupcu, ki postavlja stranko v središče pozornosti podjetja. Za učinkovitost podatkovnega rudarjenja, za potrebe analitičnega CRM, mora podjetje zadostiti osnovnim pogojem (Berry & Linoff, 2004, str. 3):

- možnost spremljanja aktivnosti posameznih strank,
- zapomniti si mora, kaj so stranke počele tekom daljšega obdobja,
- potrebo se je učiti oziroma pridobivati znanje iz tega kar smo si zapomnili,
- ustrezno delovati na podlagi pridobljenega znanja.

Za podprtje oziroma zadostitev teh pogojev potrebujemo primerno arhitekturo sistema, ki jo ponavadi sestavljajo (Berry & Linoff, 2004, str. 3-6):

- **sistem za sprotno transakcijsko obdelavo podatkov**, ki skrbi tudi za zajem podatkov o interakciji strank,
- **podatkovno skladišče**, kjer hranimo zgodovinske podatke o vedenju strank,

- **podatkovno rudarjenje** za pretvorbo zgodovinskih podatkov v načrte za bodoče aktivnosti,
- **primerno strategijo managementa odnosov z odjemalci**, s katero te načrte pretvorimo v poslovno prakso.

S pomočjo podatkovnega rudarjenja lahko odgovorimo tudi na zapletenejša vprašanja:

- Katere lastnosti so značilne za obiskovalce, ki redno kupujejo v naši e-trgovini?
- Kaj karakterizira kupce, ki se odločajo za večje nakupe?
- Katere izdelke ponuditi posameznemu kupcu?

SKLEP

Kot smo v diplomskem delu do sedaj ugotovili, multidisciplinarnost analiz podatkov spletnih rešitev po eni strani izhaja iz tehnološkega po drugi pa iz vsebinskega in poslovnega poznavanja celotnega problema.

Analiz dnevniških datotek spletnega strežnika ni za zapostaviti, se pa moramo vsekakor zavedati določenih omejitev in pomanjkljivosti, če se analize lotimo na ta način. Analize dnevniških datotek so iz vidika podatkov bolj primerne za analize spletnega prometa. Ugotovili smo, da če želimo analizirati obnašanje in posamezne odločitve spletnih obiskovalcev moramo najprej poskbeti za zajem primernih podatkov, ki bodo to vedenje na ustrezen način odražali. Zajem podatkov v spletnih programskih rešitvah, ki nastajajo neposredno ob interakciji spletnih obiskovalcev, tako v našem primeru predstavlja optimalno rešitev.

Na podlagi primera priporočljive arhitekture smo ugotovili, da je potrebno k analizam podatkov, zajetih v spletnih programskih rešitvah, pristopiti na karseda celovit način. Za optimalno učinkovitost procesa odkrivanja znanja moramo sistem za analize podatkov integrirati v naš obstoječi sistem za spletno prodajo. Tako si lahko zagotovimo primerne in dovolj kakovostne podatke, istočasno pa si lahko prihranimo precej naporov, ki jih moramo sicer vložiti v predprocesiranje, čiščenje in razumevanje podatkov, v okviru projektov odkrivanja znanja. Osnovni ideji arhitekture sistema sta omogočanje skupne rabe metapodatkov, ki jih tako lahko uporabljamo v vseh treh komponentah, in možnost

uporabe rezultatov analize, ki jih lahko tako direktno uporabimo v komponentah za definiranje poslovnega procesa ter interakcijo obiskovalcev. Sistem za zajem podatkov je ponavadi sestavni del sistema za sprotno transakcijsko obdelavo podatkov, relacijska shema OLTP sistema pa je praviloma nadzorovana s strani arhitekture, zato jo lahko na dokaj preprost način pretvorimo v najprimernejšo obliko, potrebno za naše analize.

K analizam podatkov spletnih programskih rešitev lahko pristopamo tako z uporabo poročil, OLAP orodij ali s podatkovnim rudarjenjem. Kako bomo k posameznemu problemu pristopili je predvsem odvisno od tega, na kakšna vprašanja želimo odgovoriti. Prednost poročil in OLAP orodij se odlikuje predvsem pri lažjem razumevanju poslovnih uporabnikov, prednost podatkovnega rudarjenja pa se odraža predvsem v odkrivanju vzorcev in povezav v množici podatkov, ki jih ponavadi ponujajo spletne programske rešitve.

Bistvo analiz podatkov spletnih rešitev pa ne tiči le v uporabi kompleksnih orodij, ki jih lahko pri tem uporabljamo. Bistvo se skriva tudi v raziskovanju in želji po odkrivanju novih znanj, s katerimi želimo rešiti nastali poslovni problem ali podpreti ter nekako izboljšati poslovni proces. Pri tem smo velikokrat odvisni predvsem od lastne iznajdljivosti.

Analize podatkov spletnih programskih rešitev bodo postale ključnega pomena za uspešno poslovanje vsake spletne trgovine. Potencial, ki nam ga ponujajo spletne tehnologije, vsekakor moramo izkoristiti. Številne možnosti, ki jih ponuja zajem podatkov na nivoju programskih rešitev, nam odpirajo nove načine uporabe različnih podatkov o uporabnikih in njihovi uporabi našega spletnega sistema. Prednosti, ki jih ponujajo analize na osnovi podatkov spletnih rešitev, se odražajo v nepogrešljivih povratnih informacijah, ki jih potrebujemo tako za tekoče poslovanje, ažuriranje, prepoznavanje in uresničevanje potreb po posodobitvah, posebitvah in optimizaciji delovanja spletnega mesta kot celote.

Zavedati se moramo, da je vsebina našega diplomskega dela vsekakor šele vrh ledene gore. Na primeru integracije sistema za analize podatkov smo na kratko nakazali kako poteka zajem podatkov o uporabi spletnih rešitev ter možnosti njihovih analiz. Znanje, ki ga pridobimo, pa samo po sebi nima nikakršne vrednosti, če ga ne pretvorimo v prakso. Za

prihodnje delo nam tako iz vidika analiz podatkov spletnih rešitev ostajajo še številni izzivi in odprta vprašanja, predvsem glede nadaljnje uporabe rezultatov analiz.

LITERATURA IN VIRI

1. Ansari, S., Kohavi, R., Mason, L. & Zheng, Z. (2001). *Integrating E-Commerce and Data Mining: Architecture and Challenges*. San Mateo: Blue Martini Software.
2. Berry, M., J., A. & Linoff, G., S. (2004). *Data mining techniques: For Marketing, Sales, and Customer Relationship Management*. (2nd ed.) Indiana: Wiley Publishing.
3. Berry, M., J., A. & Linoff, G., S. (2001). *Mining the Web: Transforming Customer Data into Customer Value*. New York: John Wiley & Sons.
4. Bien, S., Bogataj, J., Pirc, M., N., Prelesnik, M. & Žaucer, A. (2006). *Zakon o varstvu osebnih podatkov (ZVOP-1): s komentajem*. Ljubljana: GV založba.
5. [Islovar] *Clickstream*. Najdeno 6. julija 2008a na spletnem naslovu http://www.islovar.org/iskanje_enostavno.asp.
6. [SearchCRM.com] *Clickstream analysis*. Najdeno 6. julija 2008 na spletnem naslovu http://searchcrm.techtarget.com/sDefinition/0,,sid11_gci786594,00.htm.
7. Jaklič, J. (2002). *Upravljanje in uporaba podatkov*. Ljubljana: Ekonomska fakulteta.
8. Kimball, R. & Ross, M. (2002). *Data warehouse toolkit: Complete guide to dimensional modeling*. (2nd ed.) New York: John Wiley & Sons.
9. Kovačič, A. (1998). *Informatizacija poslovanja*. Ljubljana: Ekonomska fakulteta.
10. Larose, T., D. (2005). *Discovering Knowledge in Data: Introduction to Data Mining*. Hoboken: John Wiley & Sons.
11. Potočnik, V. (2001). *Trženje v trgovini*. Ljubljana: GV založba.
12. [Islovar] *Poslovno obveščanje*. Najdeno 4. julija 2008b na spletnem naslovu <http://www.islovar.org/izpisclanka.asp?id=4856>.
13. *Web Analytics [Softlutions]*. Najdeno 6. junija 2008 na spletnem naslovu <http://www.softlution.com/en/solutions/WebAnalytics/>.
14. *Web Intelligence [University California, Irvine Extension]*. Najdeno 5. junija 2008 na spletnem naslovu http://unex.uci.edu/certificates/it/web_intel/.
15. [Wikipedia] *Web log analysis software*. Najdeno 15. junija 2008 na spletnem naslovu http://en.wikipedia.org/wiki/Web_log_analysis_software.

PRILOGA 1:

Pomeni nekaterih najpomembnejših izrazov, ki jih ZVOP-1 (2005) uporablja

»**Osební podatek** - je katerikoli podatek, ki se nanaša na posameznika, ne glede na obliko, v kateri je izražen.«

»**Posameznik** - je določena ali določljiva fizična oseba, na katero se nanaša osebni podatek; fizična oseba je določljiva, če se jo lahko neposredno ali posredno identificira, predvsem s sklicevanjem na identifikacijsko številko ali na enega ali več dejavnikov, ki so značilni za njeno fizično, fiziološko, duševno, ekonomsko, kulturno ali družbeno identiteto, pri čemer način identifikacije ne povzroča velikih stroškov, nesorazmerno velikega napora ali ne zahteva veliko časa.«

»**Obdelava osebnih podatkov** – pomeni kakršnokoli delovanje ali niz delovanj, ki se izvaja v zvezi z osebnimi podatki, ki so avtomatizirano obdelani ali ki so pri ročni obdelavi del zbirke osebnih podatkov ali so namenjeni vključitvi v zbirko osebnih podatkov, zlasti zbiranje, pridobivanje, vpis, urejanje, shranjevanje, prilagajanje ali spreminjanje, priklicanje, vpogled, uporaba, razkritje s prenosom, sporočanje, širjenje ali drugo dajanje na razpolago, razvrstitev ali povezovanje, blokiranje, anonimiziranje, izbris ali uničenje; obdelava je lahko ročna ali avtomatizirana (sredstva obdelave).

»**Avtomatizirana obdelava** – je obdelava osebnih podatkov s sredstvi informacijske tehnologije.«

»**Zbirka osebnih podatkov** – je vsak strukturiran niz podatkov, ki vsebuje vsaj en osebni podatek, ki je dostopen na podlagi meril, ki omogočajo uporabo ali združevanje podatkov, ne glede na to, ali je niz centraliziran, decentraliziran ali razpršen na funkcionalni ali geografski podlagi; strukturira niz podatkov je niz podatkov, ki je organiziran na takšen način, da določi ali omogoči določljivost posameznika.«

»**Osebna privolitev posameznika** – je prostovoljna izjava volje posameznika, da se lahko njegovi osebni podatki obdelujejo za določen namen in je dana na podlagi informacij, ki mu jih mora priskrbeti upravljalec po tem zakonu; osebna privolitev posameznika je lahko pisna, ustna ali druga ustrezna privolitev posameznika.

»**Anonimiziranje** – je takšna sprememba oblike osebnih podatkov, da jih ni več mogoče povezati s posameznikom ali je to mogoče le z nesorazmerno velikimi napori, stroški ali porabo časa.«

PRILOGA 2:

Nekateri najpomembnejši zakonski členi

Ustava Republike Slovenije

Ustava RS, 38. člen:»Zagotovljeno je varstvo osebnih podatkov. Prepovedana je uporaba osebnih podatkov v nasprotju z namenom njihovega zbiranja. Zbiranje, obdelovanje, namen uporabe, nadzor in varstvo tajnost osebnih podatkov določa zakon. Vsakdo ima pravico seznaniti se z zbranimi podatki, ki se nanašajo nanj, in pravico do sodnega varstva ob njihovi zlorabi.«

Zakon o varstvu osebnih podatkov (ZVOP-1, 2005)

ZVOP-1, 2005, II. del (Obdelava osebnih podatkov), 1. poglavje (Pravne podlage in nameni), Splošna opredelitev, 8. člen:

*(1) Osebni podatki se lahko obdelujejo le, če obdelavo osebnih podatkov in osebne podatke, ki se obdelujejo, **določa zakon** ali če je za obdelavo določenih osebnih podatkov podana **osebna privolitev posameznika**.*

*(2) **Namen obdelave** osebnih podatkov mora biti **določen v zakonu**, v primeru obdelave na podlagi **osebne privolitve posameznika** pa mora **biti posameznik predhodno pisno ali na drug način seznanjen** z namenom obdelave osebnih podatkov.*

ZVOP-1, 2005, II. del, 1. poglavje, 8. člen (Pravne podlage v zasebnem sektorju):

*(1) **Osebni podatki v zasebnem sektorju se lahko obdelujejo**, če obdelavo osebnih podatkov in osebne podatke, ki se obdelujejo, določa zakon **ali** če je za obdelavo določenih osebnih podatkov podana osebna privolitev posameznika.*

*(2) Ne glede na prejšnji odstavek se lahko v zasebnem sektorju obdelujejo osebni podatki posameznikov, ki **so z zasebnim sektorjem sklenili pogodbo** ali pa so na podlagi pobude posameznika z njim **v fazi pogajanj za sklenitev pogodbe**, če je obdelava osebnih podatkov potrebna in primerna za izvedbo pogajanj za sklenitev pogodbe ali za izpolnjevanje pogodbe.*

*(3) Ne glede na prvi odstavek tega člena se lahko v zasebnem sektorju obdelujejo osebni podatki, če je to **nujno zaradi uresničevanja zakonitih interesov zasebnega sektorja in ti interesi očitno prevladujejo nad interesi posameznika**, na katerega se nanašajo osebni podatki.*

ZVOP-1, 2005, IV. del (Področne ureditve), 1. poglavje (Neposredno trženje), Pravice in dolžnosti upravljalca, 72. člen:

(1) Upravljalec osebnih podatkov lahko uporablja osebne podatke posameznikov, ki jih je zbral iz **javno dostopnih virov** ali **v okviru zakonitega opravljanja dejavnosti**, tudi za namene ponujanja blaga, storitev, zaposlitev ali začasnega opravljanja del z uporabo poštne storitve, telefonskih klicev, elektronske pošte ali drugih telekomunikacijskih sredstev (v nadaljnjem besedilu: **neposredno trženje**) v skladu z določbami tega poglavja, če drug zakon ne določa drugače.

(2) Za namene neposrednega trženja lahko upravljalec osebnih podatkov uporablja le naslednje osebne podatke, ki jih je zbral v skladu s prejšnjim odstavkom: **osebno ime, naslov stalnega ali začasnega prebivališča, telefonsko številko, naslov elektronske pošte in številko telefaksa**. Na podlagi osebne privolitve posameznika lahko upravljalec osebnih podatkov obdeluje tudi druge osebne podatke, občutljive osebne podatke pa le, če ima za to osebno privolitev posameznika, ki je izrecna in praviloma pisna.

(3) Upravljalec osebnih podatkov mora neposredno trženje izvajati tako, da posameznika ob izvajanju neposrednega trženja obvesti o njegovih pravicah iz 73. člena tega zakona.