

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**ZAGOTAVLJANJE KAKOVOSTI TRANSAKCIJSKIH PODATKOV
O KUPCIH ZA ANALITIČNE POTREBE**

Ljubljana, september 2010

BOŠTJAN OSEAR

IZJAVA

Študent Boštjan Osenar izjavljam, da sem avtor tega magistrskega dela, ki sem ga napisal v soglasju s svetovalcem prof. dr. Jurij Jakličem, in da v skladu s 1. odstavkom 21. člena Zakona o avtorskih in sorodnih pravicah dovolim njegovo objavo na fakultetnih spletnih straneh.

V Ljubljani, dne _____

Podpis: _____

KAZALO

UVOD	1
1 IZVOR IN UPORABA TRANSAKCIJSKIH PODATKOV	3
1.1 Nastanek transakcijskih podatkov	3
1.1.1 Izvor transakcijskih podatkov	4
1.1.2 Oblika hranjenja transakcijskih podatkov	4
1.2 Uporaba transakcijskih podatkov	4
1.2.1 Analiza transakcijskih podatkov	5
1.2.2 Povezava kupca z opravljenimi transakcijami	6
1.2.3 Izdelava napovedi na podlagi transakcijskih podatkov	7
2 KAKOVOST PODATKOV IN INFORMACIJ	9
2.1 Problemi, ki izvirajo iz slabe kakovosti informacij	11
2.2 Zagotavljanje kakovosti informacij	14
2.2.1 Merjenje kakovosti informacij	17
2.2.2 Cena informacijske (ne)kakovosti	20
3 PRIMER BELEŽENJA IN UPORABE TRANSAKCIJSKIH PODATKOV	23
3.1 Analiza transakcijskih podatkov	23
3.1.1 Izbira metode in orodja za analizo	23
3.1.2 Izvor podatkov	24
3.1.3 Priprava podatkov za analizo	25
3.1.4 Struktura podatkov za analizo	26
3.1.5 Izvedba analize	27
3.2 Rezultati analize in ugotovitve	28
3.2.1 Ustreznost paketa naročnikovim potrebam	30
3.2.2 Program zadrževanja naročnikov in ciljano oglaševanje	31
3.2.3 Dejanska uporaba rezultatov analize	32
4 ANALIZA UGOTOVLJENEGA STANJA	33
4.1 Ocena kakovosti vira podatkov	33
4.1.1 Razumljivost	35
4.1.2 Natančnost	35
4.1.3 Jasnost	36
4.1.4 Uporabnost	36

4.1.5	Jedrnatost.....	36
4.1.6	Usklajenost	36
4.1.7	Pravilnost.....	37
4.1.8	Ažurnost	37
4.1.9	Udobnost.....	37
4.1.10	Pravočasnost	38
4.1.11	Sledljivost	38
4.1.12	Interaktivnost	38
4.1.13	Dosegljivost	38
4.1.14	Varnost	39
4.1.15	Vzdrževanost	39
4.1.16	Hitrost.....	39
4.2	Vzroki in posledice nekovosti informacij	39
4.2.1	Uporaba in pomen atributov izvornih podatkov	39
4.2.2	Vpliv informacij o prekinitvah naročniških razmerij	40
4.2.3	Vpliv (ne)integracije na kakovost informacij.....	42
4.3	Stroški informacijske (ne)kovosti	44
4.4	Predlogi in izboljšave.....	44
4.4.1	Izboljšava kakovosti informacij	45
4.4.2	Uporaba pristopov za zagotavljanje kakovosti informacij	50
4.5	Ovire in omejitve	51
4.5.1	Vpliv zakonodaje na hranjenje in uporabo podatkov	51
4.5.2	Organizacijske ovire	52
SKLEP		52
LITERATURA IN VIRI.....		54

KAZALO SLIK

Slika 1:	Pregled metodologije celovitega managementa s kakovostjo informacij.....	17
Slika 2:	Procesi življenjskega cikla vira	21
Slika 3:	Diagram podatkovne baze z zbranimi transakcijskimi podatki.....	26
Slika 4:	Razdelitev populacije uporabnikov glede na uporabo storitev	29
Slika 5:	Primerjava uporabe storitev uporabnikov treh grozdov	30
Slika 6:	Porazdelitev obstoječih naročniških paketov med grozde	31
Slika 7:	Kakovost vhodnih informacij.....	35
Slika 8:	Sklenitve in prekinitve naročniških razmerij ter aneksov na tedenski ravni za naročnike, aktivne med 2007 in 2009	41
Slika 9:	Razmerje med znano in neznano starostjo naročnikov	42
Slika 10:	Starost naročnikov	43
Slika 11:	Podatek o starosti za podjetja.....	43
Slika 12:	Kritične ocene po kriterijih informacijske kakovosti	45
Slika 13:	Register matičnih podatkov	47

KAZALO TABEL

Tabela 1:	Razlika med problemi povezanimi s kakovostjo podatkov in problemi povezanimi s kakovostjo informacij	10
Tabela 2:	Problemi, za katere je vzrok slaba kakovost informacij	13
Tabela 3:	Šestnajst kriterijev informacijske kakovosti	19
Tabela 4:	Vhodni atributi za izvedbo grozdenja.....	28
Tabela 5:	Ocena kakovosti informacij	34

UVOD

Razumevanje navad in obnašanja kupcev je ključno za uspešno poslovanje (Kawasaki, 2008, str. 296-300). Pri izvajanju poslovnih procesov organizacije zberejo ogromne količine podatkov, ki se čedalje bolj pogosto uporabljajo kot osnova za izdelavo profila kupcev (Malhotra & Birks, 2007, str. 139 - 145). Z izkoriščanjem teh informacij pa se pojavi vprašanje njihove kakovosti. Najbolj pereč problem, ki ga povzročajo neakovostne informacije, je negativen vpliv na najpomembnejše vire organizacije – človeške vire. Slaba kakovost informacij jim preprečuje učinkovito izvajanje dela hkrati pa povzroča odtujitev strank z uporabo napačnih informacij (English, 1999, str. XVII).

Kupci za sabo puščamo transakcijske sledi (Parsons & Maclaran, 2009, str. 187). Primer vsakodnevne aktivnosti je med drugimi tudi nakup prehranskih izdelkov v Trgovinah Tuš. Pri uveljavitvi popusta, moramo ob nakupu priložiti identifikacijsko kartico, s katero izkazujemo članstvo v Tuš klubu. Račun na dnu vsebuje tudi podatke o zbranih bonitetah. (Trgovine Tuš, Tuš klub, 2010). To pomeni, da se pri nakupu odpovemo anonimnosti v zameno za določeno ugodnost. Trgovcem je na ta način z analizo na voljo celotna transakcijska sled, ki jo na ta način za sabo puščamo ob vsakem nakupu. Za tržnega analitika je kartica zvestobe sredstvo, s katerim je mogoče povezati lastnosti kupca z dejansko opravljenimi nakupi (Malhotra & Birks, 2007, str. 127).

To pa seveda ni edini primer zbiranja podatkov o opravljenih transakcijah. V primeru Tuš kluba lahko ob nakupu ostanemo anonimni, če se odločimo, da ne izkoristimo članskega popusta oziroma popusta zvestobe. Povsod drugje, kjer pa je opravljanje storitev vezano na pogodbo, pa možnosti, da ostanemo anonimni enostavno nimamo. Kot primer naj navedem sklenitev naročniškega razmerja z operaterjem mobilnih storitev, kjer bodoči naročnik sklene pogodbo, ki vsebuje vse zakonsko obvezne osebne podatke. Uporabnik za opravljanje storitev prejme identifikacijski modul (angl. *Subscriber Identity Module*) na odstranljivi kartici z identifikacijskim modulom, ki omogoča globalno unikatno identifikacijo uporabnika (European Telecommunications Standards Institute - Specification of the Subscriber Identity Module, 2010). Ob opravljanju klicev, prenosu podatkov in opravljanju drugih storitev, uporabnik oziroma naročnik spet pušča transakcijsko sled. Razvoj informacijske tehnologije je omogočil shranjevanje ogromnih količin podatkov, ki jih tržniki sedaj lahko uporabljajo med drugim tudi za trženje na podlagi podatkov (angl. *data driven marketing*). Namesto, da bi poskušali razumeti psihološke ali sociološke nagibe k nakupu, se v tem primeru tržniki zanašajo na matematiko, da napovedo verjetnost, s katero bo kupec izvedel določeno odločitev za nakup (Parsons & Maclaran, 2009, str. 187). Na splošno gledano gre torej za iskanje morebitnih vzorcev med preteklimi transakcijami kupcev, ki bi jih tržniki uporabili za pridobivanje novih kupcev, zadrževanje obstoječih kupcev, opustitev nekaterih obstoječih

kupcev in analizo košarice (angl. *market basket analysis*) (Parsons & Maclaran, 2009, str. 188-189).

Podatki so surova snov, so produkt enega ali več poslovnih procesov, ki jih ustvari in vzdržuje. Ustvarjeni in vzdrževani podatki postanejo gradivo ali vir (angl. *input*) za številne druge poslovne procese (English, 1999, str. 52). Ker so zbrani transakcijskih podatki kupcev osnova za izdelavo analiz, je zagotavljanje kakovosti le-teh izrednega pomena. Na kakovost in zanesljivost rezultata analize neposredno vpliva kakovost vhodnih podatkov. Bistvenega pomena je torej, da poznamo stopnjo kakovosti vhodnih informacij, da bi lahko zaupali rezultatom.

Zaradi same količine in stopnje ustvarjanja novih transakcijskih podatkov je zagotavljanje njihove kakovosti večji izziv. Uvedba samodejne kontrole pravilnosti informacij na mestu nastanka ni dovolj za zagotavljanje kakovostnih informacij. Postopki za preverjanje pravilnosti vnosa preprečujejo očitne napake kot je vpis datuma »32.12.1999« ali pa navedbo dejstva, da ima nekdo 97 otrok. Samo ljudje lahko ustvarijo točne informacije (English, 1999, str. 401). Zato je uvedba procesov za zagotavljanje informacijske kakovosti ključnega pomena za organizacijo, slaba kakovost informacij namreč povzroča visoke stroške, s tem pa ogroža obstoj organizacije. Povečevanje kakovosti informacij je poslovna nuja, stvar preživetja organizacije ter konkurenčna prednost (English, 1999, str. 13).

Namen magistrskega dela je prispevati k izboljšavi kakovosti podatkov za analiziranje ter izboljšanje procesov z vidika zagotavljanja kakovost podatkov v podjetju, ki služi za primer.

Cilj magistrskega dela je na praktičnem problemu identificirati probleme, ki jih povzroča slaba kakovost informacij. Izdelal bom analizo obnašanja naročnikov telekomunikacijskega podjetja na podlagi transakcijskih podatkov naročnikov. Analiziral bom pozitivne učinke uporabe transakcijskih podatkov naročnikov kot osnove za analizo, podati pa želim uporabne predloge in ukrepe za izboljšavo izvedbe prihodnjih analiz in poslovnih procesov ter zagotavljanje višjega nivoja kakovosti informacij.

Magistrsko delo bo v prvem delu vsebovalo pregled možnosti, ki jih ponuja analiza podatkov transakcij kupcev, predvsem predlogi iz literature. Gre za primerjavo bolj razširjenih metod za spoznavanja obnašanja kupcev z nekaterimi novejšimi metodami, ki temeljijo izključno na podatkih.

V drugem delu se bom osredotočil na teorijo, ki je povezana s kakovostjo podatkov in informacij, pregledal bom nekatere izmed načinov ugotavljanja in zagotavljanja kakovosti informacij.

V tretjem delu se bom posvetil razlagi konkretnega primera, od kje izvirajo podatki, ki jih podjetje hrani o naročnikih, kje in v kakšni obliki se hranijo, ter najpomembneje, namen

uporabe teh podatkov. Opisal bom konkreten način uporabe transakcijskih podatkov in rezultate analize le-teh.

V četrtem delu bom podrobno opisal stanje kakovosti informacij uporabljenih za analizo, ter kako se izkoriščajo in trenutno uporabljajo. Opisal bom tudi pozitivne in negativne vplive kakovosti informacij oziroma na podlagi transakcijskih podatkov pridobljenega znanja. Podal bom predloge, konkretne ukrepe za izboljšavo postopkov shranjevanja in učinkovitejše izkoriščanje zbranih podatkov. Razložil bom znane razloge, ki so pripeljali do trenutnega stanja, znane razloge, ki preprečujejo spremembe trenutnega stanja. Možne ukrepe bom podal tako na organizacijski kot informacijski ravni.

Magistrsko delo bom zaključil s sklepnimi ugotovitvami.

Pri izdelavi magistrskega dela bom uporabil znanja, pridobljena v okviru podiplomskega in dodiplomskega študija, ter znanje, ki sem ga pridobil iz praktičnih izkušenj iz dela na podobnih projektih. Pri delu se bom naslanjal na praktični primer, teorijo, primere iz literature ter uporabil metodologije za ocenjevanje in zagotavljanje informacijske kakovosti.

1 IZVOR IN UPORABA TRANSAKCIJSKIH PODATKOV

Po Moorovem zakonu moč procesiranja informacij od leta 1971 narašča eksponentno, število tranzistorjev v procesorjih se je podvojilo vsaki dve leti (Anderson, 2009). Po Kryderjevem zakonu (Scientific American – Kryder's law, 2010) pa je naraščanje procesorske moči veliko počasneje od naraščanja gostote zapisovanja informacij na trdih diskih. Ta se je od uvedbe trdega diska leta 1956 do danes povečala za faktor petdeset milijonov. Tudi MacLennan, Tang in Crivat (2009, str. 2) menijo, da se je sposobnost zapisovanja prehitela sposobnost procesiranja informacij, rezultat pa je ogromna količina ustvarjenih podatkov, ki ostajajo v podatkovnih bazah. Rezultat tega neprekinjenega zbiranja podatkov je, da so organizacije postale bogate s podatki in revne z znanjem. Zbirke podatkov, pa so tako velike, da je njihova uporaba v praksi zaradi same velikosti postala omejena.

1.1 Nastanek transakcijskih podatkov

Transakcijski podatki opisujejo notranji, zunanji dogodek ali transakcijo, ki se zgodi tekom poslovanja organizacije. Transakcijski podatki so tipično združeni v transakcijske zapise, ki vsebujejo povezane matične in referenčne podatke (McGilvray, 2008, str. 293). Transakcijska sled, kot jo imenujeta Parsons in Maclaran (2009, str. 187), pa so informacijski zapisi o preteklih dogodkih, ko na nastali kot produkt interakcije med nami, kot kupci in ponudniki blaga ali storitev.

1.1.1 Izvor transakcijskih podatkov

Izvor transakcijskih podatkov so poslovni dogodki, ki jih beleži poslovna programska oprema. Tipično so to finančne aplikacije, celovite programske rešitve (angl. *enterprise resource planning*) in sistemi za management odnosov z odjemalci (angl. *customer relationship management*).

1.1.2 Oblika hranjenja transakcijskih podatkov

Podatkovna infrastruktura večine organizacij je običajno sestavljena iz več heterogenih sistemov. Sistemi popolnoma ločeno skrbijo za podporo določenih poslovnih procesov, med seboj pa navadno niso integrirani. Zaradi pomanjkanja povezav med posameznimi sistemi se pojavljajo težave, kadar bi želeli informacije, ki zajemajo področje, ki ga pokriva več ločenih sistemov.

Še eden izmed tipičnih problemov pa je, da so sistemi načrtovani za podporo specifičnim operacijam. Celovite programske rešitve ob upoštevanju računovodskih standardov transakcijske podatke beležijo do nivoja podrobnosti, ki omogoča pripravo računovodskih poročil, vodenje natančnega stanja zaloge in podobno, kar pa je običajno preveč podrobno za učinkovito izdelavo poročil ali zagotavljanje informacij na nivoju strategije podjetja.

Za rešitev opisanih problemov organizacije uporabljajo rešitve v obliki podatkovnega skladišča (angl. *data warehouse*). Namesto spreminjanja izvornih podatkov v podatkovno skladišče shranjujejo podatke v primernejši obliki, kar običajno pomeni, da se podatki pred prenosom v podatkovno skladišče integrirajo in zberejo. Takšen podatkovni model učinkovito podpira zahtevane poizvedbe, poročila in analize.

1.2 Uporaba transakcijskih podatkov

Transakcijski podatki sami se uporabljajo v različne specifične namene, odvisno od potreb in zmožnosti organizacij: poročanje ob predhodno definiranih dogodkih, izdelava posameznih analiz in preizkušanje hipotez ter nudenje podpore pri odločanju znotraj organizacije (Malhotra & Birks, 2007; Parsons & Maclaran, 2009, str. 188).

Transakcijski podatki so navadno shranjeni v podatkovnih bazah, zato je za odkrivanje morebitnih pomembnih korelacij, vzorcev in trendov v podatkih ogična izbira podatkovno rudarjenje. Analiza podatkov, ki ne opisujejo zgolj anonimne transakcije, ampak transakcije povezujejo s kupci in njihovi profili, lahko nudi odločevalcem bolj uporabne informacije. Ker pa vse transakcije niso vedno vezane na znanega kupca, je za organizacije, predvsem trgovske verige, problematika povezave kupca s transakcijami postala bistvenega pomena. Z povezavo transakcij z znanimi kupci se poveča razumljivost

rezultatov analize opravljene na takšnih podatkih, kar pomeni dvig kakovosti iz transakcijskih podatkov pridobljenih informacij.

Pri uporabi rezultatov analize, ugotovljenih korelacij, vzorcev in trendov, pa velja ohraniti previdnost. Ocene tveganja pri uporabi ugotovitev v prihodnosti namreč ni mogoče podati. Trendi in vzorci, ki veljajo za obstoječe podatke na katerih je bila izvedena analiza, se morda v prihodnosti ne bodo pojavljali več. Po drugi strani, pa lahko redke dogodke pomotoma označimo za na primer statistično napako. Redkim dogodkom na ta način ne posvetimo dovolj pozornosti kljub njihovim morebitnim vplivom na prihodnost.

1.2.1 Analiza transakcijskih podatkov

Podatkovno rudarjenje je metoda s področja strojnega učenja, namenjena pa je reševanju problemov na podlagi analize podatkov, ki so shranjeni v podatkovnih bazah (Witten & Frank, 2005, str. 5). Osnovni namen podatkovnega rudarjenja je pridobivanje znanja iz podatkov, ki jih imamo na razpolago (MacLennan et al., 2009, str. 2). Malhotra in Birks (2007, str. 137) definirata podatkovno rudarjenje kot proces odkrivanja pomembnih korelacij, vzorcev in trendov v podatkih z uporabo tehnik prepoznavanja vzorcev ter statističnih in matematičnih tehnik.

MacLennan et al. (2009, str. 4) omenjajo naslednje poslovne primere pri katerih lahko uporabimo tehniko podatkovnega rudarjenja:

- Ustvarjanje priporočil: na podlagi obnašanja vseh kupcev določenega ponudnika izdelkov ali storitev izpeljana pravila, ki se uporabijo za ustvarjanje priporočila storitve ali izdelka, ki bi kupca zanimal.
- Prepoznavanje anomalij: izbor podatkov, ki izstopajo. Možna je uporaba tako za prepoznavanje prevar kot tudi za preverjanje ustreznosti vnosnih podatkov.
- Analiza nihanj. Lahko pomaga prepoznavati kupce, za katere je verjetno, da bodo prešli k konkurenci.
- Upravljanje s tveganjem. Tehnika podatkovnega rudarjenja se lahko uporabi za določitev stopnje tveganja za dodelitev bančnega kredita.
- Ciljano oglaševanje. Z uporabo podatkov o preteklih nakupih ali pa zgodovini obiskanih strani lahko spletne strani prikazujejo obiskovalcem oglase z vsebino, ki je izbrana glede na obnašanje obiskovalca strani.
- Napovedi. Možna je uporaba za napovedovanje količine prodaje izdelkov, napovedovanje količine zaloga izdelkov in podobno.

Glede na to, da so transakcijskih podatki v digitalni obliki hranjeni v podatkovnih bazah, je iskanje vzorcev in pravil mogoče izvesti z uporabo metod s področja strojnega učenja. Witten in Frank (2005) poudarjata, da med strojnem učenjem in statistiko ni stroge ločne črte. Večina algoritmov za strojno učenje za ustvarjanje pravil ali dreves uporablja

statistične preizkuse. Statistični preizkusi se uporabljajo tudi za preverjanje in ocenjevanje točnosti modelov, dobljenih z strojnim učenjem.

Westphal in Blaxton (1998, str. 5) sicer navajata, da se težava pri statistični analizi podatkov pojavi v primeru, da manjka precejšnje število zapisov ali pa so atributi v kvalitativni in ne v kvantitativni obliki, toda tudi algoritmi za podatkovno rudarjenje teh težav ne rešujejo sami po sebi. Potrebna je primerna priprava vhodnih podatkov, ki je specifična za različne algoritme (pretvarjanje zveznih vrednosti atributov v diskretne s pomočjo razredov in podobno).

1.2.2 Povezava kupca z opravljenimi transakcijami

Transakcijski podatki brez povezave z osebo, ki je transakcijo opravila, za tržne analitike nimajo velike vrednosti. Za pomoč pri odločitvah, povezava transakcij in kupcev predstavlja naslednje prednosti (Malhotra & Birks, 2007, str. 127):

- Ustvarjanje profila kupcev. Mogoče je spremljanje tipov posameznikov, ki jih določena prodajna mesta privabijo, merjenje zaslužka glede na tip posameznika. Mogoča je izdelava profila idealnega kupca in načrtov, kako privabiti takšen tip kupca.
- Informacije o prodaji izdelkov ali storitev. Tipe izdelkov, ki se dobro ali slabo prodajajo je mogoče natančno nadzorovati. Glede na profil kupca, je k prodajnemu programu mogoče dodati tudi druge izdelke. S kombiniranjem informacij o prodanih izdelkih in profila kupcev je obnašanje kupcev mogoče bolje razumeti.
- Komunikacijske poti, ki delujejo ali ne delujejo. Mogoče je merjenje učinkovitosti posameznih komunikacijskih poti glede na tipe kupcev, ki jih privabljajo.
- Prilagajanje metod za razpečevanje. Prodajalci lahko prilagodijo načine za prodajo blaga glede na različne profile kupcev: nekateri kupci imajo raje velike, drugi manjše trgovine, nekaterim je bližje nakupovanje preko spleta.

Seveda so imeli Malhotra & Birks (2007, str. 127) v mislih predvsem prodajalne blaga, supermarkete, ki kupce povezujejo z nakupi preko kartic zvestobe. V drugih dejavnostih povezava kupcev s transakcijami lahko poteka na popolnoma drugačen način:

- Mobilni operaterji, povezava osebe, ki opravi klic in transakcije klica: v mobilnem aparatu to povezavo opravlja naročnikov identifikacijski modul (angl. *subscriber identity module*), ki omogoča globalno enolično identifikacijo uporabnika (European Telecommunications Standards Institute - Specification of the Subscriber Identity Module, 2010)
- Varovanje kopenske državne meje, registrska tablica vozila predstavlja ključ za povezavo različnih virov podatkov, na primer lastnikov avtomobilov in njihovih naslovov (Westphal, 2009, str. 123)

Kljub popolnoma različni naravi primerov, skupna točka ostaja enaka – nujen je obstoj stične točke oziroma ključa, ki omogoča integracijo različnih virov podatkov.

1.2.3 Izdelava napovedi na podlagi transakcijskih podatkov

Transakcijski podatki v kombinaciji s podatki uporabnikov so osnova za izdelavo in preizkušanje modelov za napovedovanje, s katerimi se poskuša napovedati, kako bi se kupci odzvali na situacije v prihodnosti (Malhotra & Birks, 2007, str. 137). Parsons in Maclaran (2009, str. 187) sta mnenja, da tržniki podatkovno rudarjenje lahko uporabijo tudi za napovedovanje potrošnikove odločitve o izvedbi nakupa, namesto da bi se zanašali na psihološke ali sociološke vzgibe za izvedbo nakupa.

Taleb (2007, str. 135) pa opozarja, da je napovedovanje prihodnosti na podlagi preteklih dogodkov izjemno nezanesljivo. Orodja, ki jih uporabljamo za napovedovanje prihodnosti ne upoštevajo redkih in nepredvidljivih dogodkov. Vzorce in pravila, ki jih najdemo med preteklimi podatki veljajo za pretekle podatke, mogoče je oceniti tudi stopnjo zaupanja. Pri uporabi pravil na drugi množici podatkov, v prihodnosti, pa je stopnjo zaupanja nemogoče oceniti, kar predstavlja precejšnje tveganje. Enako je z večino algoritmov za strojno učenje, ki se uporabljajo pri podatkovnem rudarjenju. Naslednji algoritmi redke dogodke obravnavajo kot statistično napaka:

- Naivni Bayes: sistematična metoda za učenje, ki temelji na štetju dokazov. Učenje s štetjem pojavov povezanosti med opazovano spremenljivko ter vsemi ostalimi (MacLennan et al., 2009, str. 216). V rezultatu so redki dogodki zanemarjeni.
- Odločitvena drevesa: gre za rekurzivno metodo deljenja množice podatkov v podmnožice. V vsakem koraku se za vse vhodne attribute preveri, kako učinkovito delijo množice podatkov na razrede (MacLennan et al., 2009, str. 236). Za zmanjševanje velikost dreves in večjo natančnost predvidevanja se uporablja tehnika obrezovanja dreves. Z obrezovanjem dreves se izločijo tudi podatki, ki so obravnavani kot napaka (Witten & Frank, 2005). Tudi v primeru odločitvenih dreves se redki dogodki izločijo s tehniko obrezovanja.
- Časovne serije: iskanje vzorca v podatkih, ki imajo dodano časovno dimenzijo. Za iskanje vzorcev se uporablja regresijska formula, po kateri se ugotavlja kako močno faktorji vplivajo na napoved spremenljivk. Microsoftova implementacija algoritma časovnih serij temelji na implementaciji algoritma odločitvenih dreves, kar pomeni, da je tudi obravnavanje redkih dogodkov enako (MacLennan et al., 2009, str. 268).
- Asociacijska pravila: iskanje objektov, ki se znotraj transakcij pojavljajo skupaj. Najbolj tipična uporaba je za analizo nakupovalne košarice, kjer se iščejo izdelki, ki se v posameznih košaricah pojavljajo skupaj (MacLennan et al., 2009, str. 343). Na splošno je za napovedovanje algoritem asociacijskih pravil slabši v primerjavi z drugimi algoritmi (MacLennan et al., 2009, str. 350). Ker pa gre za algoritem, ki v osnovi zgolj

ugotavlja pogostost korelacij (MacLennan et al., 2009, str. 359), so redke povezave med objekti prezrte.

- Nevronske mreže: izdelava matematičnega modela nevronske povezave, ustvarjanje in učenje umetnih nevronske poti – povezav med vhodi in izhodi (MacLennan et al., 2009, str. 371). Mera za kakovost učenja je podana glede na stopnjo izgube oziroma številom primerov, ki ne ustrezajo naučenim potem (MacLennan et al., 2009, str. 389). Napake in redki dogodki so z algoritmom nevronske mreže obravnavani enako.

Izjema pri obravnavanju napak in s tem tudi redkih dogodkov je algoritem grozdenja (angl. *clustering*). Uporablja se za klasifikacijo, grupiranje podatkov v množice, ki so sestavljene iz kar se da podobnih elementov glede na njihove lastnosti. Algoritem med podatki išče attribute, ki delujejo kot skupni imenovalci (MacLennan et al., 2009, str. 291). Tehniko grozdenja je mogoče uporabiti tudi za zaznavanje napak in anomalij (MacLennan et al., 2009, str. 297).

Uporaba algoritma grozdenja je zato primerna za uporabo na podatkih, kjer lahko vrednosti v redkih primerih močno odstopajo od povprečja. Zapisi z ekstremnimi odkloni so na ta način zbrani v ločenem grozdu. Zanimarjeni bi lahko bili le v primeru, da so parametri za izvedbo algoritma nastavljeni tako, da za nastanek grozda zahtevajo minimalno podporo primerov oziroma je število grozdov močno omejeno.

Z analizo transakcijskih podatkov lahko pridemo do informacij, pravil, vzorcev in znanja. Vse to pa velja za transakcijske podatke, ki so služili za analizo. Pri prenosu iz preteklih podatkov pridobljenega znanja v prihodnost, torej praktični uporabi znanja v aktualnih situacijah, pa so možne pasti na katere lahko naletimo.

Popper (2002, str. 10) meni, da je potrebno poudariti, da je teorija, ki je proces potrjevanja prestala pozitivno, lahko le začasno potrjena oziroma veljavna. Naknadne negativne ugotovitve lahko teorijo še vedno zavrnejo. Če Popperjev princip prenesemo na pravila in vzorce obnašanja, ki smo jih pridobili na podlagi preteklih transakcijskih podatkov, je pri njihovi uporabi na primerih v prihodnosti, na mestu določena mera dvoma. Ne glede na moderen pristop pri pridobivanju vpogleda v obnašanje kupcev, še vedno obstaja določeno tveganje, ki pa ga je nemogoče oceniti vnaprej.

Tako transakcijski podatki porabe storitev kot podatki gibanja finančnih trgov predstavljajo velikanske zbirke podatkov. Z samo velikostjo zbirk podatkov pa narašča tudi možnost, da se ustvarijo navidezno predvidljivi in ponavljajoči se vzorci. Moč naključij je dovolj, da se ustvarijo lažni vzorci in cikli, ki se zdijo predvidljivi in vnovčljivi. Toda finančni trg je še posebno nagnjen k takšnim statističnim prividom. Mehurčki in zlomi so del finančnih trgov. So neizogibna posledica človeške potrebe po iskanju vzorcev tam, kjer v resnici ne obstajajo (Mandelbrot & Hudson, 2004, str. 21).

Včasih je ogromno podatkov lahko brez pomena, drugič pa lahko zgolj en sam kos informacije nosi ogromen pomen. Ta princip nesimetrije informacij je preučeval Popper (2002), razvil je tehniko s katero je mogoče ovreči teorijo (angl. *Falsifiability*), ki pa je v praksi vse prej kot enostavna za izvedbo. Na problem asimetrije moramo biti pozorni predvsem pri interpretaciji rezultatov in potrjevanju njihove veljavnosti.

Primer, ki prikazuje to asimetrijo znanja, navaja Taleb (2008), nanaša pa se na čas, preden so Evropejci odkrili Avstralijo, torej pred začetkom 17. stoletja. Takrat je veljalo splošno prepričanje, da so vsi labodi beli, kar so potrjevali tudi empirični dokazi. Polged na prvega črnega laboda je bil za takratne ornitologe precejšnje presenečenje. Primer ilustrira našo omejitev pri učenju na podlagi dokazov ali izkušenj ter krhkost znanja. Eno samo opazanje lahko ovrže splošno trditev, ki je bila izpeljana na podlagi tisočletnih opazanj milijonov labodov, ki so potrjevala, da so le-ti beli. Vse kar je potrebno, da je trditev ovržena, je en sam črn labod.

Pri izvedbi podatkovnega rudarjenja celotno množico zapisov razdelimo na vzorčno podmnožico in podmnožico za kontolo pravil in rezultatov, pridobljenih na vzorčni podmnožici. Velikost vzorčne podmnožice je navadno trideset odstotkov celotne množice, je pa velikost in način vzorčenja mogoče prilagajati pri vsaki izvedbi algoritma za podatkovno rudarjenje (MacLennan et al. 2009). Iskanje vzorcev in povezav je tako omejeno le na zapise, ki so izbrani v vzorec, kar omogoči pojav asimetrije znanja. Najden vzorec morda ne velja, ali pa je stopnja zaupanja manjša, pri celotni populaciji. V primeru, da pa se vzorec ni pojavil, ne moremo sklepati, da vzorec na celotni populaciji ne obstaja. Problem asimetrije je pri podatkovnem rudarjenju mogoče omiliti z navzkrižnim preverjanjem in deljenjem množice zapisov na več množic.

Pri ocenjevanju kakovosti informacij pa je problem precej podoben. Navadno zaradi same količine prometnih informacij ni praktično izvedljiva podrobna analiza celotne množice in obsega informacij. Pri vzorčni analizi in interpretaciji dobljenih rezultatov velja enako pravilo, kakovostne informacije v vzorčni množici informacij ne pomenijo nujno kakovosti vseh informacij. Pri zaznani nekakovosti informacij v vzorčni množici, pa je verjetno, da gre za problem informacijske kakovosti na nivoju celotne organizacije.

2 KAKOVOST PODATKOV IN INFORMACIJ

Crosby (1979, str. 9) definira kakovost kot skladnost z zahtevami. Skladnost z zahtevami pa je natančno merljiva. To pomeni, da so merila za kakovost vezana na uporabo podatkov ali informacij. Z drugimi besedami, če so transakcijski podatki primarno namenjeni podpori osnovnim poslovnim procesom, so tudi primarni kriteriji za kakovost vezani na zahteve, s katerimi morajo biti podatki skladni. Problem pa se lahko pojavi pri uporabi transakcijskih podatkov v namene, ki prvotno niso bili načrtovani. S spremembo zahtev, ki

jih morajo podatki izpolnjevati za sekundarno rabo, se lahko spremeni tudi raven izmerjene kakovosti.

Jasna definicija problema izdatno pripomore k njegovi rešitvi (Berkun, 2009). Glede na to, da se tako problemi kot rešitve s področja kakovosti podatkov in kakovosti informacij popolnoma razlikujejo je potrebno jasno ločiti med podatki in informacijam. Podatek je le zapis brez pripisanega konteksta. Podatki v tem smislu postanejo informacije, ko jih povežemo z drugimi podatki: razumljiva množica stavkov, ki tvorijo sporočilo (Eppler, 2006, str. 22).

Razlika med pojmovanjem podatkov in informacij je najbolje razvidna iz problemov vezanih na težave s kakovostjo ter pristopom k reševanju teh težav, ki so opisani v Tabeli 1.

Tabela 1: Razlika med problemi povezanimi s kakovostjo podatkov in problemi povezanimi s kakovostjo informacij

Problem, vezan na kakovost podatkov	Problem, vezan na kakovost informacij
Duplikati, več virov podatkov	Nasprotujoči si predlogi v študiji ali analizi
Neobstoje relacij med podatki	Nejasne vzročne zveze v diagnozi
Vnosi brez pomena	Obsežna poročila brez logičnega toka
Napačno črkovanje	Nejasen jezik s slovničnimi napakami
Zastareli podatki	Analize niso ažuriranje v skladu z najnovejšimi dejstvi ali spremembami v organizacijskem smislu
Nedosledne oblike zapisa podatkov ali poimenovanja	Nedosledna postavitve ali navigacijska struktura
Založeni podatki shranjeni v napačni bazi podatkov	Izgubljeni ali založeni dokumenti
Komplicirane procedure za poizvedbo	Težavna navigacija po informacijah ali oteženo pridobivanje informacij
Napačno kodirani ali označeni podatki	Nezadostna ali neustrezna kategorizacija
Napačno vneseni podatki zaradi pomanjkljive kontrole vnosa	Negotovi zaključki na podlagi pomanjkljivih dejstev
Spreminjanje shranjenih podatkov	Spreminjanje odločitvenega procesa (zmeda, odvrčanje pozornosti, ipd)

Vir: M. J. Eppler, Managing Information Quality: Increasing the Value of Information in Knowledge-Intensive Products and Processes (2nd ed.), 2006, str. 28.

Probleme, povezane s kakovostjo podatkov, je mogoče rešiti z algoritmi čiščenja podatkov, statistično kontrolo procesa, postopki uparjanja s slovarjem ter drugimi metodami, medtem ko problemov, vezanih na kakovost informacij, navadno ni mogoče

rešiti z avtomatiziranim postopkom. Za odpravo slednjih so običajno potrebni naslednji elementi (Eppler, 2006, str. 29):

- priporočila za načrtovanje
- politika za objavljane vsebine
- izobraževanje na področju uredništva vsebin
- pravila za preverjanje pravilnosti vira informacij
- dodatne informacijske storitve in nakup infrastrukture
- Sprememba procesa za pregled in popravljanje informacij
- ipd.

2.1 Problemi, ki izvirajo iz slabe kakovosti informacij

Redman (2001, str. 39) navaja primer, za katerega je v celoti kriva slaba kakovost informacij – ameriško bombardiranje kitajske ambasade med vojno na Kosovu. Dogodki so si sledili v naslednjem vrstnem redu:

- ZDA in zavezniki pričnejo z napadi z zraka, da bi Srbe pregnali s Kosova, kar je potekalo med junijem in julijem 1999.
- Vladne agencije izberejo in predlagajo možne tarče za bombardiranje s strani ameriških zračnih sil.
- CIA je izbrala tarčo, za katero je menila, da je jugoslovansko skladišče orožja.
- Tarča je bila locirana na podlagi starega naslova. Namesto skladišča orožja, se je na tem naslovu sedaj nahajala kitajska ambasada.
- Običajno uspešen proces za pregled tarče, napake ni odkril. Analitik, ki je sumil v pravilno izbiro lokacije, je bil v kritičnem trenutku odličitve izven pisarne.
- Bomba je zadela stavbo, ki je bila ciljana, s tem pa uničila kitajsko ambasado. Ubiti so bili trije kitajski državljani, dvajset je bilo ranjenih.
- ZDA so priznale napako in na koncu plačale 28 milijonov dolarjev odškodnine kitajski vladi.

Običajno kakovost informacij nima tako grozovitih posledic, primer pa jasno ilustrira naslednje:

- Pravi problem, pridobitev informacij pravega naslova tarče in naslova kitajske ambasade, se je zgodil veliko višje v organizacijski strukturi od mesta, kjer je bila povzročena škoda.
- Napake je izjemno težko odpraviti.
- Resnična višina stroškov, ki jih povzroča slaba kakovost informacij, je neizmerljiva.

Al-Hakim (2007) omenja še več primerov incidentov z različnih področij, ki jih je zakrivila slaba kakovost informacij:

- Vesoljsko plovilo, ki ga je 11. decembra 1999 izstrelila NASA, da bi beležilo spremembe v ozračju Marsa, je bilo izgubljeno med pristankom 23. septembra 1999. Ugotovljeno je bilo, da je bil glavni razlog za izgubo plovila, napačno pretvarjanje angleških v metrične enote v segmentu programske opreme, ki je skrbel za navigacijo plovila na tleh. Problem informacijske kakovosti v tem primeru se nanaša na informacije, ki so bile pridobljene iz dveh različnih merilnih sistemov.
- Eksplozija v naftni rafineriji BP, Texas City ZDA, 23. marca 2005, je terjala 15 življenj, v eksploziji je bilo poškodovanih 170 ljudi. Glede na poročilo raziskave, ki je sledila nesreči, je bilo tragedijo zakrivilo neupoštevanje predpisanega postopka. Postopek je le tip informacije, problem informacijske kakovosti pa se v tem primeru nanaša na dosegljivost in uporabnost informacij.
- Teroristi so 11. septembra 2001 zrušili stolpnici WTC in v napadu poškodovali stavbo Pentagona. Poročilo komisije razkriva neučinkovito širjenje informacij – svaril in neuspešno povezovanje informacij, ki so jih posedovali posamezni agenti s področja nacionalne varnosti kot nekaj možnih vzrokov za neukrepanje ob terorističnih napadih.

Problemi, ki jih povzroča informacijska nekakovost, Eppler (2006, str. 30) deli na dve skupini. Prva skupina predstavlja probleme, pri katerih so informacije obravnavane kot produkt (neuporabnost, dvoumnost, nepopolnost in nepovezanost), druga pa predstavlja probleme, ki izvirajo iz procesa pridobivanja ali obdelave informacij (neustrezna predstavitev, nezanesljivost, nedosegljivost in izkrivljene informacije):

- Omejena uporabnost informacij zaradi preobremenitve uporabnikov informacij zaradi pomanjkanja ali zanemarjanja prečiščevanja, vzdrževanja in analize oziroma sinteze informacij.
- Dvoumnost informacij, ki vodi do različnih ali napačnih interpretacij zaradi netočnosti, uporabe okrajšav ali žargona ali različnega zornega kota uporabnikov informacij.
- Nepopolnost informacij, ki lahko vodi k pomanjkljivim odločitvam. Glavni razlog za takšne probleme je prevelika razdrobljenost dela, kot posledica razdrobljenosti pa specializacija.
- Nepovezane informacije, ki vodijo k zmedi. Vzroki za nepovezane ali nasprotujoče si informacije so pomanjkljivo usklajevanje med avtorji in distributerji informacij, nedorečena področja odgovornosti ali uporaba več nepovezanih virov informacij.
- Neustrezen format predstavitve informacij. Informacije so predstavljene v obliki, ki onemogoča neposredno uporabo brez sprememb.
- Informacije, ki so nezanesljive ali niso vredne zaupanja. Gre za povečano tveganje, da vsebujejo napake, ali pa njihovega izvora ni mogoče preveriti.
- Nedosegljivost informacij zaradi njihove izgube ali pa sprememb v tehnologiji.

- Izkrivljanje informacij, informacije dostavljene uporabnikom se razlikujejo od originalnih, izvornih.

Tipični problem vezani na neakovostne informacije in posledice, ki iz njih izvirajo in predlagane rešitve so zbrani v Tabeli 2. Gre za tipične probleme, ki se pojavljajo skozi celoten življenjski cikel informacij: od nastanka, shranjevanja in vzdrževanja do njihove uporabe. Izvzeti so problemi, ki so strogo vezani na informacijsko tehnologijo - podatkovne baze in podatkovna skladišča (Eppler, 2006, str. 32).

Tabela 2: Problemi, za katere je vzrok slaba kakovost informacij

Naziv problema	Opis	Posledice	Rešitev
Več virov	Več virov enakih informacij ustvari različne vrednosti in vodi k zmedi, manjši kredibilnosti in sprejemljivosti.	Upadanje uporabe informacij. Uporabniki informacijam ne zaupajo, vzdrževanje je oteženo in drago.	Uvedba enotnih definicij in doslednih postopkov.
Subjektivno ustvarjanje	Informacije so ustvarjene z uporabo subjektivnih presoj, ki vodijo k pristranskosti informacij.	Objektivnost informacij se zmanjša, oteženo je ocenjevanje informacij.	Več izobraževanja, boljša pravila, ekspertni sistemi.
Napake pri ustvarjanju	Sistemske napake pri ustvarjanju informacij vodijo k izgubi informacij.	Poveča se iskanje in popravljanje informacij.	Izboljšave procesov, spodbujanje, kontrola.
Preveč informacij	Velika količina shranjenih informacij oteži dostop do informacij v realnem času.	Preveč porabljenega časa za pridobitev in povzemanje informacij.	Analiza informacijskih potreb, redna izdelava povzetkov ali podmnožic pomembnih informacij.
Porazdeljeni sistemi	Porazdeljeni, heterogeni sistemi so vzrok za nedosledne definicije, neprimerno obliko in napačne vrednosti.	Informacij ni več mogoče združiti ali kombinirati zaradi nezdržljivosti in razlik v obliki.	Integracija sistemov ali uvedba zgolj enega sistema.

se nadaljuje

nadaljevanje

Naziv problema	Opis	Posledice	Rešitev
Sprememba potreb za nalogo	Spremenijo se naloge uporabnikov informacij, spremeni se organizacijsko okolje, informacije ki so pomembne in uporabne se spremenijo.	Razvijejo se neujemanja med informacijami, ki so na voljo in informacijami, ki so potrebne za opravljanje naloge.	Predvidevanje sprememb v nalogah in izvedba ustreznih popravkov procesov in sistemov še preden pride do kritičnega neujemanja.
Zahteve za varnost in zasebnost informacij	Enostaven dostop do informacij lahko vodi do konflikta glede varnostnimi zahtev, zagotavljanja zasebnosti in zaupnosti informacij.	Varnostni mehanizmi onemogočajo ali otežijo dostop, s tem pa se zmanjšuje vrednost informacij.	Razvoj doslednih politik in postopkov za varovanje informacij.
Pomanjkanje virov za obdelavo	Infrastruktura informacijske tehnologije je nezadostna in omejuje dostop do informacij.	Zaposleni, ki za svoje delo potrebujejo informacije niso motivirani, ne morejo delati učinkovito.	Razvoj politik nadgrajevanja informacijske tehnologije.

Vir: M. J. Eppler, *Managing Information Quality: Increasing the Value of Information in Knowledge-Intensive Products and Processes* (2nd ed.), 2006, str. 33.

Nekakovost pa je tudi glavni krivec za neuspeh podatkovnih skladišč. Podatkovna skladišča ne uspejo zaradi množice razlogov, vsi pa imajo skupni vzrok: nekakovost. Odpoved podatkovnega skladišča je neizbežna v primeru, da podatkovna skladišča uporabnikom informacij ne dostavljajo zanesljivih informacij, ki bi jim zadovoljivo pomagale pri odločitvah in strateških procesih (English, 1999, str. 4).

Pri uvedbi managementa matičnih podatkov, Loshin (2009, str. 18) opozarja, da je kakovost informacij ključnega pomena za uspeh. V nasprotnem primeru narašča tveganje, da produkt programa informacijskega managementa na nivoju celotne organizacije postane zgolj eden od mnogih neusklajenih informacijskih odložišč.

2.2 Zagotavljanje kakovosti informacij

Pr pristope za zagotavljanje kakovosti Redman (2001, str. 54) razdeli v štiri skupine:

1. Soočenje s posledicami informacijske nekakovosti.

2. Izvedba čiščenja informacij z različnimi metodami na nivoju organizacije.
3. Izvedba sprotnih čiščenj informacij. Čiščenje informacij ob nastanku, ali pa preden se uporabijo.
4. Iskanje virov napačnih informacij višje v informacijski verigi, kjer se informacije ustvarijo.

Organizacije, ki ignorirajo probleme s kakovostjo informacij, se nezavedno odločajo za prvi pristop, soočijo se s posledicami – slabimi odločitvami, nezadovoljnimi odjemalci in nezaupanjem znotraj organizacije.

V praksi se veliko organizacij odloča za drugi pristop, čiščenje podatkovnih baz, navadno s pomočjo avtomatiziranih orodij. Osnovni problem takšnega pristopa je v tem, da enostavno ne deluje na dolgi rok. Čeprav so informacije po čiščenju na zadovoljivem nivoju kakovosti, to ne velja za informacije, ki se po čiščenju dodajo oziroma spreminjajo. Postopek se nikoli ne konča. Druge pomanjkljivosti postopka očiščevanja so navadno visoki stroški izvedbe in lažen občutek, da so vse informacije pravilne, saj se nekatere informacije lahko izmuznejo skozi postopke in pravila za ugotavljanje napak.

Tretja možnost, izvedba sprotnih čiščenj informacij ima dva načina: čiščenje ob nastanku ali pa čiščenje informacij neposredno pred uporabo. Čiščenje informacij na ta način postane del običajnega delovnega procesa. Organizacije, ki se pogosto poslužujejo tega principa so navadno odvisne od najnovejših informacij, kot so na primer posredniške hiše.

Četrta, zadnja možnost zahteva, da se ugotovi in popravi izvor napačnih informacij. Pristop se od prejšnjih treh razlikuje v spremembi pozornosti od iskanja in odpravljanja napak, ki so se že zgodile, k preprečevanju napak. Ta pristop ima očitne dolgoročne prednosti.

Za preprečevanje nastanka problemov v procesih, ki temeljijo na znanju (angl. *knowledge-intensive processes*), Eppler (2006) predlaga uvedbo ogrodja za zagotavljanje informacijske kakovosti (ang. *information quality framework*). Ogrodje je sestavljeno iz podrobno razčlenjenih nivojev informacijske kakovosti, kriterijev za ocenjevanje kakovosti, faz za obdelavo informacij in principov za upravljanje. Samo merjenje kakovosti informacij pa za zagotavljanje njihove kakovosti ni dovolj, potreben je celovit pristop k informacijski kakovosti. V ta namen se organizacije lahko poslužujejo že znanih metodologij.

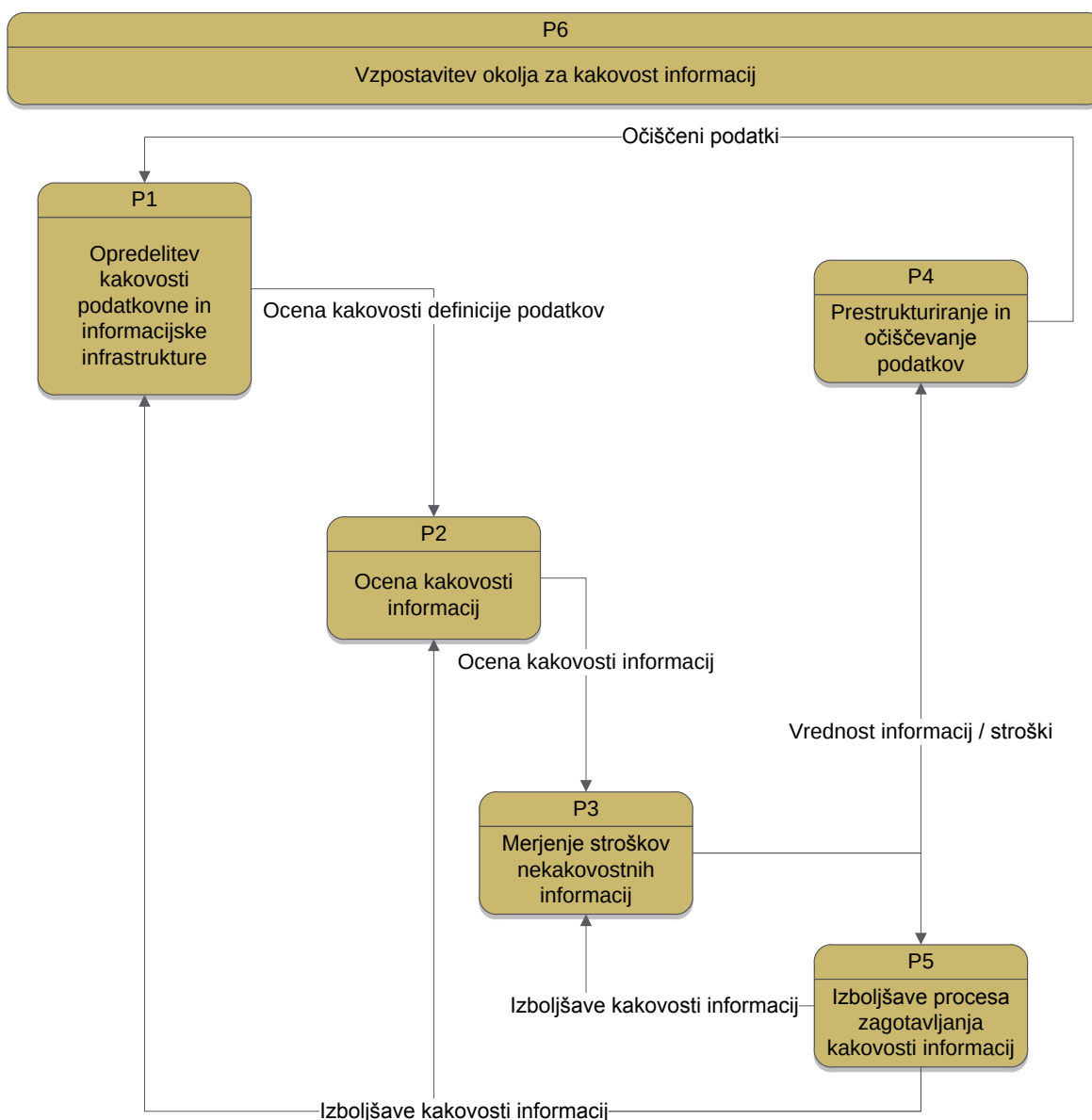
Metodologija merjenja in izboljševanja kakovosti informacij je opredeljena kot niz navodil in tehnik. Začne se pri zbiranju informacij o določenem predmetu in opisuje razumen proces od uporabe informacij do meritev in izboljšav kakovosti informacij skozi dane faze in odločitvene točke (Batini & Scannapieca, 2006, str. 161).

Batini in Scannapieca (2006, str. 170) kot metodologijo najbolj primerno managerjem predstavljata metodo celovitega managementa s kakovostjo informacij (angl. *total information quality management*). Metoda celovitega managementa kakovosti informacij je bila najprej načrtovana za uporabo na projektih podatkovnega skladiščenja, zaradi širokega obsega in nivojev podrobnosti, pa je uporabna tudi širše. Področja, na katerih pa je metoda izvirna in bolj razumljiva v primerjavi z drugimi metodologijami (na primer Istat in CIHI. Istat je zasnovana na italijanskem inštitutu za statistiko, ital. *Istituto Nazionale di Statistica*, CIHI pa nosi ime po kanadskem inštitutu za zdravje na katerem je bila razvita, angl. *Canadian Institute for Health*), pa je na področju stroškovne analize in managerskega pogleda na problematiko (Batini & Scannapieca, 2006).

Procesi, ki sestavljajo metodologijo za celovit management s kakovostjo informacij so (English, 1999, str. 69):

- Proces opredelitve kakovosti podatkovne in informacijske infrastrukture: izvedba procesa je pogoj, da je mogoče nadaljevati z merjenjem kakovosti informacij. Nemogoče je namreč ugotavljati kakovost proizvoda brez da bi se pred tem prepričali o točnosti specifikacij le-tega.
- Proces ocenitve kakovosti informacij: sam proces je na moč podoben procesu ugotavljanja kakovosti izdelkov . Obstajajo tehnična merila po katerih se ugotavlja, v kakšni stopnji informacije ustrezajo podani specifikaciji. Mogoče je tudi meriti stopnjo zadovoljstva uporabnikov in v kakšni meri so njihova pričakovanja izpolnjena.
- Proces meritve stroškov neakovostnih informacij: stroški neakovostnih informacij se odražajo v zmanjšanju prihodka in zasluka, kar pa je mogoče natančno izmeriti.
- Proces prestrukturiranja in čiščenja podatkov: gre za izboljšavo informacije kot produkta, pomanjkljive informacije pa je potrebno izboljšati do stopnje sprejemljivega nivoja kakovosti.
- Proces vpeljave izboljšav v proces zagotavljanja kakovosti informacij: ta proces je navada organizacij z visoko stopnjo kakovosti informacij. Gre za preprečevanje napak z planiranjem aktivnosti in analizo znanih problemov ter vzrokov za napake.
- Proces vzpostavitve okolja za kakovost informacij: predstavlja systemske, organizacijske in kulturne zahteve za zagotavljanje trajnega okolja za zagotavljanje kakovostnih informacij. Brez kulturnih značilnosti so lahko pobude za dvig informacijske kakovosti učinkovite zgolj na kratek rok.

Slika 1: Pregled metodologije celovitega managementa s kakovostjo informacij



Vir: L. P. English, *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, 1999, str. 70.

2.2.1 Merjenje kakovosti informacij

Organizacije, ki ignorirajo probleme vezane na kakovost informacij, se ukvarjajo le z odpravljanjem njihovih posledic (Redman, 2001, str. 54).

English (1999, str. 138) navaja tri razloge za opravljanje ocenjevanja kakovosti informacij:

- Prvi razlog je potrditev, da proces ustvarjanja informacij zagotavlja zanesljive informacije za odjemalce informacij.

- Drugi razlog je ugotavljanje zanesljivosti vira informacij.
- Tretji razlog pa je odkrivanje morebitnih problemov s kakovostjo informacij, ki jih bo potrebno rešiti.

Merjenje pa brez izbranih ustreznih kriterijev ni možno. Za zagotavljanje učinkovitosti, preprostosti in simetrije ogrođa za zagotavljanje informacijske kakovosti, Eppler (2006) predlaga izbor šestnajstih kriterijev za merjenje kakovosti, štiri za vsak posamezen nivo. Kriteriji so združeni v naslednje nivoje kakovosti:

- Nivo skupnosti (relevantnost): informacije so relevantne, če zadoščajo potrebam ciljnih uporabnikov oziroma skupnosti, ki jih potrebuje. Izbrani kriteriji na nivoju skupnosti: razumljivost, natančnost, jasnost in uporabnost.
- Nivo proizvoda (celovitost): informacije so celovite, kadar ne vsebujejo nepotrebnih elementov oziroma so jedrnate, informacije si ne nasprotujejo in uporabljajo enako obliko za predstavitev, ne vsebujejo napak ali napačnih elementov, na volj pa ni novejših informacij (informacije niso zastarane). Izbrani kriteriji na nivoju proizvoda: jedrnatost, usklajenost, pravilnost in ažurnost.
- Nivo procesa: kriteriji za nivo proizvoda informacije obravnavajo kot proces, ki se smatra za optimiziranega, če ustreza naslednjim kriterijem: ustvarjanje, administracija in dostava so kar se da udobni za vse deležnike (avtorje, administratorje in uporabnike). Izbrani kriteriji na nivoju procesa so: udobnost, pravočasnost, sledljivost in interaktivnost.
- Nivo infrastrukture: kriteriji zadnjega nivoja se nanašajo na zahteve informacijske infrastrukture. Informacijska infrastruktura je zanesljiva kadar ni izpadov ali obdobj, ko informacije niso dosegljive, je varna (ni nepooblaščenih dostopov do informacij), da je vzdrževana (informacije bodo enostavno dosegljive tudi v prihodnosti) in omogoča hitro interakcijo med shranjenimi informacijami in uporabniki informacij. Izbrani kriteriji nivoja infrastrukture so: dosegljivost, varnost, vzdrževanost in hitrost.

Tabela 3: Šestnajst kriterijev informacijske kakovosti

Nivo informacijske kakovosti	Kriterij za kakovost informacij	Opis	Si nasprotuje z
Nivo skupnosti (relevantnost)	Razumljivost	Ali je obseg informacij primeren? (ne preveč, ne premalo)	Jasnost Jedrnatost Pravočasnost Vzdrževanost
	Natančnost	Je informacija dovolj natančna in dovolj blizu realnemu stanju?	Jasnost Jedrnatost Pravočasnost
	Jasnost	So informacije razumljive ciljni skupini?	Natančnost Razumljivost
	Uporabnost	So informacije neposredno uporabne?	
Nivo proizvoda (celovitost)	Jedrnatost	So informacije očiščene vseh nepotrebnih elementov?	Razumljivost Točnost
	Usklajenost	Informacije ne vsebujejo nasprotujočih si dejstev ali dogovorov?	Pravočasnost
	Pravilnost	So informacije brez popačenj, pristranskosti in napak?	Pravočasnost
	Ažurnost	So informacije posodobljene in niso zastarele?	
Nivo procesa	Udobnost	Ali pridobivanje informacij ustreza potrebam in navadam uporabnikov?	Varnost
	Pravočasnost	So informacije obdelane in dostavljene brez zamika?	Razumljivost Točnost Jedrnatost Pravilnost
	Sledljivost	Je vir informacij viden (avtor, datum nastanka ipd)?	
	Interaktivnost	Ali uporabnik informacij lahko prilagodi informacijski proces?	
Nivo infrastrukture	Dosegljivost	Ali obstaja neprekinjen in neoviran dostop do informacij?	Varnost
	Varnost	So informacije zaščitene pred izgubo in nepooblaščenim dostopom?	Udobnost Dosegljivost
	Vzdrževanost	Je informacije mogoče urediti in posodobiti za uporabo v prihodnosti?	Razumljivost
	Hitrost	Je infrastruktura kos tempu uporabnikov?	Varnost

Vir: M. J. Eppler, *Managing Information Quality: Increasing the Value of Information in Knowledge-Intensive Products and Processes* (2nd ed.), 2006, str. 76.

Popoln seznam izbranih kriterijev z kratkimi opisi je podan v Tabeli 3. Naslednji kriteriji, pa imajo lahko dvoumen pomeni, zato je zanje potrebna natančnejša opredelitev (Eppler, 2006, str. 77):

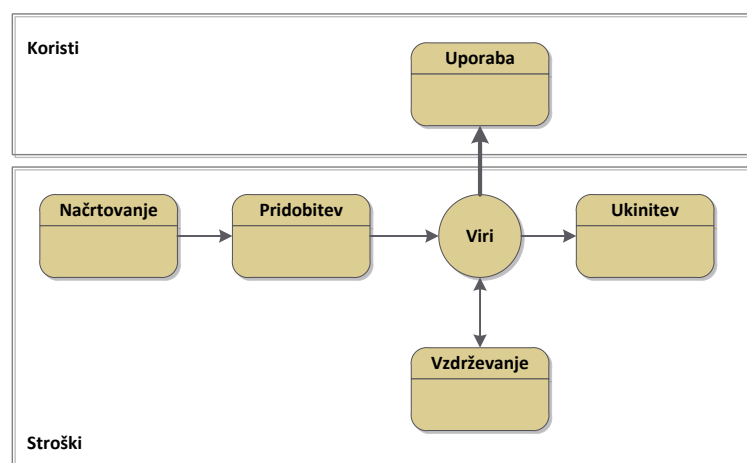
- Točnost: kako blizu resničnemu stanju je informacija. Tehnično je točnost opredeljena kot stopnja ujemanja z merilom ali dejanskim stanjem.
- Sledljivost: bistvena za kredibilnost informacije, mogoče je razbrati vir in metodologijo.
- Interaktivnost: sposobnost, prilagodljivost procesa za ravnanje z vsebino in sistema, da uporabniku zagotovi optimalen vmesnik za dostop oziroma predstavitev informacij.
- Udobnost: enostavnost uporabe. Udobnost je tesno povezana z interaktivnostjo, čeprav se kriterija razlikujeta. Preveč interaktivnosti lahko zmanjša udobnost oz. enostavnost uporabe.
- Uporabnost: opredeljena kot neposredna uporabnost v določenem kontekstu. Uporabna informacija je že urejena v obliko za delovanje in je ni potrebno pretvarjati ali preoblikovati, da bi se jo dalo uporabiti.
- Vzdrževanost: lastnost informacijskega sistema in informacij, shranjenih v njem, da jih je mogoče vzdrževati z razumnimi stroški. Določa lastnost informacijskega okolja, da ga je mogoče z razumni stroški upravljati glede na volumen, frekvenco in zahtevano kakovost vsebine ter infrastrukturo.

To ogrodje je mogoče uporabiti za sistematično urejanje kriterijev za kakovost informacij in sistematično preverjanje kakovosti informacij, informacijskih procesov in sistemov. Kot primer uporabe ogrodja, Eppler (2006, str. 180) navaja analizo vrzeli v dojemanju kakovosti med različni tipi informacijskih uporabnikov (na primer med vodstvom in strokovnjaki za informacijsko tehnologijo). Kriteriji za informacijsko kakovost so bili pretvorjeni v vprašanja, ki so bila zastavljena uporabnikom, s tem pa so uporabniki podali oceno kakovosti po kriterijih. Na tak način je mogoče določiti vrzel, ki jo predstavlja razlika med povprečnimi ocenami, ki sta jih podali obe skupini uporabnikov.

2.2.2 Cena informacijske (ne)kakovosti

Za učinkovito upravljanje English (1999, str. 203) predlaga enak pristop k ravnanju z informacijami, kot za vse ostale vire podjetja (človeški, finančni, osnovna sredstva, material in izdelki). Življenjski cikel vira je sestavljen iz petih procesov, ki so potrebni za njegovo upravljanje, to pa so: načrtovanje, pridobitev, uporaba, vzdrževanje in ukinitve virov. Za določitev vrednosti oziroma donosnosti virov v osnovi velja, da je pridobljena vrednost enaka razliki med koristmi in stroški.

Slika 2: *Procesi življenjskega cikla vira*



Vir: L. P. English, *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, 1999, str. 201.

Z informacijami kot virom so skozi življenjski cikel povezani naslednji donosi in stroški po procesih (English, 1999):

- **Načrtovanje:** načrtovanje se začne z razumevanjem poslanstva organizacije, ciljev in strategij. Potrebno je opredeliti informacijske zahteve na enak način kot bi izvedli načrtovanje človeških, finančnih ali drugih virov. Proces načrtovanja določi fiksne stroške informacij, ki nastanejo enako kot pri proizvodnih stroških, še preden shranimo ali uporabimo prvo informacijo.
- **Pridobitev:** poslovni procesi in aplikacije ustvarjajo podatke, ki se z njihovo uporabo pretvorijo v znanje in informacije.
- **Vzdrževanje:** procesi vzdrževanja informacij so sestavljeni iz poslovnih procesov, ki posodablajo oziroma ažurirajo informacije. Informacije, ki s časom ne odražajo več realnega stanja, so neakovostne, posledično pa povzročijo odpoved procesov, ki jih uporabljajo. Stroški vzdrževanja informacij neposredno ne povečujejo vrednosti informacijam, so pa nujne za preprečevanje odpovedi procesov, ki ustvarjajo vrednost. S tem znižamo stroške vzdrževanja informacij in informacijske infrastrukture.
- **Ukinitve:** ko organizacija informacij ne potrebuje več, jih je potrebno zbrisati. S tem sprostimo vire, ki so jih te informacije uporabljale. V nekaterih primerih brisanje informacij ni možno, lahko pa se informacije prestavijo na drug medij ipd.
- **Uporaba:** osnova za ustvarjanje vrednosti je uporaba vira. Dokler vir ni uporabljen, ima zgolj potencialno vrednost. Enako kot za vse vire, velja tudi za informacije: vrednost ustvarimo z uporabo informacij. Poslovni procesi, ki uporabljajo informacije in ustvarjajo vrednost so osnova za nastanek vrednosti informacij.

McGilvray (2008, str. 288) življenjski cikel informacij sicer deli v šest faz oziroma aktivnosti, ki sestavljajo akronim POSMAD (angl. *Plan – Obtain – Store and share – Maintain – Apply – Dispose*), ki pa se od definicije, ki jo podaja English (1999) razlikuje le v tem, da eksplicitno navaja dodatno fazo hranjenja in deljenja informacij, ki bi bila po Engliševi opredelitvi lahko razdeljena med procesa vzdrževanja in uporabe.

English (1999, str. 209) stroške, ki jih povzročijo neakovostne informacije deli v tri skupine:

- Stroški povezani z odpovedjo procesov: stroški nastanejo, ko so neakovostne informacije vzrok za nepravilno delovanje procesov. Primera:
 - Nenadomestljivi stroški, nastanejo pri pošiljanju drugega izvoda iste brošure kupcu zaradi podvojitve zapisa o kupcu kot stranki.
 - Stroški ponovne pridobitve kupcev ali odpravljanja težav med kupci in organizacijo, nastanejo kot posledica nezadovoljstva kupcev zaradi problemov s kakovostjo informacij (napačne cene, naslov za dostavo ipd.).
- Stroški zavračanja in ponovnega pridobivanja informacij: enako kot pri proizvodnji izdelkov je potrebno ukrepati pri informacijah z napako. Primeri stroškov povezanih z informacijami z napako:
 - Stroški rokovanja z podvojenimi informacijami
 - Stroški povezani z iskanjem manjkajočih informacij
 - Stroški povezani z ponovno izvedbo poslovnih procesov, ki se niso izvedli pravilno zaradi nezadovoljive kakovosti informacij
 - Stroški zaradi zmanjšanja produktivnosti
 - Stroški overjanja, uporabniki ne zaupajo informacijam
 - Stroški popravkov programske informacijske opreme
 - Stroški čiščenja in popravljanja informacij
- Stroški z naslova zgrešenih ali zamujenih priložnosti: gre za nerealizirane prihodke in dobiček zaradi slabe kakovosti informacij. Te stroške pa je potrebno analizirati na ravni donosnosti kupca skozi njegov celoten življenjski cikel v organizaciji.

Dokler se natančno ne določi višina teh stroškov in stroškov neakovostnih informacij na splošno, vodstvo organizacije lahko dojema te stroške kot običajne stroške poslovanja – dokler druge organizacije ne izrabijo priložnosti in dvignejo standarde s pomočjo visoko kakovostnih informacij (English, 1999, str. 235).

3 PRIMER BELEŽENJA IN UPORABE TRANSAKCIJSKIH PODATKOV

Praktičen primer izvira iz podjetja, ki sicer želi ostati neimenovano, je pa eno izmed vodilnih telekomunikacijskih podjetij v Sloveniji. Zaradi varovanja poslovnih skrivnosti so prikriti tudi nekateri podatki v grafih in tabelah. Podatki, ki so predmet analize, so zapisi o opravljenih storitvah, ki so jih opravili uporabniki storitev, ki imajo sklenjeno pogodbeno naročniško razmerje.

Ta primer analize transakcijskih podatkov je služil kot vzorčna analiza v organizaciji, tako z vidika izvedbe podatkovnega rudarjenja, kot tudi ocenjevanja kakovosti informacij. Transakcijski podatki, vir za analizo, so majhna podmnožica celotne zbirke informacij znotraj organizacije. Njihov primarni vir je sistem za zaračunavanje storitev, ki je bistvenega pomena za delovanje organizacije. Na podlagi vzorčne analize bom ocenil kakovost informacij, ki so služile kot vir, ter podal predloge za izboljšavo informacijske kakovosti.

3.1 Analiza transakcijskih podatkov

Primarno se transakcijski podatki o opravljenih storitvah uporabljajo kot osnova za zaračunavanje opravljenih storitev. V tem primeru pa gre za poskus uporabe teh podatkov v zbrani obliki za analizo obnašanja kupcev oziroma naročnikov. Kupci oziroma naročniki storitev niso anonimni, kar zagotavljata dve stvari: sklenjeno pogodbeno razmerje in ena ali več mobilnih telefonskih števil, ki so vezane na pogodbo. Uporabniki dovoljujejo uporabo transakcijskih podatkov za obdelavo, kar je pogoj za izvedbo tovrstne analize.

Cilj je bil iz obstoječih transakcijskih podatkov pridobiti naslednje:

- informacije za pomoč pri odločitvah o spremembi obstoječih, uvajanju novih in ukinitvi nekaterih obstoječih naročniških paketov,
- informacije za podporo pri izvedbi odločitev v procesu zadrževanja naročnikov,
- informacije, ki bi služile kot osnova za uvedbo ciljanega oglaševanja obstoječim naročnikom ter
- informacije za podporo pri procesu izdelave oglaševalskih akcij.

Temelj za pridobivanje tovrstnih informacij so podatki o uporabi storitev v kombinaciji z nekaterimi demografskimi podatki naročnikov.

3.1.1 Izbira metode in orodja za analizo

Z metodo grozdenja je mogoče doseči vse zastavljene cilje (analiza naročniških paketov, informacije za zadrževanje naročnikov, informacije za ciljano oglaševanje in izdelavo oglaševalskih akcij). Grozdenje je preprosta, naravna in celo samodejna človeška

operacija, kadar gre za razvrščanje objektov z relativno majhnim številom lastnosti. Z naraščanjem števila različnih lastnosti oziroma atributov pa narašča tudi kompleksnost problema, človek pa ga samostojno zaradi kompleksnosti ne more več rešiti. MacLennan et al. (2009, str. 291) menijo, da človek, strokovnjak, ki podrobno pozna podatke, lahko ustvari grozde pri pet do šest dimenzijah (atributih), medtem ko jih sodobne zbirke podatkov tipično vsebujejo precej več.

Problem je rešljiv z izbiro ustreznega orodja, v tem primeru sem izbral Microsoft Server 2009 zaradi enostavne uporabe, dobrih izkušenj z orodjem in kar je najpomembneje, zaradi popolne integracije z orodjem za podatkovno rudarjenje – Microsoft Business Intelligence Development Studio. Za grozdenje sem uporabil Microsoftovo implementacijo algoritma, Microsoft Clustering. Ena od alternativ bi bila izvedba grozdenja z odprotokodno programsko opremo Weka 3 (Weka, 2010), ki je na voljo popolnoma brezplačno za razliko od Microsoftovih orodij.

3.1.2 Izvor podatkov

Želene informacije je mogoče pridobiti iz transakcijskih podatkov, ki vsebujejo časovno dimenzijo, podatke o opravljeni storitvi in nekatere demografske podatke naročnika. Vsi podatki, zbrani na mesečnem nivoju, so se nahajali v podatkovnem skladišču, vanj pa so se prenašali iz sistema za obračunavanje storitev. Sistem za obračunavanje storitev beleži vsako opravljeno storitev glede na obračunsko enoto oziroma interval:

- Glasovni klic, osnova za obračun je čas trajanja.
- Kratko besedilno sporočilo (angl. *short message service* ali SMS), osnova za obračun je količina poslanih sporočil.
- Multimedijско sporočilo (angl. *multimedia message service* ali MMS), enako kot pri kratkih besedilnih sporočilih je osnova za obračun število poslanih sporočil.
- Podatkovni klic, pri podatkovnih klicih pa je osnova za obračun količina prenesenih podatkov.

Prenos podatkov iz podatkovnega skladišča v podatkovno bazo za analizo je bil mogoč le preko izvoza podatkov v datoteko ter naknadnega uvoza v podatkovno bazo.

3.1.2.1 Uvoz podatkov

Uvoz podatkov sem izvedel z v ta namen razvito preprosto aplikacijo. Podatki iz podatkovnega skladišča so bili namreč organizirani tako, da je vsak zapis vseboval podatke o porabi storitve za določen mesec v letu, tip storitve, telefonsko številko in identifikator uporabnika. Ta oblika zapisa je povsem primerna za podatkovno rudarjenje, izbrano orodje za izvedbo podatkovnega rudarjenja omogoča izvedbo algoritmov tudi na vgnezenih tabelah.

Zaradi zagotavljanja večje fleksibilnosti in učinkovitosti pri kasnejšem izvajanju poizvedb, sem vgnezdene tabele pretvoril v dimenzijske tabele, ki so na tabelo dejstev vezane preko ključev (Slika 3). Podatkovna shema v obliki zvezde je običajna za podatkovna skladiča, saj struktura zagotavlja učinkovitost poizvedb. Dimenzijski model opisuje proces s pomočjo dejstev in dimenzij. Pri tem so dejstva merila, ki opisujejo proces, dimenzije pa dejstva postavljajo v kontekst (Adamson, 2006, str. 6).

S pregledom podatkov v dimenzijskih tabelah je mogoče hitro in na enostaven način dobiti predstavo o naravi podatkov za analizo ter jih po potrebi preoblikovati ali izločiti. Takoj po uvozu sem izločil vse podatke, ki so bili vezani na naročniške pakete, ki obstajajo zgolj zaradi načina delovanja sistema za zaračunavanje storitev.

3.1.3 Priprava podatkov za analizo

Zaradi velike količine podatkov in preveč analitičnega opisa dogodkov, je bilo potrebno pred izvedbo podatkovnega rudarjenja podatke zbrati zgolj na letnem nivoju posameznega uporabniškega računa (telefonske številke) naročnika za leto 2009. Število zapisov za analizo je na ta način zmanjšano na 2.803.077.

Z izvedbo zbiranja zapisov na nivoju uporabniškega računa se je število zapisov za analizo zmanjšalo na obvladljivih (s trenutno informacijsko tehnologijo in infrastrukturo) 112.749 zapisov.

Hkrati z dodatnim zbiranjem podatkov sem tudi pretvoril zapise z zbranimi količinami opravljenih storitev v attribute. Za vsak tip storitve sem uvedel nove attribute: število klicev, količina klicev, število kratkih sporočil, število multimedijskih sporočil, število opravljenih podatkovnih klicev in količina prenesenih podatkov med podatkovnimi klici. Na ta način sem se izognil problemu gnezdenih tabel (angl. *nested table*), kjer več zapisov opisuje dogodke za isti uporabniški račun. Podatki o uporabi storitev za posamezen račun se sedaj nahajajo v le enem zapisu.

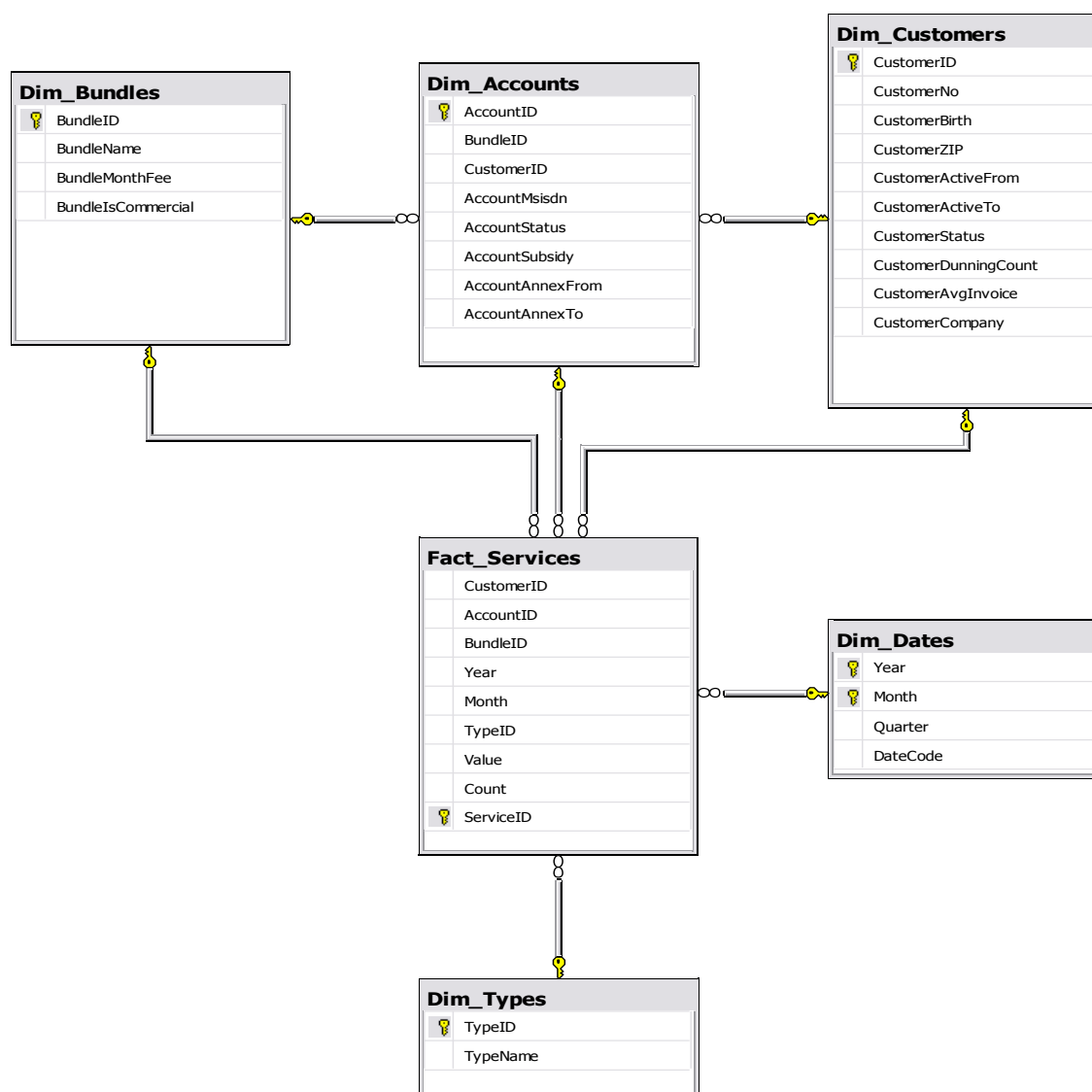
Zaradi zagotavljanja čim bolj primerljivih podatkov na mesečnem nivoju, namreč nekateri izmed naročnikov niso bili aktivni celotno leto 2009 – nekateri so naročniško razmerje sklenili po začetku leta, nekateri pa so naročniško razmerje prekinili pred iztekom leta 2009, je bilo potrebno povprečne vrednosti računati glede na dejansko število mesecev, v katerih so bili naročniki aktivni. Da je naročnik v izbranem mesecu aktiven sem sklepal v primeru, da je ta mesec opravil vsaj eno storitev glasovnega klica, saj se je podatek o dnevu prekinitve naročniškega razmerja izkazal za izjemno nezanesljivega.

3.1.4 Struktura podatkov za analizo

Podatki so shranjeni v podatkovni baze, shema podatkovne baze je v zvezdasti obliki (Slika 3). Tabela dejstev (Fact_Services) vsebuje vse podatke vezane na poslovni dogodek oziroma opravljeno storitev. Dimenzijske tabele pa vsebujejo naslednje podatke:

- Podatki o naročnikih (Dim_Customers), vsebuje podatke o nosilcih naročniških pogodb
- Podatki o uporabniških računih (Dim_Accounts), uporabniški račun je vezan na telefonsko številko naročnika. Naročnik ima lahko več uporabniških računov.
- Podatki o naročniških paketih (Dim_Bundles), vsak naročniški uporabniški račun je vezan na enega od naročniških paketov.
- Podatki časovne dimenzije (Dim_Dates), obdobje v katerem je bila storitev opravljena.
- Podatki o tipu storitve (Dim_Types), tipi storitev so: glasovni klic, kratko besedilno sporočilo, multimedijsko sporočilo in podatkovni klic.

Slika 3: Diagram podatkovne baze z zbranimi transakcijskimi podatki



Transakcijski podatki, so bili za analizo zbrani na mesečnem nivoju za vsak uporabniški račun oziroma telefonsko številko uporabnika, naknadno so bili dodatno zbrani na letnem nivoju. Atributi naročnikov so omejeni zgolj na rojstni datum in pošto številko ter indikator, ki pove, ali gre za fizično ali pravno osebo. V primeru, da gre za pravno osebo, naj bi bil datum rojstva prazen – atribut je v tem primeru brez pomena. Dimenzijska tabela naročnikov vsebuje tudi informacije o časovnem obdobju, v katerem je bil naročnik aktiven oziroma je imel vsaj eno veljavno pogodbo za naročniško razmerje. Vsak naročnik ima lahko več naročniških razmerij, ki so zapisana v obliki uporabniških računov.

Uporabniški račun pa vsebuje naslednje ključne attribute:

- Telefonsko številko uporabnika (angl. *Mobile Subscriber Integrated Services Digital Network Number*), ki jo uporabnik uporablja trenutno. Telefonska številka sama je prenosljiva, kar pomeni, da jo je uporabnik med trajanjem naročniškega razmerja lahko zamenjal. Informacije o pretekli rabi telefonskih števil v tem primeru niso bile zajete, prav tako manjka informacija o datumu zadnje spremembe telefonske številke.
- Višina zneska zadnje subvencije za nakup mobilnega aparata. Višina subvencije je odvisna od polne cene aparata in dolžine obdobja sklenitve aneksa k naročniški pogodbi. Podatki o polni ceni in tipu aparata niso zajeti.
- Datum sklenitve in prekinitve zadnjega aneksa, uporabniki aneks k naročniški pogodbi sklenejo zaradi pridobitve subvencije za nakup mobilnega aparata. Uporabniki, ki plačajo pono ceno aparata ne sklenejo aneksa k pogodbi, v tem primeru podatki o datum sklenitve in prekinitve ne smejo biti izpolnjeni.

Ostale dimenzijske tabele vsebujejo podatke o tipu sklenjenega paketa, tipu uporabljene storitve in časovnem obdobju. Pri dimenzijski tabeli za časovno obdobje sem poleg atributov, ki so bili med izvornimi podatki (mesec in leto) dodal še atribut časovne kode in atribut za oznako četrletja. Časovna koda je datumski zapis primeren za računalniško obdelavo, informacijo o četrletju pa dodal za enostavnejše izvajanje morebitnih dodatnih poizvedb v prihodnosti, ki bi bile vezane na četrletna obdobja.

3.1.5 Izvedba analize

Zbirni zapisi za posamezne uporabniške račune predstavljajo vir podatkov za podatkovno rudarjenje. Z zbiranjem in pretvorbo zapisov o opravljenih storitvah v posamezne attribute se kompleksnost izvedbe grozdenja zmanjšala, posledično pa se je zmanjšal tudi čas izvedbe. Hkrati pa je oblika zapisov, kjer so za posamezen uporabniški račun podatki zbrani v enem zapisu, bolj človeško berljiva.

Po večkratnem izvajanju algoritma grozdenja z zbranimi podatki naročnikov sem izločil vse attribute, ki so se izkazali za premalo pomembne pri delitvi naročnikov na skupine (na primer številka pošte ali regija, razlike med navadami porabe storitev naročnikov po

posameznih regijah so zanemarljive) ali pa so bili podatki premalo kakovostni, da bi jih bilo smiselno uporabiti (na primer rojstni datum naročnika in s tem starost).

Tabela 4: Vhodni atributi za izvedbo grozdenja

Vhodni atribut	Opomba
Višina subvencije	Subvencionirani znesek pri nakupu telefonskega aparata ob sklenitvi naročniškega razmerja
Povprečno število podatkovnih klicev	Povprečne količine enot porabljenih storitev
Povprečna količina prenesenih podatkov	
Povprečno število multimedijskih sporočil	
Povprečno število kratkih besedilnih sporočil	
Povprečno število govornih klicev	
Povprečna doba trajanja govornih klicev	

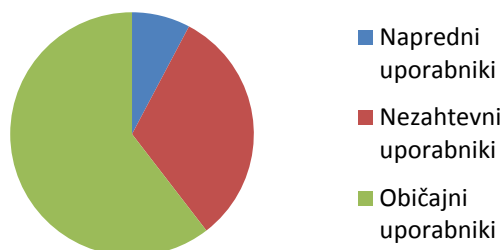
Zaradi večje razumljivosti sem število posameznih skupin, na katere algoritem grozdenja razdeli vzročne zapise, omejil na tri. Pri večjem številu skupin so si bile te preveč podobne, da bi bilo smiselno vztrajati na takšni delitvi.

Razumljivost delitve je v tem primeru pomembnejša od identifikacije zapisov z izjemnimi odkloni vrednosti. Zato je število skupin oziroma grozdov lahko zmanjšano na tri, prav tako je določena minimalna podpora primerov za grozd oziroma minimalno število zapisov, ki lahko tvorijo svoj grozd.

3.2 Rezultati analize in ugotovitve

Rezultat grozdenja vzorčnih podatkov so tri skupine uporabnikov, čeprav gre natančneje za skupine uporabniških računov. Naročnik ima v veliki večini zgolj en uporabniški račun oziroma mobilno telefonsko številko, tudi v primeru, da imajo več računov pa še vedno posamezne telefonske številke lahko uporabljajo različni uporabniki. Zato sem predpostavljal, da je osnovni zapis za analizo, prav uporabniški račun, in ne nosilec naročniškega razmera.

Slika 4: Razdelitev populacije uporabnikov glede na uporabo storitev



Opomba: število posameznih članov skupin zaradi zaupnosti informacij ni prikazano.

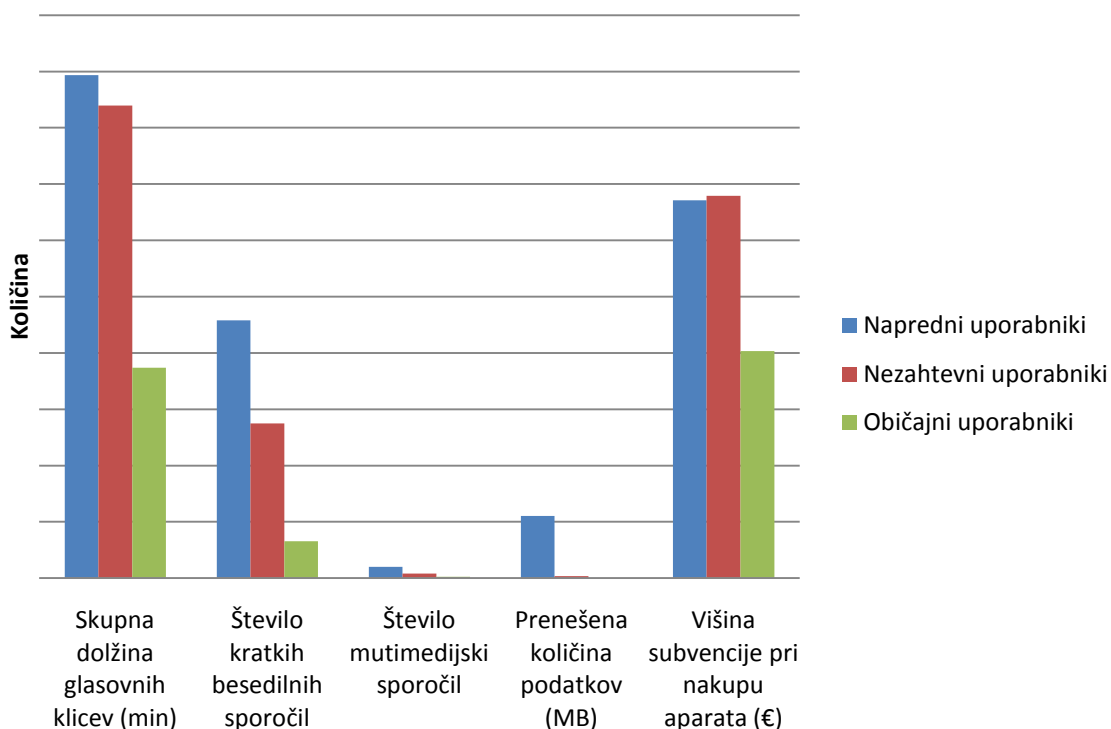
Skupine uporabnikov so privzeto poimenovanje zgolj s številkami od 1 do 3. Zaradi lažjega razlikovanja med skupinami sem jih preimenoval v skupine naprednih, nezahtevnih in običajnih uporabnikov. Kimbal, Ross, Thorntwaite, Mundy in Becker (2006, str. 496) navajajo primer v podjetja Claritas, ki je uporabnike s pomočjo grozdenja razdelilo v več kot šestdeset različnih skupin, nato pa so problem razlikovanja med njimi rešili z poimenovanjem tipičnih predstavnikov posameznih skupin. Nekaj tipičnih imen skupin, ki služijo za učinkovito razlikovanje med seboj:

- 02: Posestva modre krvi; star denar, velika posestva (angl. *Blue Blood Estates*)
- 51: Puške in poltovornjaki; delavski razred, velike družine, mobilni domovi (angl. *Shotguns and Pickups*)
- 60: Starci s parka; skromen prihodek, umirjeni, čez dan gledajo televizijo (angl. *Park Bench Seniors*)

Uporabniki so razdeljeni na tri skupine (običajni, nezahtevni in napredni uporabniki), katerih člani imajo naslednje lastnosti:

- Običajni uporabniki: najštevilčnejša skupina uporabnikov, ki so najmanj zahtevni uporabniki v smislu uporabe storitev. Poleg glasovnih klicev pošljejo tudi nekaj kratkih besedilnih sporočil, naprednejših storitev (multimedijska sporočila in podatkovni klici) pa skorajda ne uporabljajo.
- Nezahtevni uporabniki: opravijo veliko glasovnih klic ter pošljejo precej kratkih besedilnih sporočil, uporaba storitev pošiljanja multimedijskih sporočil in podatkovnih klicev je minimalna.
- Napredni uporabniki: uporabljajo največ storitev tako po količini kot tipu. Skupina uporabnikov, ki za razliko od drugih dveh skupin uporablja podatkovne klice in pošilja več multimedijskih sporočil. So številčno najmanj zastopana populacija.

Slika 5: Primerjava uporabe storitev uporabnikov treh grozdov



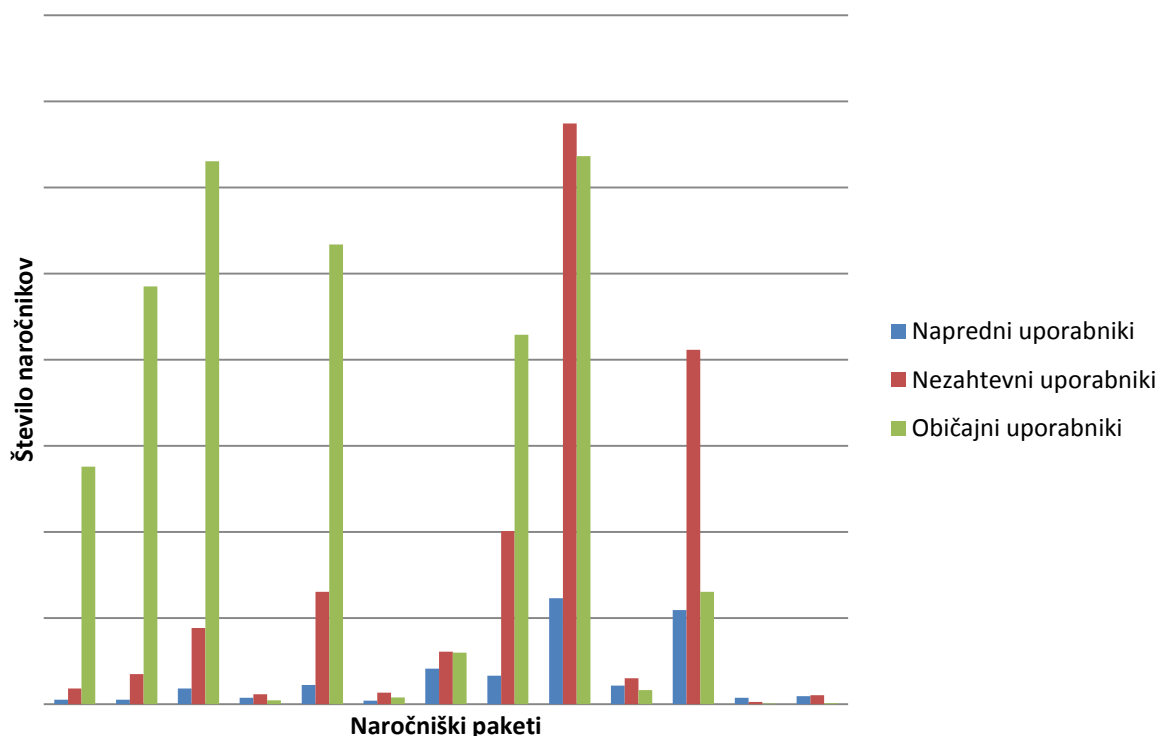
Opomba: zaradi zaupnosti informacij so oznake na ordinatni osi odstranjene.

Rezultati analize so skoraj v celoti dosegli cilje zastavljene pred začetkom izvedbe analize. Rezultati kažejo na to, da so uporabljeni transakcijski podatki primerni za izdelavo tovrstnih analiz, kar bom podrobje prdstavil v nadaljevanju.

3.2.1 Ustreznost paketa naročnikovim potrebam

Ali imajo naročniki tudi naročniški paket, ki jim ustreza, je mogoče le posredno ugotoviti s primerjavo popularnosti uporabe obstoječih naročniških paketov in profila grozdov. Na Sliki 6 je prikazana popularnost posameznih obstoječih naročniških paketov med naročniki, ki so razdeljeni na skupine po zahtevnosti (običajni, nezahtevni in napredni uporabniki). Cena in količina v naročnino že vključenih naprednih storitev narašča iz leve proti desni. Skrajno desno so paketi, ki že vključujejo relativno veliko količino tako glasovnih klicev kot tudi naprednejših storitev, med njimi pa presenetljivo ni veliko naprednih uporabnikov, saj so paketi v prvi vrsti namenjeni prav njim.

Slika 6: Porazdelitev obstoječih naročniških paketov med grozde



Opomba: Zaradi zaupnosti informacij so odstranjena imena naročniških paketov na abscisni osi in oznake na ordinatni osi.

Glede na obstoječo popularnost paketov in dejansko uporabo storitev med skupinami uporabnikov je mogoče sklepati, da zahtevni in delno tudi nezahtevni uporabniki ne uporabljajo naročniških paketov, ki so jim bili namenjeni. Razvidna so tudi odstopanja pri strukturi naročnikov za posamezne pakete.

Razlogi in vzroki iz te analize niso razvidni, sama kakovost obstoječih podatkov pa v trenutku nastanka analize ni omogočala natančnejšega vpogleda v na primer demografsko strukturo skupin uporabnikov, ki bi lahko vplivala na spremembo v strategiji trženja naročniških paketov.

3.2.2 Program zadrževanja naročnikov in ciljano oglaševanje

Glede na to, da je analiza uporabnikov pokazala možno delitev obstoječih uporabnikov na tri skupine, bi bilo ugotovitve mogoče uporabiti v programu zadrževanja naročnikov. Naročnikom, ki se pogodbeno naročniško razmerje izteka bi bilo mogoče ponuditi njihovi dosedanji porabi prilagojeno ponudbo, namesto običajne ponudbe enake za vse.

Naročniki so v konkretnem primeru privolili v uporabo transakcijskih podatkov v namene ciljanega oglaševanja, kar je nujno glede na Zakon o elektronskih komunikacijah (Ur.l. RS,

št. 13/2007-UPB1). Ta narekuje, da se ciljano oglaševanje na podlagi transakcijskih podatkov lahko izvaja le s predhodnim soglasjem uporabnika.

Tudi sicer so uporabniki čedalje bolj pripravljeni del zasebnosti zamenjati za prilagojeno obliko oglaševanja, ki ustreza njihovemu življenjskemu cilju. To potrjuje tudi raziskava podjetja IBM iz leta 2008, ki je zajemala 2.800 ljudi iz šestih držav (IBM Press room, 2010).

3.2.3 Dejanska uporaba rezultatov analize

Kljub potencialno uporabnim rezultatom in ugotovitvam analize, so bili le-ti s strani vodstva organizacije popolnoma prezrti. Prav tako s strani vodstva nisem prejel nobenih odgovorov ali pojasnil. Dejansko se ni uporabil noben del rezultatov dobljenih s to analizo, prav tako so prezrte tudi vse ugotovitve in predlogi vezani na kakovost informacij.

Ker nisem uspel dobiti konkretnih odgovorov s strani vodstva, sem možne odgovore za (ne)reakcijo poiskal v literaturi. Berkun (2009) navaja možne razloge, ki so vezani na sprejem nove ideje. Peter in Hull (1994) pa navajata razloge, ki so vezani na postopno odpoved učinkovitosti dela v hierarhičnih organizacijah.

3.2.3.1 Sprejem novih idej

Ideja o uporabi rezultatov in zagotavljanju kakovost informacij, ki sega čez njihovo primarno uporabo, znotraj organizacije ni bila sprejeta. Berkun (2009, str. 59) navaja seznam pogledov na nove ideje, ki so v večini primerov krivi za negativen sprejem nove ideje:

- ego/zavidanje: ideje ne sprejemam, ker ni moja
- ponos in politika: sprejem ideje slabo vpliva na mojo podobo v organizaciji
- strah: strah me je sprememb, ki jih ideja prinaša
- prioriteta: imam veliko inovativnih idej, virov pa le dovolj za izvedbo ene
- lenivec: sem len, in nočem razmišljati oziroma opravljati več dela kot je to nujno potrebno
- varnost: s sprejemom ideje lahko izgubim, kar ne želim izgubiti
- pohlep: z zavrnitvijo ideje lahko zaslužim
- usklajenost: ideja krši moja prepričanja (ne glede na njihovo smiselnost)

Posledica teh pogledov na sprejem nove ideje, ne glede na to ali so upravičeni ali ne, je enaka. Berkun (2009, str. 60) kot primer zavrnitve ideje glede na zgoraj navedene poglede navaja trditev Galileja in njegovo trditev, da je sonce središče sončnega sistema. Galilejo je bil deležen obtožb Cerkve in zahodnega sveta zaradi enakih razlogov.

3.2.3.2 Kvocient zrelosti hierarhije

Na podlagi analize stotine primerov strokovne nesposobnosti je Dr. Lawrence J. Peter formuliral pravilo, ki je tudi imenovano po njemu (angl. *Peter's principle*): v hierarhični organizaciji vsak zaposleni napreduje do svojega nivoja inkompetence (Peter & Hull, 1994, str. 25).

Pogoj za izpolnitev pravila je zadostno število mest v hierarhiji in dovolj časa. Peter (1994) namreč meni, da zaposleni napredujejo na višje delovno mesto v organizaciji s tem, da pokažejo strokovno sposobnost na trenutnem delovnem mestu in ne zaradi sposobnosti opravljaja dela na delovnem mestu na katerega napredujejo. Z napredovanjem na delovno mesto, ki presega posemeznikove strokovne sposobnosti se napredovanje ustavi.

Koeficient zrelosti hierarhije, v enačbi (1) je označen kot MQ (angl. *maturity quotient*), je mera za neučinkovitost organizacije, definiran pa je kot koeficient med številom zaposlenih na nivoju inkompetence in skupnim številom zaposlenih v hierarhiji.

$$MQ = \frac{\text{Št.zaposleni } h \text{ v organizaciji na nivoju inkompetence}}{\text{Število vseh zaposleni } h} * 100 \quad (1)$$

Ko koeficient zrelosti doseže vrednost 100, organizacija ne opravlja več dela. Ocena koeficienta zrelosti v konkretni organizaciji presega okvire tega magistrskega dela. Vkolikor bi se ob podrobni analizi izkazalo, da je koeficient relativno visok, bi to lahko bil eden od možnih vzrokov za preztje rezultatov analize.

4 ANALIZA UGOTOVLJENEGA STANJA

Analiza transakcijskih podatkov je bila vzorčni primer, ki je služil za ocenitev stanja informacijske kakovosti. V nadaljevanju bom podal oceno trenutnega stanja kakovosti informacij, ki so služile za analizo. Poiskal sem vzroke in omejitve, ki so najbolj pripomogli k nekovosti in podal predloge za odpravo pomanjkljivosti ter izboljšave.

4.1 Ocena kakovosti vira podatkov

Kakovost vhodnih podatkov, ki so bili osnova za izvedbo analize je pomanjkljiva, kritična je predvsem na nivoju proizvoda (usklajenost in pravilnost). Da slaba kakovost ni bila zaznana že prej, pa gre pripisati predvsem dejstvu, da transakcijski podatki zadovoljivo služijo svojemu prvotnemu namenu - na njihovi osnovi se naročnikom zaračunavajo storitve.

Zaradi popolne ločenosti sistema za zaračunavanje od sistema za management odnosov z odjemalci, pa se je pojavil tudi problem neuskklajenosti informacij. Vira informacij nista integrirana, zato je potreben dodaten napor za pridobitev informacij z obeh strani.

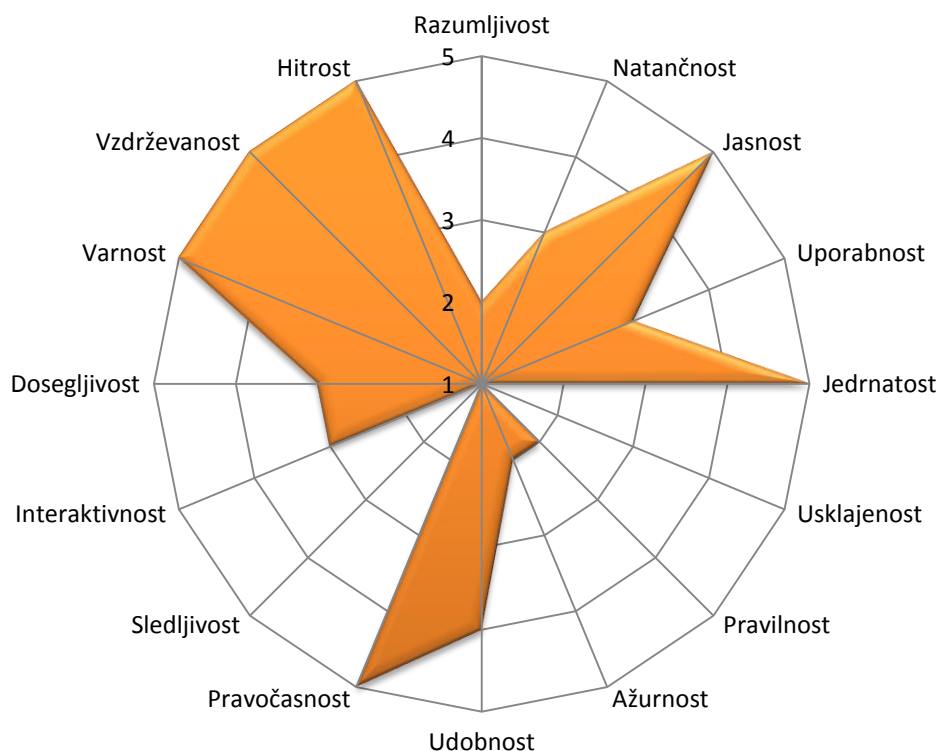
Še en ključni problem pa izhaja iz samega načina uporabe sistema za zaračunavanje, ob prekinitvi naročniškega razmerja se vsi podatki naročnika brišejo iz sistema. Razlog je v sistemu zaračunavanja licenčnin za programsko opremo. Prekinitev naročniškega razmerja je sicer zavedena v sistemu management odnosov z odjemalci, a zaradi neobstoja integracije do teh informacij ni enostavno dostopati.

Stanje kakovosti informacij, ki so služili kot vir za izvedeno analizo, sem ocenil glede na izbranih šestnajst kriterijev informacijske kakovosti (Eppler, 2006). Pregled in kratka utemeljitev ocene se nahaja v Tabeli 5, podrobnosti pa so opisane v nadaljevanju.

Tabela 5: Ocena kakovosti informacij

Nivo	Kriterij	Ocena	Opomba
Nivo skupnosti (relevantnost)	Razumljivost	2	Obseg podatkov je pomanjkljiv. Manjkajo tako dodatne informacije k obstoječim atributom kot demografski podatki, podatki o interakcij z naročniki, ipd.
	Natančnost	3	Informacije, ki so osnova za zaračunavanje so zapisane dovolj točno za potrebe analize.
	Jasnost	5	Informacije so namenjene računalniški obdelavi.
	Uporabnost	3	Potrebno je predhodno čiščenje.
Nivo proizvoda (celovitost)	Jedrnatost	5	Ni nepotrebnih elementov.
	Usklajenost	1	Informacije so popolnoma neusklajene razen informacij za zaračunavanje storitev.
	Pravilnost	2	Večina informacij je močno popačenih. Izjema so informacije o porabi storitev.
	Ažurnost	2	Informacije vezane na prekinitve naročniški razmerij niso posodobljene kljub obdelavi podatkov za predhodno poslovno leto.
Nivo procesa	Udobnost	4	Priprava podatkov se izvede na zahtevo.
	Pravočasnost	5	Informacije so za preteklo obdobje, hitrost dostave ni kritična.
	Sledljivost	1	Informacij o nastanku zapisov ali samem viru ni.
	Interaktivnost	3	Uporabnik nima vpliva na proces pridobitve informacij, vendar to ni kritično.
Nivo infrastrukture	Dosegljivost	3	Dostop je omejen na pooblaščen osebe, ki imajo za dostop do informacij tudi pisno pooblastilo.
	Varnost	5	Varovan je dostop do informacij, vzpostavljen je tudi sistem za samodejno izdelavo varnostnih kopij.
	Vzdrževanost	5	Informacije so v hranjene v digitalni obliki v sistemu, ki zagotavlja nemoteno uporabo v prihodnosti.
	Hitrost	5	Odziv sistema je kos nalogam v realnem času.

Slika 7: *Kakovost vhodnih informacij*



4.1.1 Razumljivost

Problem nezadostne razumljivosti je vezan predvsem na attribute, ki nepopolno opisujejo dejansko stanje na primer obdobje veljavnosti ankesa k naročniški pogodbi. Pomen je mogoče napačno interpretirati, ker gre v resnici le obdobje zadnjega sklenjenga aneksa. V primeru, da je naročnik imel tudi pred tem obdobjem sklenjen aneks, vezan na isti uporabniški račun, je ta informacija izpuščena.

Atribut, ki predstavlja ključ za povezavo uporabniškega računa z drugimi sistemi je telefonska številka, ki brez obdobja veljava oziroma točnega datuma zajema, ni zanesljiv podatek. Telefonska številka je namreč prenosljiva in brez informacij, kdaj je bil ta podatek zajet, je uparjanje podatkov na podlagi telefonske številke lahko napačno.

Obseg informacij pa bi bil lahko tudi razširjen, vendar le na račun jedrnatosti. Glavni težo za slabo oceno na področju razumljivosti informacij nedvomno nosijo nepopolne informacije.

Ocena: 2/5

4.1.2 Natančnost

Pri kriteriju natančnosti oziroma točnosti gre za stopnjo ujemanja z merilom ali dejanskim stanjem (Eppler, 2006, str. 77). Vse informacije, ki so vezane na porabo

storitev, natančno opisujejo dejansko stanje. Izračunane skupne vrednosti in povprečja pa so predstavljene dovolj natančno za potrebe analize. Ostali podatki so zapisani v obliki, ki verno predstavlja dejansko stanje in jih je tako mogoče tudi interpretirati.

Pri podatkih, ki pa ne izvirajo iz sistema za zaračunavanje storitev, pa bi zaradi neusklajenosti lahko rekel, da gre za popolnoma napačne podatke. Razlog za napake ni v nenatančni predstavitvi ali zapisu informacij, ampak v neusklajenosti z drugimi viri informacij.

Ocena: 3/5

4.1.3 Jasnost

Informacije so namenjene nadaljni računalniški obdelavi, kot takšne pa so popolnoma jasne in razumljive ciljni skupini – računalniku. Format datoteke je popolnoma prilagojen računalniški obdelavi in ni namenjen branju s strani človeka.

Ocena: 5/5

4.1.4 Uporabnost

Informacije niso neposredno uporabne, pred uporabo jih je potrebno uvoziti v sistem za analizo. Še pred izvedbo analize pa je potrebno izločiti vse zapise, ki za analizo niso pomembni, prav tako pa je potrebno izvesti zbiranje zapisov na nivoju, ki je uporaben za analizo.

Ocena: 3/5

4.1.5 Jedrnatost

Informacije so očiščene vseh nepotrebnih elementov, kar pomeni da so jedrnate. Ker pa je na račun jedrnatosti okrnjen osbeg informacij, ki so na voljo za analizo, trpi ocena razumljivosti. Visoka stopnja jedrnatosti bi lahko vplivala tudi na slabšo točnost informacij, vendar to v tem primeru ni problem.

Ocena: 5/5

4.1.6 Usklajenost

Informacije so popolnoma neusklajene z drugimi viri. Gre za to, da zaradi enostavnosti izvirajo iz enega samega vira podatkov: sistema za zaračunavanje storitev. Posledica pa je, da so informacije, ki niso neposredno vezane na problematiko zaračunavanja storitev, popolnoma neusklajene. Podatki o uporabnikih niso usklajeni s sistemom za management odnosov z odjemalci, tako da so informacije vezane na osebne podatke naročnikov precej manj točne, kot bi bile lahko sicer.

Zaradi tega, ker usklajenost ni bila izrecno zahtevana, so nezanesljivi tudi atributi, ki predstavljajo ključne za upravljanje zapisov različnih virov. Prenosljiva telefonska številka, ki je ključ za uporabniški račun je tak primer. To pa onemogoča uparjanje podatkov z drugimi in s tem usklajevanje podatkov in rezultatov v prihodnosti.

Ocena: 1/5

4.1.7 Pravilnost

Pravilnost informacij je pri podatkih za analizo podedovana iz njihovega vira, podatkovnega skladišča. Informacije vezane na uporabo storitev so pravilne. Informacije o prekinitev aneksov naročnikov oziroma informacije o prekinitvi aktivnosti med naročnikom in organizacijo, pa so izračunane povsem posredno. Prekinitev aktivnosti se namreč v podatkovno skladišče zabeleži v primeru, da je naročnik odstranjen iz sistema za zaračunavanje. Na ta način pridobljen podatek je v resnici napačen. Pogoji, da je zabeležen datum prekinitve naročniškega razmerja za naročnika, ki je naročniško razmerje v resnici prekinil je, da je odstranjen iz sistema za zaračunavanje. Nato se ob prenosu podatkov iz sistema za zaračunavanje v podatkovno skladišče kot datum prekinitve naročniškega razmerja zabeleži datum izvedbe obdelave, ta obdelava pa se izvede le na zahtevo in ne ob vnaprej določenih rednih intervalih.

Na tak način zabeležena informacija o datumu prekinitve naročniškega razmerja je napačna, pravilna je le posredna informacija, da je naročnik razmerje prekinil do datuma, ki je naveden kot datum prekinitve naročniškega razmerja.

Ocena: 2/5

4.1.8 Ažurnost

Podatki za analizo so vezani na storitve naročnikov, izvedene prejšnjem letu (2009). Kljub pomankljivostim pa ti podatki še vedno niso ažurni, oziroma ne odražajo dejanskega stanja, ki je veljalo za leto 2009. Dokler viri podatkov niso integrirani in usklajeni, podatki za analizo v celotni ne bodo odražali pravega stanja.

Ocena: 2/5

4.1.9 Udobnost

Pridobivanje podatkov za analizo poteka na zahtevo, kar ustreza trenutnim potrebam. Zaradi varovanja informacij pa je pred tem potrebno pridobiti odobreno zahtevo za vpogled in obdelavo podatkov. Gre za postopek, ki je skladen z Zakonom o varovanju osebnih podatkov (Ur.l. RS, št. 94/2007-UPB1) in Zakonom o elektronskih komunikacijah (Ur.l. RS, št. 13/2007-UPB1). Udobnost je v tem primeru na račun varnosti zmanjšana.

Ocena: 4/5

4.1.10 Pravočasnost

Podatki za analizo so za preteklo poslovno leto in pravočasnost dostave ni kritična, prav tako je zamik odstave popolnoma znotraj zahtevanih in pričakovanih časovnih okvirjev.

Ocena: 5/5

4.1.11 Sledljivost

Z beleženjem vsakega koraka pri pridobivanju informacij in njihovi uporabi skupaj z razlogi in namenom uporabe, je mogoče rekonstruirati celoten proces obdelave informacij v njihovem celotnem življenjskem ciklu. Sledljivost je tako predpogoj za kredibilnost informacij. Omogoča bolj celovito interpretacijo informacij, ker je mogoče razkriti vir in metodologijo, ki je bila uporabljena pri obdelavi (Eppler, 2006, str. 77).

Podatki za analizo ne vsebujejo nobenega elementa sledljivosti, vir podatkov tudi v podatkovnem skladišču ni eksplicitno naveden. Potrebno je zaupanje, da so se podatki v nespremenjeni obliki prenesli iz sistema za zaračunavanje. Pri prenosu pa so informacije očiščene tudi celotne zgodovine sprememb, kar negativno vpliva na sledljivost, hkrati pa negativno vpliva tudi na celovitost informacij.

Edini element sledljivost pa je papirna sled, za obdelavo podatkov je potrebno imeti odobreno pisno dovoljenje z izraženim namenom obdelave. Vendar pa to ne vpliva na zagotavljanje sledljivosti nastanka in sprememb informacij.

Ocena: 1/5

4.1.12 Interaktivnost

Podatki so uporabniku dostavljeni v obliki, ki ni človeško berljiva ampak je namenjena računalniški obdelavi. Uporabnik nesporedno ne more vplivati na prikaz informacij ali jih prilagoditi svojim potrebam. Ne glede na to, da je prilagodljivost na ta način skoraj popolnoma okrnjena, pa to nima posebnega vpliva na uporabo podatkov.

Ocena: 3/5

4.1.13 Dosegljivost

Dosegljivost je na račun varnosti zmanjšana. Za dostop do informacij je potrebno pisno pooblastilo z izraženim namenom. Obdelava za pripravo podatkov je nato sprožena ročno, na zahtevo.

Ocena: 3/5

4.1.14 Varnost

Dostop do informacij je varovan z uporabniški imeni in gesli za dostop, prav tako je kontrolirana tudi izdaja uporabniških imen in gesel za dostop. Dodatno je za pridobitev informacij potrebno pisno dovoljenje. Vzpostavljen je sistem za samodejno izdelavo varnostnih kopij. Informacij so varovane tako z vidika dostopa do njih, kot tudi pred izgubo, kar pa po drugi strani zmanjšuje udobnost dostopa do njih.

Ocena: 5/5

4.1.15 Vzdrževanost

Informacije se hranijo v digitalni obliki, kar omogoča varno in zanesljivo hrambo ter njihovo vzdrževanje. Glede na trend povečevanja kapacitete za hranjenje digitalnih podatkov ob čedalje nižjih stroških, bo informacije ob zanemarljivo nizkih mejnih stroških mogoče hraniti tudi v prihodnosti.

Ocena: 5/5

4.1.16 Hitrost

Hitrost sistema za dostavo podatkov popolnoma zadošča potrebam, upoštevati pa je treba, da je šlo za zbrane transakcijske podatke. Število zapisov je bilo na ta način precej zmanjšano. V primeru, da bi šlo za obdelavo transakcijskih podatkov na nivoju posamezne transakcije, pa bi se lahko pokazale pomankljivosti v sistemu dela na principu izvoza in uvoza podatkov.

Ocena: 5/5

4.2 Vzroki in posledice nekovosti informacij

Ugotovljena nekovost informacij, ki so služile za vzorčno analizo, pomeni, da je z vidika kakovosti problematičen tudi sam vir informacij – podatkovno skladišče. Informacije si namreč nasprotujejo z informacijami, ki se nahajajo v sistemu za management odnosov z odjemalci. Informacije, ki se hranijo v podatkovnem skladiču, pa so osnova za izdelavo različnih poročil in vir informacij, na podlagi katerih se izvaja program za zadrževanje naročnikov.

V nadaljevanju bom opisal bistvene vzroke za slabo kakovost informacij ter neposredne posledice.

4.2.1 Uporaba in pomen atributov izvornih podatkov

Pri atributih izvornih podatkov so pomanjkljivosti, med katerimi nekatere vplivajo neposredno na kakovost izvedbe analize, druge pa na možnost uporabe rezultatov.

Tipični atributi, ki ne ponujajo celovitih informacij, so tisti, ki vsebujejo zgolj zadnjo vrednost atributa, kot so telefonska številka uporabnika, datum sklenitve aneksa, višina subvencije ipd. Pri datumu sklenitve aneksa in višini subvencije na primer, ni problem zgolj v pomenu podatka, saj je naročnik aneks lahko podaljšal večkrat, s tem pa se mu je spremenil tudi skupni subvencionirani znesek za nakup aparatov. Atributi vsebujejo zgolj zadnje vrednosti, kar lahko pomeni, da odražajo drugačno stanje od dejanskega stanja. Samo odstopanje od dejanskega stanja pa je mogoče ugotoviti zgolj z dodatno analizo, ali pa je podatke potrebno izvoziti iz podatkovnega skladišča v dopolnjeni obliki, če se tam ta informacija tudi nahaja.

Pri atributu, ki vsebuje telefonsko številko uporabnika, pa je problem precej večji. Namreč, pri izvedbi grozdenja uporabnikov na podlagi uporabniških računov, ki jih enolično določa telefonska številka, so rezultati praktično neuporabni zaradi prenosljivosti telefonske številke. V času po izvedbi analize je uporabnik lahko zamenjal telefonsko številko, v tem primeru bo podatek, ki je bil v analizi uporabljen, popolnoma neuporaben. Pri uparjanju podatkov iz grozdov in podatki iz sistema za management odnosov z odjemalci, bi prišlo do napačnega uparjanja za uporabnike, ki so telefonsko številko zamenjali, ali pa so že uporabljeno številko dobili kot novo. Več časa kot preteče od izvedbe analize do uporabe podatkov, več zapisov bi bilo lahko potencialno neuporabnih.

4.2.2 Vpliv informacij o prekinitvah naročniških razmerij

Bistveni del življenjskega cikla uporabnika je sklenitev samega naročniškega razmerja, aktivacija uporabnika v sistemu za zaračunavanje, s čimer je stranki omogočena uporaba storitev ter sklenitev aneksa za subvencioniranje telefonskega aparata. Bodisi ob sklenitvi naročniškega razmerja ali naknadno, lahko naročnik sklene aneks k pogodbi za subvencioniranje telefonskega aparata, aneks se sklene za obdobje 12 ali 24 mesecev. Naročniško razmerje je prekinjeno, ko je prekinjena pogodba o uporabi storitev med organizacijo in uporabnikom.

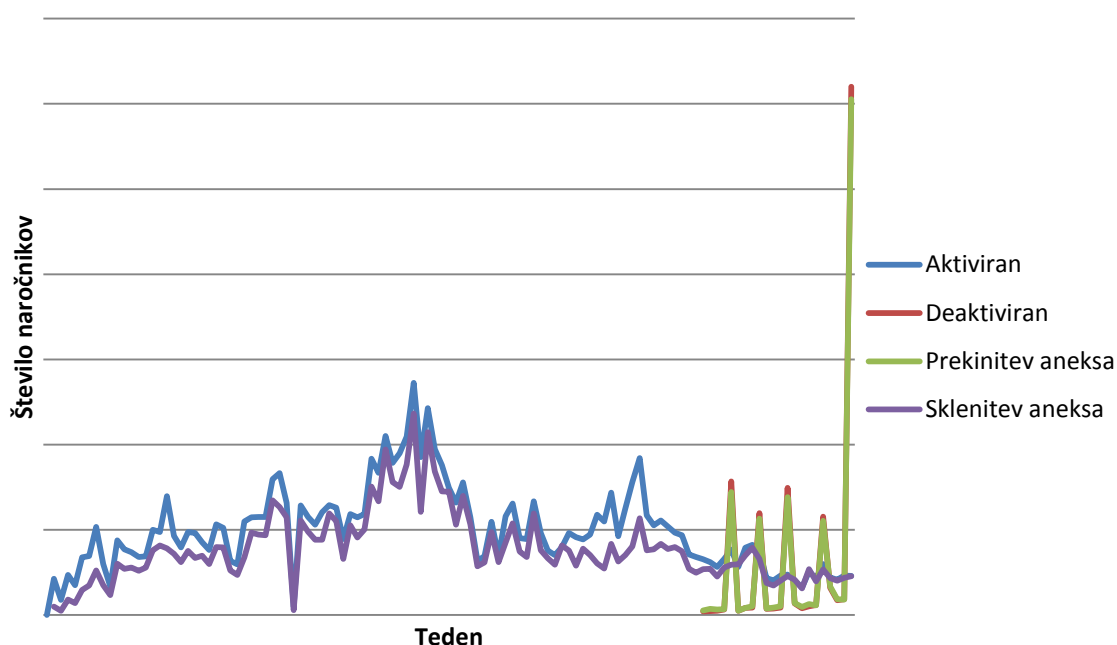
Ob prekinitvi naročniškega razmerja se podatki naročnika brišejo iz sistema za zaračunavanje (še vedno ostanejo v sistemu za management odnosov z odjemalci). Pri prenosu podatkov iz sistema za zaračunavanje v podatkovno skladišče, bi se zato podatki o naročnikih, ki so naročniško razmerje prekinili, prenehali ažurirati. Točnega datuma prekinitve naročniškega razmerja na ta način ni več mogoče zapisati v podatkovno skladišče. Ta problem je bil rešen tako, da obdelava, ki prenaša podatke iz sistema za zaračunavanje v podatkovno skladišče, ob ugotovitvi, da podatkov naročnika ni več v sistemu za zaračunavanje, v podatkovnem skladišču vpiše datum izvedbe obdelave kot datum prekinitve naročniškega razmerja. Rešitev ima očitne pomanjkljivosti:

- Obdelava se ne izvaja dovolj pogosto, da bi dovolj ažurno zaznavala prekinitve naročniških razmerij

- Podatki o naročnikih se ne odstranjujejo dovolj pogosto in v rednih intervalih iz sistema za zaračunavanje, da bi bila dogodka (prekinitev razmerja in odstranitev zapisa) zanesljivo povezana
- Spremembe na podatkih v podatkovnem skladišču na ta način niso zabeležene

Popravki podatkov so razvidni iz Slike 8, ki prikazuje število sklenitev in prekinitev naročniških razmerij in aneksov po tednih, ter število vklopov in izklopov uporabnikov na teden iz omrežja za obdobje med 2007 in 2009. Grafa sklenitev naročniških razmerij in vklopov uporabnikov v sistem sta pravilna, prekinitve razmerij in izklopi uporabnikov pa ne odražajo dejanskega stanja, ampak se pojavijo šele v zadnjem delu leta 2009.

Slika 8: Sklenitve in prekinitve naročniških razmerij ter aneksov na tedenski ravni za naročnike, aktivne med 2007 in 2009



Opomba: zaradi zaupnosti informacij so odstranjene oznake na abscisni in ordinatni osi.

Slaba kakovost informacij bi lahko na dolgi rok povzročila še veliko problemov, ki pa do sedaj niso bili zaznani. Podatki so hkrati vir za analizo, katere rezultati služijo kot pomoč obstoječemu programu zadrževanja naročnikov. Nezaanesljivost informacij o prekinitvah naročniških razmerij tu nima vpliva, saj so v program zadrževanja vključeni samo naročniki, ki imajo še vedno sklenjeno pogodbeno razmerje. Je pa na ta način onemogočena na primer analiza naročnikov, ki so naročniško razmerje prekinili in podobno.

4.2.3 Vpliv (ne)integracije na kakovost informacij

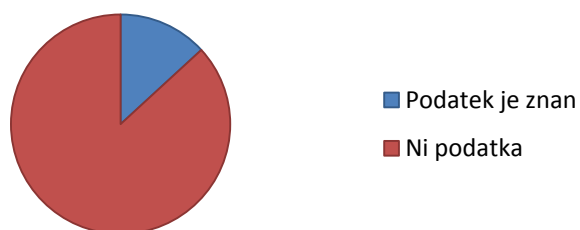
Podatki za analizo izvirajo iz sistema za zaračunavanje, kjer se nahajajo točne informacije o porabi storitev. Podatki o naročnikih, kot so na primer demografski podatki, podatki o pritožbah in reklamacijah, pa se nahajajo v sistemu za management odnosov z odjemalci. Podatki iz obeh sistemov se prenašajo v podatkovno skladišče, kjer pa je povezava virov zaradi neustrezne strukture onemogočena zaradi povezave na skupni točki - številke naročnika. V podatkovnem skladišču je uporaba storitev vezana na telefonsko številko uporabnika, s tem pa nastane problem pri naknadnem uparjanju zbranih transakcijskih podatkov iz podatkovnega skladišča s podatki iz sistema za management odnosov z odjemalci.

Zaradi neupoštevanja prenosljivosti številke tako pride do problemov pri povezavah različnih sistemov, telefonska številka namreč ne določa enolično uporabniškega računa, zato je kot ključ za povezavo neuporabna.

Zaradi neobstoječe integracije na točki preden podatki vstopijo v podatkovno skladišče se posledično močno zniža kakovost informacij v podatkovnem skladišču, saj so ključne informacije za uparjanje izgubljene.

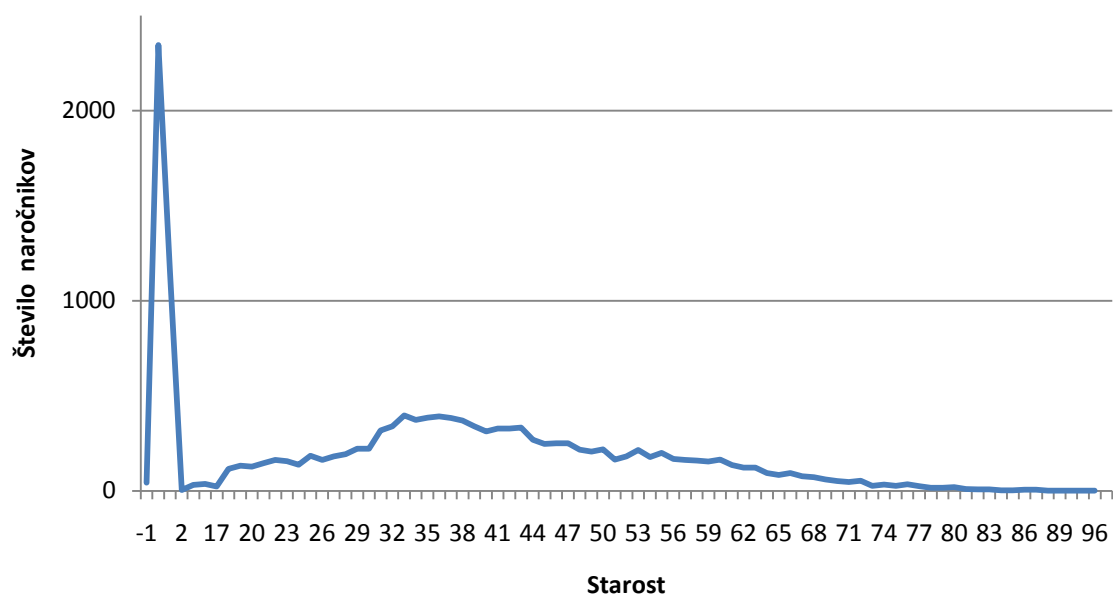
Za ilustracijo problema sem vzel informacije o starosti naročnikov. Ker podatek o rojstnem dnevu naročnika ni obvezen na strani vira podatkov, je starost mogoče določiti le za približno petnajst odstotkov naročnikov. Odstotek velja za podatke, ki so predmet analize in ne za celotno bazo naročnikov.

Slika 9: Razmerje med znano in neznano starostjo naročnikov



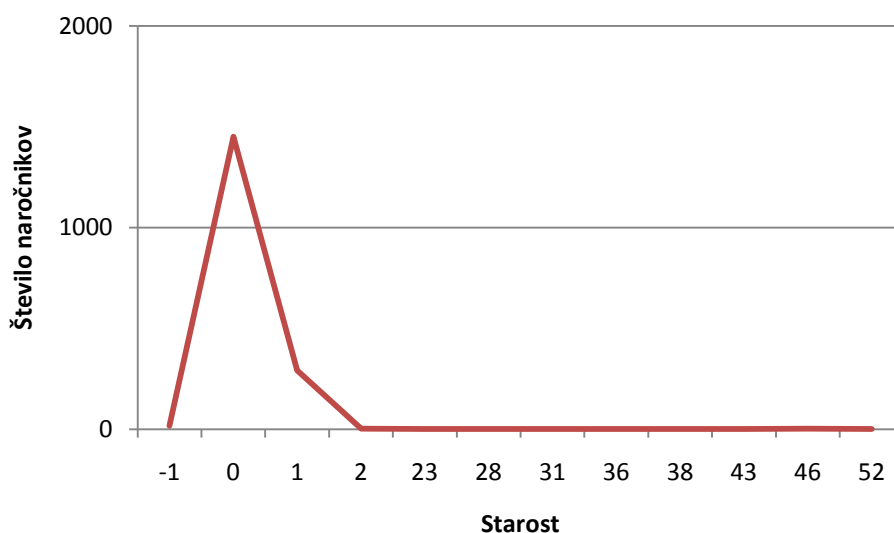
Še ena pomanjkljivost pa je vezana na naročnike, za katere je starost znana. Na Sliki 10 je prikazana distribucija starosti naročnikov, kjer se jasno pokaže anomalija: več kot 30% naročnikov je starih do dve leti. Več kot očitno je, da podatek o starosti naročnikov niti približno ne odraža realnega stanja.

Slika 10: Starost naročnikov



Po podrobnem pregledu podatkov je bilo mogoče ugotoviti, da je velik delež zapisov z napako vezan na podjetja. Večina podjetij ima namesto praznega podatka o rojstnem dnevu le-tega izpolnjenega. Glede na najpogostejše vrednosti, kjer se starost giblje med +1 in -1, gre verjetno za napako, ki se je zgodila ob samem vnosu podatka ali pri morebitni obdelavi podatkov v sistemu za zaračunavanje.

Slika 11: Podatek o starosti za podjetja



Obstoj duplikatov, več virov podatkov ima za posledico nasprotujoče si predloge v študiji ali analizi (Eppler, 2006, str. 28), kar se v tej analizi dogaja z podatki naročnikov. Podatki, ki so služili kot vir za analizo, so bili v zbrani in preneseni iz sistema za zaračunavanje, ki je

edini vir podatkov o uporabi storitev. Osebni podatki, kot je rojstni datum, pa se nahajajo tudi v sistemu za management odnosov z odjemalci, kjer je omenjeni podatek obvezen.

Poleg integracije virov je potrebno določiti tudi zaupanja vredne vire posameznih informacij, za kar bi bilo potrebno opraviti dodatno analizo, ki pa presega okvire tega magistrskega dela.

4.3 Stroški informacijske (ne)kakovosti

Organizacija v času nastanka tega dela, ugotovitev o slabi kakovost informacij ter njihovem vplivu ni upoštevala, kar bi bila osnova za ugotavljanje stroškov povezanih s kakovostjo informacij.

Glede na ugotovljene pomankljivosti, pa bi se v prihodnosti lahko pojavili naslednji stroški, posledica slabe kakovosti informacij:

- Nedlovanje oziroma neuspešnost procesa za zadrževanje kupcev, stroški ponovne pridobitve naročnikov, ki bi lahko nastali zaradi neustrezne kakovosti informacij, na podlagi katerih se pripravlja ponudba za naročnike. Tovrstni stroški so lahko že nastali, vendar analiza za ugotavljanje njihovega morebitnega nastanka še ni bila izvedena.
- Stroški ponovnega prenosa informacij v podatkovno skladišče, stroški popravkov in prilagoditev poročil ter postopkov za poročanje v primeru spremembe strukture, oblike in vsebine podatkov hranjenih v podatkovnem skladišču, ter stroški čiščenja informacij. Glede na ugotovljeno kakovost informacij, ki izvirajo iz podatkovnega skladišča, je nastanek tovrstnih stroškov v prihodnosti skoraj zanesljiv.
- Stroški z naslova zgrešenih ali zamujenih priložnosti:
 - izgubljeni naročniki zaradi ponudbe, ki ni prilagojena njihovim potrebam in navadam ter
 - neukrepanje ob prekinitvah naročniških pogodb

Dokler pa organizacija ne prizna obstoja problemov, povezanih s kakovostjo informacij, bodo stroški, povezani s kakovostjo informacij, predstavljeni kot stroški običajnega poslovanja organizacije.

4.4 Predlogi in izboljšave

Glede na ugotovljeno stanje kakovosti informacij, je logičen naslednji korak izboljšava kakovosti informacij. Ključna pomanjkljivost obstoječega stanja je nepovezanost in podvajanje virov informacij, z odpravo teh pomanjkljivosti pa bi bilo mogoče zagotavljati cilje, ki so bili prvotno zastavljeni za izvedbo analize transakcijskih podatkov.

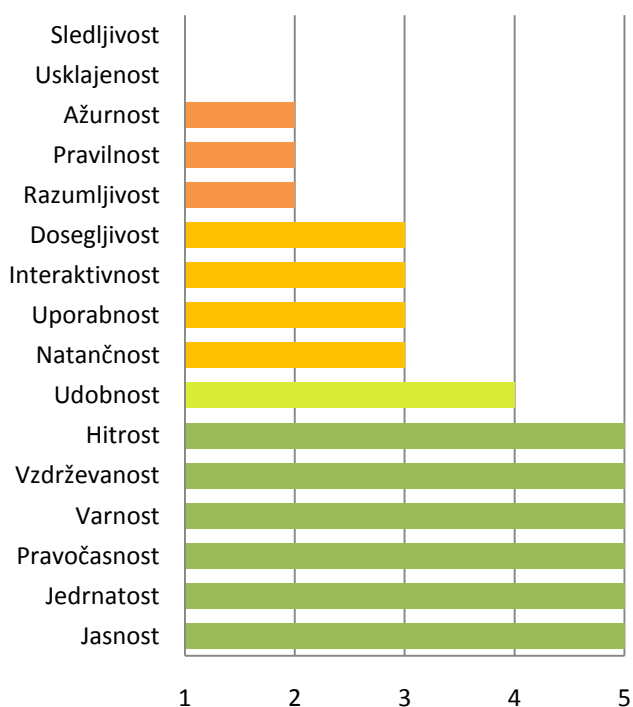
4.4.1 Izboljšava kakovosti informacij

Informacije slabe kakovosti povzročajo visoke stroške, s tem pa ogrožajo organizacijo. Edini namen za izboljšavo informacijske kakovosti je zagotavljanje učinkovitosti poslovnega procesa z povečevanjem zadovoljstva kupcev in drugih v proces vpletenih deležnikov. To pa ima za posledico povečevanje dobička in vrednosti vlagateljev. Povečevanje kakovosti informacij je poslovna nuja, stvar preživetja organizacije in konkurenčna prednost (English, 1999, str. 13).

Kakovost informacij sem ocenil glede na izbranih šestnajst kriterijev informacijske kakovosti po Epplerju (2006). Kriteriji, kjer kakovost po mojem mnenju ni dosegala najvišjega nivoja, so:

- Razumljivost, ocena: 2/5
- Natančnost, ocena: 3/5
- Uporabnost, ocena: 3/5
- Usklajenost, ocena: 1/5
- Pravilnost, ocena: 2/5
- Ažurnost, ocena: 2/5
- Sledljivost, ocena: 2/5
- Interaktivnost, ocena: 3/5
- Dosegljivost, ocena: 3/5

Slika 12: Kritične ocene po kriterijih informacijske kakovosti



Predloge za povišanje kakovosti informacij po posameznih kriterijih, po katerih sem kakovost informacij tudi ocenil, sem podal v nadaljevanju.

4.4.1.1 Sledljivost

Sledljivost informacij je zaenkrat popolnoma prezrt problem. Poled zagotavljanja beleženja sprememb, časovnega žigosanja le-teh ter hranjenja informacij o avtorjih sprememb oziroma o viru je nujno potrebno vzpostaviti tudi postopek oziroma politiko, za zagotavljanje enovitega beleženja dostopov in sprememb. Hkrati z uvedbo sledljivosti bi pripomogli tudi k večji kredibilnosti informacij.

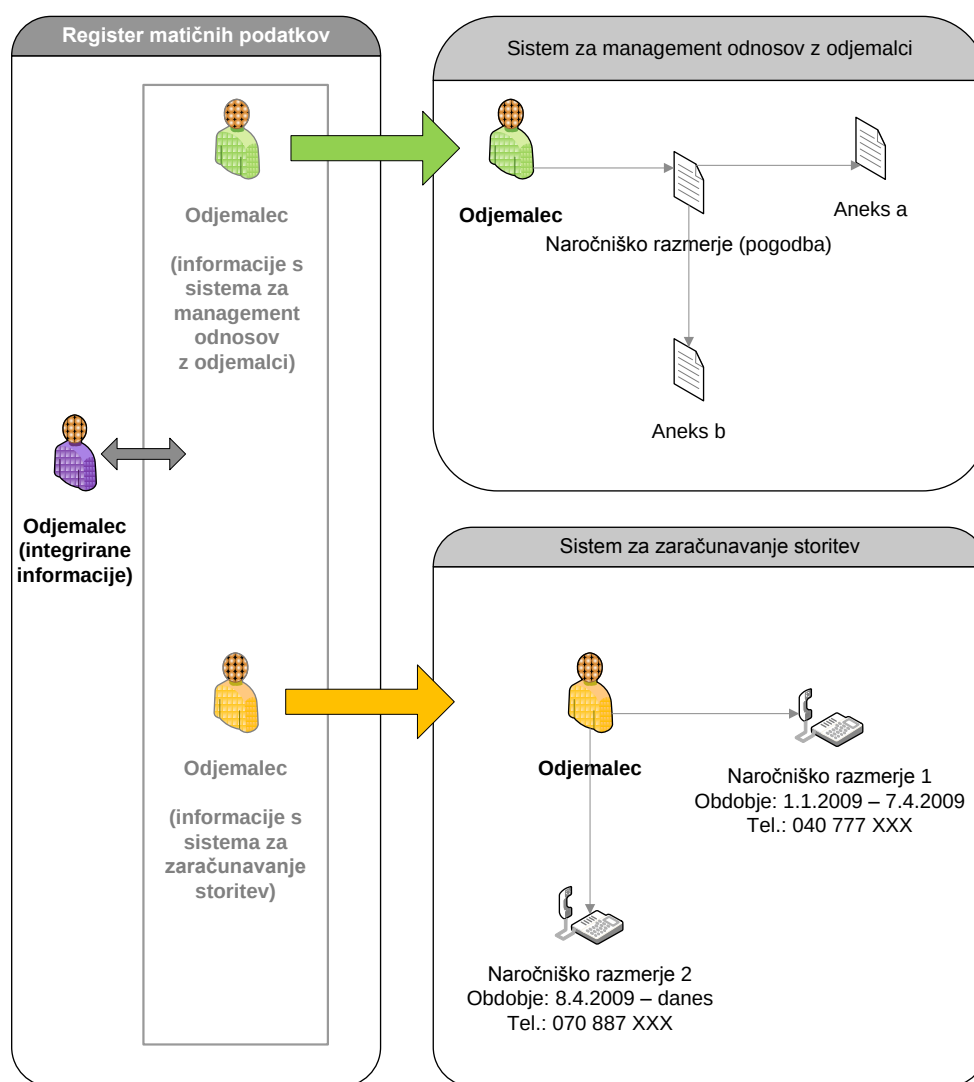
4.4.1.2 Usklajenost

Med seboj neodvisni sistemi, ki skrbijo za podporo posameznim procesom poslovanja, so vir informacij, ki so med seboj tako kot sistemi, popolnoma neusklajene. Loshin (2009) predlaga dve, oziroma tri arhitekturne rešitve za management matičnih podatkov:

- Navidezni seznam matičnih podatkov, implementiran v obliki registra
- Popolnoma poenotena zbirka matičnih podatkov za usklajen dostop, implementiran kot transakcijsko vozlišče
- Kombinacija obeh pristopov v hibridnem modelu

Transakcijsko vozlišče (angl. *transaction hub*) je eno samo odložišče, ki se uporablja za ravnanje z vsemi vidiki matičnih podatkov. Podatkovni elementi niso objavljeni v zunanjih aplikacijskih sistemih. Ker obstaja samo ena kopija matičnih podatkov, so vsi sistemi prilagojeni tako, da dostopajo neposredno do vozlišča (Loshin, 2009, str. 168). Uvedba vozlišča za dostop do matičnih podatkov je praktično neizvedljiva ravno zaradi posega v delovanje vseh obstoječih sistemov. Sistem za management odnosov z odjemalci in sistem za zaračunavanje storitev, kot tipična sistema, ki bi morala imeti poenoten register matičnih podatkov o odjemalcih, sta sistema namenjena za splošno uporabo in nista razvita zgolj za potrebe konkretne organizacije, kar bi dodatno otežilo morebitno prilagajanje delovanja.

Slika 13: Register matičnih podatkov



Pri arhitekturi v obliki registra pa je rešitev v obliki seznama, ki hrani le minimalen nabor atributov, ki so potrebni za enolično identifikacijo entitete. Glavni seznam, ki ga zagotavlja register, vsebuje kazalce na izvorni sistem za vsako edinstveno določeno entiteto (Loshin, 2009, str. 166). Z uvedbo arhitekture v obliki registra je mogoče zagotoviti enovit seznam matičnih podatkov, ki omogoča preslikavo različnih instanc iste entitete po vseh sistemih znotraj organizacije. Berson in Dubov (2007) rešitev priporočata organizacijam, ki nameravajo še nekaj časa obdržati obstoječe porazdeljene sisteme, sredstva za razvoj pa želijo nameniti posodobitvam starejših sistemov.

Z uporabo enotnega registra matičnih podatkov in preslikavo podatkov, ki enolično določajo naročnike, bi bilo mogoče uskladiti podatke različnih sistemov bodisi po tem, ko vstopijo v podatkovno skladišče ali pa pred tem, kar bi uporabo podatkov iz podatkovnega skladišča dodatno poenostavilo.

4.4.1.3 Ažurnost

Za zagotavljanje ažurnosti informacij je potrebno vzpostaviti avtomatiziran prenos podatkov v podatkovno skladišče. Edino tako je mogoče zagotoviti, da bodo informacije v podatkovnem skladišču kar se da ažurne. Pogoji za avtomatiziran prenos pa je pravilno načrtovanje, saj je podatke pred prenosom v podatkovno skladišče potrebno ustrezno obdelati. Stopnja podrobnosti na nivoju posameznih transakcij navadno ni primerna za podatkovno skladišče. Poročila, ki se tipično izdelujejo na podlagi informacij iz podatkovnega skladiča, so tipično vezana na mesečno obdobje, zato tipično tudi ni potreben prenos na nivoju posameznih transakcij.

4.4.1.4 Pravilnost

Uskladitev informacij bo imela pozitiven vpliv tudi na pravilnost informacij. Osebni podatki o naročnikih so ažurni in pravilni v sistemu za management odnosov z odjemalci, v sistemu za zaračunavanje pa žal niso. Z zajemom informacij iz pravega vira se odpravi večina nepravilnih informacij. V sistemu za zaračunavanje storitev so neažurne predvsem tiste informacije, ki niso strogo vezane na porabo storitev in njihovo zaračunavanje. Ker se te informacije dodatno ne uporabljajo v nobenem kritičnem poslovnem procesu, njihovo ažuriranje trenutno ni prioriteto.

Enako je s problemom napačnih datumov prekinitev ankesov k naročniškim pogodbam, v sistemu za management odnosov z odjemalci so zavedeni pravilno, skupaj s celotno zgodovino sprememb.

4.4.1.5 Razumljivost

Za povečanje razumljivosti je potrebno v podatke za analizo vključiti tudi vse spremembe in ne samo zadnje vrednosti atributov. Poleg zadnjega datuma sklenitve aneksa k naročniški pogodbi je nujno dodati tudi vse predhodnje aktivnosti, vse v preteklosti sklenjene anekse.

Za boljše razumevanje obnašanja naročnikov, tudi tistih, ki so naročniško razmerje že preknili, pa bi bilo morda smiselno v analizo vključiti tudi informacije z drugih virov podatkov, ko so na primer podatki o morebitnih reklamacijah ali klicih v klicni center.

4.4.1.6 Dosegljivost

Prosta dosegljivost do informacij, ki so uporabljene za analizo zaradi zaupnosti ni mogoča. Možno pa bi bilo omogočiti manj varovan dostop do omejenega obsega informacij, ki so anonimizirane. V primeru, da bi informacijski uporabniki potrebovali bolj natančne informacije, bi dostop do njih potekal na enak način kot dosedaj: s pisnim pooblastilom in pripravo podatkov na zahtevo.

4.4.1.7 Interaktivnost

Informacijski uporabnik zaenkrat nima možnosti prilagoditve informacij. Ne nadzira in ne vpliva na proces oblikovanja informacij iz podatkovnega skladišča. Za prilagajanje informacij oziroma možnost vpliva na proces pridobivanja informacij ter njihovo obliko pa bi bilo potrebno vzpostaviti napreden vmesnik, ki bi služil tovrstnim funkcijam

4.4.1.8 Uporabnost

Informacije, ki se za analizo pripravljajo sedaj so le posredno uporabne, podatki pa niso človeško berljivi. Seveda je glavni razlog v tem, da je tovrstna analiza podatkov prototipna. Za neposredno uporabnost informacij, bi bilo potrebno vnaprej pripraviti poročila, ki bi jih informacijski uporabniki dobivali periodično, ob vnaprej predvidenih časovnih intervalih, ali pa na zahtevo. Poročila bi morala biti prilagojena različnim tipom informacijskih uporabnikom, predvsem kar se tiče nivoja podrobnosti. Vsi uporabniki ne potrebujejo enako podrobnega pregleda nad informacijami. Glede na rezultate analize transakcijskih podatkov bi se grozdenje lahko izvajalo na mesečnem nivoju, vodstveni delavci pa bi dobivali le poročila v kar se da jedrnatih obliki, za zaposlene v oddelku trženja pa bi bilo potrebno pripraviti bolj podrobna poročila, da bi na podlagi prejetih informacij lahko sprejemali bolj kakovostne odločitve.

4.4.1.9 Natančnost

Podoben problem kot pri kriteriju uporabnosti se pojavlja tudi pri kriteriju za natančnost. V primeru avtomatiziranih poročil z večimi nivoji podrobnosti bi bilo potrebno ustrezno prilagoditi tudi nivoje natančnosti. Vodstveni delavci nedvomno ne potrebujejo enake stopnje natančnosti kot na primer delavci iz področja trženja.

4.4.1.10 Udobnost

Dokler je za dostop in obdelavo transakcijskih podatkov predvidena izjemno stroga stopnja varovanja je praktično nemogoče povečati stopnjo udobja, brez povečevanja varnostnega tveganja.

Eden od možnih načinov pa je morda manj varovan dostop do omejenega obsega informacij, ki so anonimizirane, najbolj kritične informacije pa prikrite. S tem bi omogočili bolj prost in s tem udoben dostop do informacij brez pisnih odobritev in podobno. S tem pa se na drugi strani dodatno oteži priprava in dostava informacij, ki mora skrbeti za streženje informacij v več oblikah. Zagotoviti je potrebno tudi managiranje pravic in dostopov v digitalni obliki namesto v papirni, kot to poteka sedaj.

4.4.2 Uporaba pristopov za zagotavljanje kakovosti informacij

Implementacija avtomatiziranih postopkov za urejanje in preverjanje vnosa podatkov v aplikacijah lahko zagotovi osnovno raven informacijske kakovosti. Postopki za urejanje in preverjanje lahko preprečijo očitne napake kot je vpis datuma »32.12.1999«, ali navedbo dejstva, da ima nekdo 97 otrok. Samo ljudje lahko ustvarijo točne informacije (English, 1999, str. 401). Z drugimi besedami, informacijska tehnologija sama ni dovolj za zagotavljanje celovite kakovosti informacij, ključni del odgovornosti za informacijsko kakovost sloni na ljudeh.

Ustvarjalci informacij bodo ustvarili informacije, ki bodo kakovostne le do nivoja za katerega so naučeni in odgovorni. Resnične in trajnostne izboljšave informacijske kakovosti je mogoče doseči le z uvedbo managementa odgovornosti za informacije na enak način, kot za druge poslovne produkte in vire. Če managerji ne nosijo odgovornosti za informacijsko kakovost, ne obstaja spodbuda za odstranjevanje ovir, ki ustvarjalcem informacij preprečujejo ustvarjanje točnih, popolnih in pravočasnih informacij (English, 1999, str. 402). Analiza kakovosti informacij kaže na to, da je kakovost informacij zadovoljiva le do nivoja, za katerega se informacije primarno uporabljajo, kar pa pomeni, da na višjih nivojih organizacije ni potrebne predanosti za zagotavljanje informacijske kakovosti.

Ocena kakovosti informacij s pomočjo šestnajstih izbranih kriterijev kakovosti je pokazala pomanjkljivosti, skupni imenovalec simptomov pa je pomanjkanje odgovornosti za skrb nad kakovostjo informacij na vseh nivojih organizacije. Skrbništvo nad informacijsko kakovostjo na vseh nivojih organizacije je ključen element okolja za zagotavljanje informacijske kakovosti.

Glede na to, da je eden od bistvenih vzrokov za nekakovost prav porazdeljenost informacij po različnih sistemih znotraj organizacije ter njihovo podvajanje, je nujna uvedba managementa matičnih podatkov. Pri tem je potrebno poudariti, da bistvo problema ne leži v informacijskih tehnologiji ampak v pristopu vodstva k problematiki kakovosti informacij. Pogoji za uvedbo managementa matičnih podatkov so izpolnjeni naslednji pogoji (Loshin, 2009, str. 15):

- efektivna tehnična struktura za sodelovanje
- organizacijska pripravljenost
- sprejetje s strani celotne organizacije in integracija
- merljiva visoka kakovost informacij
- nadzor nad procesi managementa matičnih podatkov (in drugih) s pomočjo politik in postopkov za upravljanje z informacijami.

Uvedba managementa matičnih podatkov torej predvideva že obstoječo visoko kakovost informacij in pripravljenost ter sodelovanje celotne organizacije za uvedbo sprememb. Za dvig kakovosti pa bo največ potrebno postoriti ravno na organizacijskem področju organizacije. English (1999) za zagotavljanje informacijske kakovosti zagovarja metodologijo Kaizen – nenehen proces izboljšave, kjer so vsi člani organizacije odgovorni za izvajanje nenehnega procesa za izboljšavo celotnega poslovanja. Dokler pa organizacijsko okolje ne bo vzpodbujalo izboljšav na vseh organizacijskih nivojih in določalo odgovornosti ter skrbništva za kakovost informacij, se korenite spremembe ne bodo izvedle.

4.5 Ovine in omejitve

Na obstoječe stanje kakovosti informacij in njihovo uporabo imajo pomemben vpliv tudi zakonske določbe, zakonodaja s področja elektronskih komunikacij in varovanja osebnih podatkov namreč natančno določa pravila rokovanja s transakcijskimi in osebnimi podatki. Organizacijske ovire pa izvirajo predvsem iz dejstva, da nihče v podjetju ni odgovoren za kakovost informacij.

4.5.1 Vpliv zakonodaje na hranjenje in uporabo podatkov

Transakcijski podatki naročnikov bi morali biti anonimizirani takoj, ko je obdelava za zaračunavanje storitev končana, kar določa Zakon o elektronskih komunikacijah (Ur.l. RS, št. 13/2007-UPB1, v nadaljevanju ZEKom). ZEKom na ta način preprečuje povezavo najdenih vzorcev v podatkih s konkretnim naročnikom. Iskanje vzorcev je na ta način omejeno na anonimne podatke oziroma na podatke, ki ne razkrivajo identitete uporabnikov, razen v primeru, da naročniki soglašajo z uporabo in obdelavo njihovih transakcijskih podatkov v namene njim prilagojenega trženja.

Glede na ZEKom je kakršno koli trženje dovoljeno samo na podlagi naročnikovega ali uporabnikovega predhodnega soglasja. Naročniki oziroma uporabniki morajo biti pri tem predhodno obveščeni o vrstah prometnih podatkov, ki se obdelujejo ter trajanju takšne obdelave pred pridobitvijo soglasja. Uporabnik ali naročnik pa ima pravico, da kadarkoli prekliče to soglasje.

Zakon o varovanju osebnih podatkov (Ur.l. RS, št. 94/2007-UPB1, v nadaljevanju ZVOP-1) anonimiziranje opredeljuje kot takšno spremembo oblike osebnih podatkov, da jih ni več mogoče povezati s posameznikom ali pa je le-to mogoče le z nesorazmerno velikimi naporji, stroški ali porabo časa.

Zakonske določbe torej narekujejo rokovanje z transakcijskimi in osebnimi podatki, ki vpliva na delovanje celotne organizacije.

4.5.2 Organizacijske ovire

Organizacija se problemov vezanih na kakovost informacij ne zaveda ali pa jih zavestno ignorira. S tem je organizacija izbrala strategijo soočenja s posledicami slabe informacijske kakovosti. Morda največji problem je sama togost organizacije, saj se z informacijsko kakovostjo ne ukvarja nobeno izmed področij organizacije, posledično nihče znotraj organizacije ne nosi odgovornosti za probleme, ki jih je in jih bo informacijska nekovost povzročila v prihodnosti.

Zaposlene čaka resnično delo: zadovoljevanje potreb odjemalcev, doseganje proizvodnih norm, izdelavo strategij in pogovori z zahtevnim vodstvom. Koga skrbijo biti in bajti, nihče se namreč ne pritožuje (Redman, 2001, str. 3). Dokler je informacijska kakovost z vidika organizacije na zadovoljivem nivoju, se nihče ne ukvarja z iskanjem alternativne rešitve. Ta pojav De Bono (1990) opisuje kot blokado z odprtjem (angl. *blocked by openness*). Gre za pojav, ko je prvotna pot ali rešitev zadovoljiva, zato nihče niti ne pomisli na drugačno rešitev, kaj šele, da bi jo aktivno iskal.

Hillard (2010) opaza, da je v zadnjem desetletju močno naraslo število uspešno speljanih projektov vezanih na management informacij, kot so podatkovna skladišča in odložišča dokumentov (angl. *document repositories*). Medtem kot so strokovnjaki s področja informacijske tehnologije izpopolnili tehnične rešitve, management še vedno s težavo spoznava pomembnost vloge, ki jo v poslovanju igrajo informacije. V prihodnosti bodo zmagovalci tiste organizacije, ki bodo razumele vlogo informacij in jo bodo med prvimi na trgu uporabile kot vzvod za pridobivanje konkurenčne prednosti. Managerji pa morajo po drugi strani prav tako sprejeti odgovornost za managiranje informacij prav tako kot so prevzeli odgovornost za managiranje poslovnih operacij.

Cameron in Green (2009) poudarjata pomembnost periferne vida, kot ga imenujeta (angl. *peripheral vision*). Vodstvo bi moralo biti bolj dovzetno za dogajanje okrog njih. To pomeni, da bi morali biti pozorni na dogajanje, ki ni popolnoma očitno in neposredno vidno ter biti pozorni na gibanje in spremembe. Periferni vid pa je mogoče pridobiti s časom, z komunikacijo in predvsem poslušanjem širokega profila ljudi, z postavljanjem dobrih vprašanj in vzdrževanjem odprtih odnosov. Kar pomeni, da naj bi vodstvo nemanilo manj časa sprejemanju odločitev in več časa namenilo zbiranju informacij.

SKLEP

Izvedba vzorčne analize ter naknadna ocena kakovosti informacij lahko omogoči dober vpogled v dejansko stanje kakovosti informacij. Dokler so informacije namenjene samo uporabi v primaren namen, torej podpora ključnih poslovnih procesov v organizaciji, je tudi nivo kakovosti zadovoljiv le za učinkovito doseganje prvotnega namena.

Količina zbranih podatkov strmo narašča, posledično pa so organizacije bogate s podatki in revne z znanjem. Pomembnejše od samega zbiranja podatkov je sposobnost pretvarjanja podatkov v znanje z iskanjem informacij, vzorcev in pravil. Tu nedvomno pozitivno vpliva razvoj informacijske tehnologije, žal pa nekakovosten vir negativno vpliva na kakovost ugotovitev.

Z večanjem količine shranjenih informacij je smiselno iskati tudi nove načine njihove uporabe, da bi iz informacij pridobili dodatno vrednost. Današnja tehnologija omogoča obdelavo čedalje večje količine informacij v realnem času in bila bi potrata, če informacij in tehnologije ne bi izkoristili v prid boljšim odločitvam v prihodnosti. Hkrati pa je na mestu previdnost pri uporabi dobljenih rezultatov, saj nekakovostne informacije na vходу v proces analize lahko postrežejo z le na videz točnimi rezultati.

Izvedba vzorčne analize na primeru transakcijskih podatkov naročnikov, uporabnikov mobilne telefonije, je pokazala na probleme z informacijsko kakovostjo. Vzroki zanje so poleg tehničnih predvsem organizacijski. Bistven korak za zagotavljanje kakovosti pa je ocena trenutnega nivoja informacijske kakovosti. Brez zavedanja, da informacije niso popolnoma kakovostne, do ukrepov za zagotavljanje kakovosti v prihodnosti ne more priti.

Strategija, ki jo je organizacija izbrala za reševanje problemov z informacijsko nekakovostjo je soočenje s posledicami, kar je na dolgi rok povsem neučinkovita strategija. Da bi zagovorili kakovost informacij na dolgi rok, je edina rešitev lociranje virov informacijske kakovosti ter izvedba ukrepov za zagotavljanje informacijske kakovosti na samem viru. Potrebno je poskrbeti, da do problemov z informacijsko kakovostjo sploh ne pride, kar je veliko bolj učinkovito kot odprava njihovih posledic.

Sama informacijska tehnologija lahko poskrbi le za dolčen nivo kakovosti, bistveni so ustvarjalci in uporabniki informacij. Uvedba organizacijske kulture, ki podpira kakovost na vseh nivojih organizacije, pomeni temelj za uvedbo managementa matičnih podatkov. Dokler pa do sprememb ne bo prišlo, bodo stroški informacijske nekakovosti vodstvu predstavljeni kot stroški običajnega poslovanja.

Zagotavljanje kakovosti informacij je naloga vseh posameznikov v organizaciji, organizacija pa mora vzpodbujati Kaizen, proces nehnega izboljševanja kakovosti na vseh nivojih organizacije. Informacijska tehnologija sama ne mora zagotavljati informacijske kakovosti, pač pa jo lahko zagotovijo le ustvarjalci informacij – ljudje.

LITERATURA IN VIRI

1. Adamson, C. (2006). *Mastering Data Warehouse Aggregates: Solution for StarSchema Performance*. Indianapolis: Wiley Publishing.
2. Al-Hakim, L. (2007). *Information Quality Management: Theory and Practice*. Hershey: Idea Group Publishing.
3. Anderson, C. (2009). *Free: How Today's Smartest Businesses Profit by Giving Something for Nothing*. London: Random House.
4. Batini, C., & Scannapieca, M. (2006). *Data Quality: Concepts, Methodologies and Techniques (Data-Centric Systems and Applications)*. Berlin: Springer.
5. Berkun S. (2009). *The Myths of Innovation*. Sebastopol: O'Reilly Media.
6. Berson A., & Dubov L. (2007). *Master Data Management and Customer Data Integration for a Global Enterprise*. New York: McGraw-Hill.
7. Cameron, E., & Green, M. (2009). *Making Sense of Change Management: A Complete Guide to the Models, Tools & Techniques of Organizational Change (2nd ed.)*. London: Kogan Page.
8. Crosby, P. B. (1979). *Quality is Free: The Art of Making Quality Certain*. New York: Penguin Group.
9. De Bono, E. (1990). *Lateral Thinking*. London: Penguin Group.
10. English, L. P. (1999). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. New York: John Wiley & Sons, inc.
11. Eppler, M. J. (2006). *Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes (2nd ed.)*. Berlin: Springer.
12. *European Telecommunications Standards Institute - Specification of the Subscriber Identity Module*. Najdeno 29.7.2010 na spletnem naslovu <http://www.ttfn.net/techno/smartcards/gsm11-11.pdf>
13. Hillard, R. (2010). *Information-Driven Business: How to Manage Data and Information for Maximum Advantage*. Hoboken: John Wiley & Sons.
14. Kimball, R., Ross, M., Thorntwaite, W., Mundy, J., & Becker, B. (2006). *The Data Warehouse Lifecycle Toolkit (2nd ed.)*. Indianapolis: Wiley Publishing.
15. *IBM Press room – Tiskovno središče podjetja IBM*. Najdeno 16.6.2010 na spletnem naslovu <http://www-03.ibm.com/press/us/en/pressrelease/26077.wss>
16. Kawasaki, G. (2008). *Reality check: The Irreverent Guide to Outsmarting, Outmanaging and Outmarketing Your Competition*. New York: Penguin Group.
17. Loshin, D. (2009). *Master Data Management*. Burlington: Morgan Kaufmann.
18. MacLennan, J., Tang Z., & Crivat B. (2009). *Data Minig with Microsoft SQL Server 2008*. Indianapolis: Wiley Publishing.
19. Malhotra N. K., & Birks D. F. (2007). *Marketing Research: An Applied Approach (3rd ed.)*. London: Prentice Hall.

20. Mandelbrot, B. B., & Hudson, R. L. (2004). *The (Mis)behavior of Markets: A Fractal View of Financial Turbulence*. New York: Basic Books.
21. McGilvray, D. (2008). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Burlington: Morgan Kaufmann.
22. Parsons E., & Maclaran P. (2009). *Contemporary Issues in Marketing and Consumer Behaviour*. Oxford: Butterworth - Heinemann.
23. Peter, L. J., & Hull, R. (1994). *The Peter Principle: Why Things Always Go Wrong*. New York: William Morrow & Co.
24. Popper K. (2002). *The Logic of Scientific Discovery*. London: Routledge.
25. Redman, T. C. (2001). *Data Quality: The Field Guide*. Oxford: Butterworth-Heinemann.
26. *Scientific American* - Kryder's law. Najdeno 26.5.2010 na spletnem naslovu <http://www.scientificamerican.com/article.cfm?id=kryders-law&ref=sciam>
27. Taleb N. N. (2007). *Fooled by Randomness: The Hidden Role of Chance in Life and in the Market*. New York: Penguin Group.
28. Taleb N. N. (2008). *The Black Swan: The Impact of the Highly Improbable*. New York: Penguin Group.
29. *Trgovine Tuš, Tuš klub – Spletna stran podjetja*. Najdeno 29.4.2010 na spletnem naslovu <http://www.tus.si/index.php?Nivoid=23>
30. *Weka – Programska oprema*. Najdeno 15.6.2010 na spletnem naslovu <http://www.cs.waikato.ac.nz/ml/weka/>
31. Westphal C., & Blaxton T. (1998). *Data Mining Solutions: Methods and Tools for Solving Real-World Problems*. New York: Wiley Computer Publishing.
32. Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques (2nd ed.)*. San Francisco: Morgan Kaufmann.
33. Zakon o elektronskih komunikacijah. *Uradni list RS* št. 13/2007-UPB1.
34. Zakon o varstvu osebnih podatkov. *Uradni list RS* št. 94/2007-UPB1.