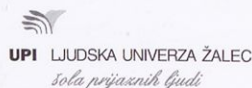
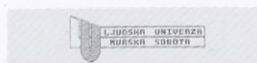
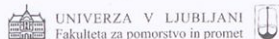




# Testni centri ECDL

**ECDL** (European Computer Driving License), ki ga v Sloveniji imenujemo evropsko računalniško spričevalo, je standardni program usposabljanja uporabnikov, ki da zaposlenim potrebno znanje za delo s standardnimi računalniškimi programi na informatiziranem delovnem mestu, delodajalcem pa pomeni dokazilo o usposobljenosti. V Evropi je za uvajanje, usposabljanje in nadzor izvajanja ECDL pooblaščen ustanova ECDL Foundation, v Sloveniji pa je kot član CEPIS (Council of European Professional Informatics Societies) to pravico pridobilo Slovensko društvo INFORMATIKA. V državah Evropske unije so pri uvajanju ECDL močno angažirane srednje in visoke šole, aktivni pa so tudi različni vladni resorji. Posebej pomembno je, da velja spričevalo v več kot osemdesetih državah, ki so vključene v program ECDL. Doslej je bilo v svetu izdanih že več kot tri milijone indeksov, v Sloveniji okoli 1700 in podeljenih okoli tisoč spričeval. Za testne centre ECDL so se v Sloveniji usposobile organizacije, katerih logotipi so natisnjeni na tej strani.



# U P O R A B N A I N F O R M A T I K A

2004 ŠTEVILKA 1 JAN/FEB/MAR LETNIK XII ISSN 1318-1882

## ☒ Uvodnik

## ☒ Razprave

- Tomaž Dogša: **Problematika sistemov za vodenje popraviljanja in vzdrževanja** 5
- Stjepan Pervan, Martina Kern: **Reševanje problema pomanjkanja unikatnih naslovov internetnega protokola** 11
- Krista Rizman Žalik: **Podatkovna skladišča in kakovost podatkov** 19
- Miro Ribič, Andrej Kovačič: **Sistem za upravljanje delovnih procesov** 30
- Marko Hölbl, Petra Grm, Boštjan Brumen, Tatjana Welzer, Izidor Golob: **Natančnost uvrščanja slovenskih besedil** 42

## ☒ Poročila

- Ladislav Mikola: **Uporaba desetiških SI predpon v informatiki** 47
- Zsuzsana Toszegi: **Digitaliziranje madžarske kulturne dediščine** 50

## ☒ Koledar prireditev

51

ISSN 1318-1882

**Ustanovitelj in izdajatelj:**

Slovensko društvo INFORMATIKA  
Vožarski pot 12  
1000 Ljubljana

**Predstavniki**

Niko Schlamberger

**Odgovorni urednik:**

Andrej Kovačič

**Uredniški odbor:**

Marko Bajec, Vesna Bosilj Vukšič, Dušan Caf, Aljoša Domijan, Janez Grad, Jurij Jaklič, Milton Jenkins, Andrej Kovačič, Tomaž Mohorič, Katarina Puc, Vladislav Rajkovič, Heinrich Reinermann, Ivan Rozman, Niko Schlamberger, John Taylor, Ivan Vezočnik, Mirko Vintar, Tatjana Welzer - Družovec

**Recenzenti prispevkov za objavo v reviji Uporabna informatika:**

Marko Bajec, Tomaž Banovec, Vladimir Batagelj, Marko Bohanec, Vesna Bosilj Vukšič, Dušan Caf, Srečko Devjak, Aljoša Domijan, Tomaž Erjavec, Matjaž Gams, Tomaž Gornik, Janez Grad, Miro Gradišar, Jože Gričar, Jozsef Györkos, Marjan Heričko, Jurij Jaklič, Milton Jenkins, Andrej Kovačič, Iztok Lajovic, Tomaž Mohorič, Katarina Puc, Vladislav Rajkovič, Heinrich Reinermann, Ivan Rozman, Niko Schlamberger, Ivan Vezočnik, Miro Vintar, Tatjana Welzer - Družovec, Franc Žerdin

**Tehnična urednica**

Mira Turk Škraba

**Oblikovanje**

Bons

**Prelom**

Dušan Weiss, Ada Poklač

**Tisk**

Prograf

**Naklada**

700 izvodov

**Naslov uredništva**

Slovensko društvo INFORMATIKA  
Uredništvo revije Uporabna informatika  
Vožarski pot 12, 1000 Ljubljana  
www.drustvo-informatika.si/posta

Revija izhaja četrtletno. Cena posamezne številke je 4.500 SIT. Letna naročnina za podjetja 17.800 SIT, za vsak nadaljnji izvod 11.900 SIT, za posameznike 5.900 SIT, za študente 2.800 SIT.

Revijo sofinancira Ministrstvo za šolstvo, znanost in šport RS.

Revija Uporabna informatika je od številke 4/VII vključena v mednarodno bazo INSPEC.

Revija Uporabna informatika je pod zaporedno številko 666 vpisana v razvid medijev, ki ga vodi Ministrstvo za kulturo RS.

© Slovensko društvo INFORMATIKA

## Navodila avtorjem

Revija Uporabna informatika objavlja izvirne prispevke domačih in tujih avtorjev na znanstveni, strokovni in informativni ravni. Namenjena je najširši strokovni javnosti, zato je zaželeno, da so tudi znanstveni prispevki napisani čim bolj poljudno.

Članke objavljamo praviloma v slovenščini, prispevke tujih avtorjev v angleščini.

Prispevki so obojestransko anonimno recenzirani. Vsak članek za rubriko Razprave mora za objavo prejeti dve pozitivni recenziji. O objavi samostojno odloča uredniški odbor.

Prispevki naj bodo lektorirani, v uredništvu opravljamo samo korekturo. Po presoji se bomo posvetovali z avtorjem in članek tudi lektorirali. Prispevki za rubriko Razprave naj imajo dolžino do 40.000, prispevki za rubrike Rešitve, Poročila do 30.000, Obvestila pa do 8.000 znakov.

Naslovu prispevka naj sledi ime in priimek avtorja, ustanova, kjer je zaposlen in elektronski naslov. Članek naj ima v začetku do 10 vrstic dolg izvleček v slovenščini in angleščini, v katerem avtor opiše vsebino prispevka, dosežene rezultate raziskave. Abstract se začne s prevodom naslova v angleščino. Članku dodajte kratek avtorjev življenjepis (do 8 vrstic), v katerem poudarite predvsem delovne dosežke.

Pišite v razmaku ene vrstice, brez posebnih ali poudarjenih črk, za ločilom na koncu stavka napravite samo en prazen prostor, ne uporabljajte zamika pri odstavkih.

Revijo tiskamo v črno-beli tehniki s folije, zato barvne slike ali fotografije kot originali niso primerne. Objavljali tudi ne bomo slik zaslonov, razen če niso nujno potrebne za razumevanje besedila. Slike, grafikoni, organizacijske sheme ipd. naj imajo belo podlago. Po možnosti jih pošiljajte posebej, ne v datoteki z besedilom članka. Disketi z besedom priložite izpis na papirju.

Prispevke pošiljajte po elektronski ali navadni pošti na naslov uredništva revije: ui@drustvo-informatika.si, Slovensko društvo INFORMATIKA, Vožarski pot 12, 1000 Ljubljana. Za dodatne informacije se obračajte na tehnično urednico Miro Turk Škraba.

Po odločitvi uredniškega odbora o objavi članka bo avtor prejel pogodbo, s katero bo prenesel vse materialne avtorske pravice na Slovensko društvo INFORMATIKA. Po izidu revije pa bo prejel nakazilo avtorskega honorarja po veljavnem ceniku ali po predlogu odgovornega urednika.

## Evropa

Pričujoča številka revije *Uporabna informatika* je zadnja pred pridružitvijo Slovenije Evropski uniji. Ob tem se nam porajajo razmisleki, povezani z različnimi predstavami o naši prihodnosti. Spominjamo se navdušenja ob ustanovitvi lastne države, ki smo jo razumeli med drugim tudi kot vrnitev v Evropo, prepričanja, da je naše mesto v Evropski uniji, kar smo dokazali kot priznana najbolje pripravljena država izmed vseh kandidatke za članstvo, pa nekoliko bolj senčnih tonov, ko smo ugotavljali, da ima tudi ta medalja dve plati. Od novega položaja vsi predvsem pričakujemo, da bomo živeli bolje. To je verjetno dokaj idealiziran pogled na prihodnost, ki ga moramo uskladiti z resničnostjo.

Kaj pomeni živeti bolje? Upoštevati moramo, da tudi Evropska unija od novih članic marsikaj pričakuje. V novi zvezi gre za bistveno močnejši potencial gospodarstev držav članic in tistih, ki pristopajo. Pristopnice lahko pri tem precej pridobijo, saj bo mednarodna menjava bistveno olajšana. Gre za spretnost in znanje pri izkoriščanju potencialov, ki niso porazdeljeni enakomerno in prav v tem je naša priložnost. Ob tem mislimo tudi na informacijsko družbo. Evropska unija je smer razvoja določila z Bangemannovim poročilom in z dokumenti ter akcijami, ki so mu sledile. Države pristopnice so v času, ko je bilo objavljeno, razumele tak razvoj kot neizogiben, vendar so imele različna izhodišča in zato tudi različne načine, kako se vključiti vanj. Tega so se lotevale bolj ali manj spretno, z več ali manj pomoči razvitejših, ob močnejši ali skromnejši podpori svojih subjektov. Pomembno vlogo, ki je bila marsikje spregledana, ima pri tem civilna družba.

Slovensko društvo *INFORMATIKA*, ki se bliža že trideseti obletnici ustanovitve, je ustanova civilne družbe par excellence. Ustanovitelji društva so se zavedali pomena informatike še v času in v družbi, ki take vrste podvigom ni bila naklonjena in zato moramo njihov pogum in vizijo tem bolj ceniti. Pomembna poteza je bila tudi začetek izdajanja znanstvene revije *Informatica*, kjer so prispevki v angleškem jeziku, s čimer so napravili društvo mednarodno vidno in spoštovano. Organizacijska znanost je spoznala, da se institucije spreminjajo, njihove funkcije pa ostajajo. To lahko vidimo v proizvodnji, izobraževanju in upravi. Proizvodi in storitve nastajajo zdaj tu, zdaj tam, odvisno od možnosti in poslovnih interesov. Izobraževalne ustanove nastajajo, se združujejo in spreminjajo, uvajajo nove načine usposabljanja in učenja, ki jih omogočajo nove tehnologije; spričevala in diplome se bodo podeljevale tudi v prihodnje. Državne institucije izvajajo iste funkcije v različnih organizacijskih kombinacijah in pod različnimi nazivi.

Civilna družba je neodvisen sogovornik oblasti. Tako razmerje je pri nas relativno novo in še ne uveljavljeno, vendar telo, ki ga imenuje oblast, tej ne more biti enakopraven sogovornik, kaj šele, da bi jo opozarjalo na slabe rešitve in ji nakazovalo boljše. Tak položaj lahko zavzame le neodvisna entiteta in sčasoma postane enakovredna tradicionalnim trem vejam oblasti. Sedanja negotovost civilne družbe glede lastne vloge in moči je verjetno razlog, da v državah pristopnicah ni bilo pričakovanih rezultatov od forumov za informacijsko družbo, ki so nastali po zgledu Bangemannovega foruma. Skromna denarna sredstva, s katerimi so forumi razpolagali, so gotovo tudi med razlogi za to, vendar ne glavni. Najpomembnejša vloga civilne družbe pa je ohranjanje kontinuitete naroda, pri čemer je jezik bistveni element. Slovensko društvo *INFORMATIKA* je to spoznalo in rezultat tega je bila ustanovitev revije *Uporabna informatika*, ki izhaja že deset let, in sekcije za jezik. Jezik ni samo jezik leposlovja, temveč tudi stroke in znanosti; slediti in spremljati mora tekoča dogajanja in dosežke in jih razumljivo zapisovati, prejemati in ohranjati za sporazumevanje in sporočanje.

Kaj imajo torej skupnega naše bližnje članstvo v Evropski uniji, informacijska družba in civilna družba? V družbi razvitejših vedno obstaja nevarnost, da se bodo manj razviti in šibkejši podrejali močnejšim in razvitim. Pri tem ne gre za pritisk in nasilje, dejstvo je, da boljši in učinkovitejši sistemi izpodrivajo slabše in manj učinkovite. Poglejmo npr. amerikanizacijo evropskega življenja: ali se dogaja na silo? Ne, Amerika vstopa v naš svet nenasilno in z našim tihim pristankom, morda celo s sodelovanjem. V podobnem položaju se bomo znašli v Evropi. Informacijska družba, od katere pričakujemo vse ugodnosti industrijske, je zapeljiva in privlačna podoba. Če ne razumemo, da pri tem sicer lahko materialno pridobimo, kulturno pa tudi kaj izgubimo, naša perspektiva naroda, ki se je obdržal dva tisoč let, ni rožnata. Predvsem civilna družba je tista, ki bo vplivala na ohranitev jezika kot bistvenega nosilca narodove kulture in v tem pogledu moramo kot društvo storiti vse, da bo revija privlačna za avtorje in bralce tudi v naši evropski prihodnosti.

Pri tem se moramo zavedati svoje dolžnosti do jezika tudi posamezniki. Nedavno sem na tržnici v državi, kjer živi manjšina, sicer mnogo številčnejša od Slovencev, prebral dvojezični napis: Jezik ne izgine zato, ker se ga tisti, ki ga ne znajo, ne učijo, temveč zato, ker ga tisti, ki ga znajo, ne govorijo. Če to razumemo, se nam tudi v družbi najrazvitejših ni treba bati prihodnosti.

Niko Schlamberger

Ponovno vas vabimo, da si rezervirate čas za udeležbo na posvetovanju

## **XI. DNEVI SLOVENSKE INFORMATIKE**

**14.–16. aprila 2004**

**Kongresni center Grand hotela Emona, Portorož**

predavanja domačih in tujih strokovnjakov

- okrogle mize

- delavnice

- razstave

- družabni dogodki

### **Program posvetovanja:**

poslovna informatika in elektronsko poslovanje

- informatijske tehnologije in internet

- informatijska kultura in družba

Informacije: **[www.dsi2004.org](http://www.dsi2004.org)**

# Problematika sistemov za vodenje popraviljanja in vzdrževanja

Tomaž Dogša

cV&Vs Center za verifikacijo in validacijo sistemov

Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru, Smetanova 17, 2000 Maribor

tdogsa@uni-mb.si

## Povzetek

Vsaka organizacija, ki želi delovati v skladu s smernicami glede kakovosti, mora vzpostaviti sistem, ki bo omogočal beleženje in nadzor nad nepravilnostmi ter spremembami (v nadaljevanju sistem SVPV). Kakovosten sistem SVPV ne omogoča samo nadzora nad zahtevki za spremembo, ampak tudi razne analize, s katerimi lahko ugotavljamo uspešnost oziroma slabosti razvoja. V prispevku je opisana problematika teh sistemov in na kratko komentiran standard IEEE 1044-1993.

## Abstract

### Difficulties of Problem Tracking and Maintenance Systems

Each organisation intending to work according to the quality guidelines has to introduce a system of problem tracking and control of changes. The problem tracking system (PTS) of high quality does not support only data capturing, but also enables comparative analysis of effectiveness and weaknesses of software development cycle. In this paper a brief description of issues concerning PTS is presented and IEEE 1044-1993 Standard is also discussed.

## 1 Uvod

**Ena izmed značilnosti informacijskih sistemov je, da se pogosto spreminjajo. Formalni razlog, s katerim se zahteva sprememba komponente ali sistema, bomo poimenovali zahtevek za spremembo ali na kratko zahtevek. V mnogih primerih so obravnavane samo nepravilnosti in je zahtevek za spremembo v bistvu opažena nepravilnost ali problem. Zahtevki za spremembo so pomemben vir podatkov, na podlagi katerega lahko sklepamo o trenutni kakovosti produkta in razvoja, napovedujemo končno kakovost, razporejamo resurse in napovedujemo datum, ko bo produkt predan naročniku.**

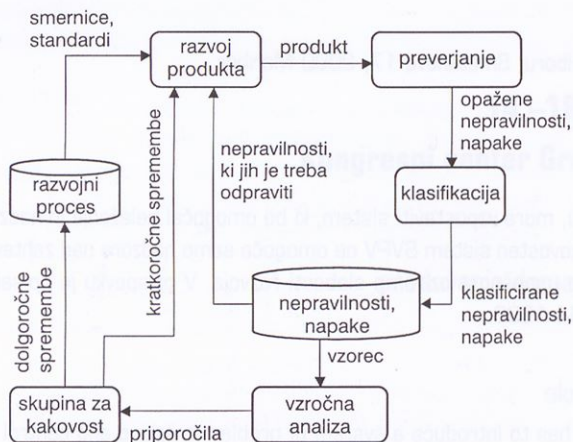
Ker zahtevek sproži niz aktivnosti, ki lahko pomembno vplivajo na stroške, roke in kakovost sistema, je zelo pomembno, da imamo dober nadzor nad celotno življenjsko dobo zahtevka. Glede na razvojni cikel se spremembe sistema pojavljajo predvsem pri preverjanju in vzdrževanju. Vzroke za spremembo strukture sistema lahko uvrstimo v naslednje kategorije: nepravilno delovanje, izboljšava, prilagajanje spremembam v okolju in nove naročnikove zahteve. Prvi dve kategoriji kakovost dvigujeta, medtem ko jo zadnji dve samo ohranjata. Ko so vzroki za zahtevek odpravljeni, se konča njegov življenjski cikel.

Vsaka organizacija, ki želi delovati v skladu s smernicami glede kakovosti, mora vzpostaviti sistem, ki bo omogočal beleženje in nadzor nad nepravilnostmi ter

spremembami. Ta sistem bomo v nadaljevanju imenovali SVPV (sistem za vodenje popraviljanja in vzdrževanja). Ker je kakovost produkta odvisna tudi od kakovosti procesa, se to odraža v nekaterih standardih, ki so povezani s kakovostjo. ISO 9001:2000 (sekcija 4.2.4 Obvladovanje zapisov) zahteva od dobavitelja, da mora vzpostaviti in vzdrževati postopke za prepoznavanje, zbiranje, urejanje, dopolnjevanje, shranjevanje, vzdrževanje in odstranjevanje zapisov o kakovosti. V ISO 9000-3 (Poglavje 5.7.3 Preskušanje) je zahtevano, da mora biti vsaka zaznana težava zapisana in njen status nadzorovan, dokler ni odpravljena. Tudi zrelostni model CMM (Capability Maturity Model) govori o procesiranju zapisov o ugotovljenih nepravilnostih. Nivo 2 zahteva samo beleženje in nadzor nad odpravo nepravilnosti, medtem ko nivo 3 zahteva še analizo nepravilnosti. Z vzročno analizo<sup>1</sup> najdenih nepravilnosti in napak lahko dobimo vpogled v slabosti razvojnega procesa (glej zgled na preglednici 1). Na nivoju 4 je dodana še ocena napovedovanja zelo verjetnih nepravilnosti. Najvišji nivo (5) zahteva, da morajo biti vzpostavljene tudi aktivnosti za preprečevanje nepravilnosti.

<sup>1</sup> Več o tej problematiki glej v članku [CARD, 1998].

Slika 1 prikazuje tipični razvojni proces, v katerem sistematično zapisujemo zahteve. Ker je sistem za zapisovanje zahtevkov ključni element, od katerega so odvisne druge aktivnosti, se bomo v tem prispevku osredotočili predvsem na njegovo problematiko.



Slika 1. Z vzročno analizo lahko izboljšujemo kakovost razvojnega procesa

Na tržišču obstaja niz sistemov<sup>2</sup> SVPV. To so podatkovno podprte aplikacije, ki so namenjene nadzoru nad popraviljanjem in vzdrževanjem. Ker gre za relativno preproste podatkovne baze, so mnoge organizacije izdelale svoje sisteme SVPV. V večini primerov imajo dva ločena sistema za nadzor nad spremembami; prvi se uporablja v fazi razvoja in preverjanja, drugi pa v fazi validacije in kasnejšega vzdrževanja. Za učinkovito obvladovanje kakovosti je smiselno, da vsaj delno združimo oba, saj je mnogo aktivnosti enakih.

Na kratko bomo opisali potek zapisovanja in opozorili na probleme, ki smo jih srečali pri implemen-

Aktivnost	Projekt 1	Projekt 2	Projekt 3
Pregled	10 %	10 %	2 %
Presoja	10 %	8 %	15 %
Inšpekcija	20 %	22 %	18 %
Testiranje	40 %	38 %	5 %
Validacija	5 %	2 %	10 %
Uporaba	15 %	20 %	50 %

Preglednica 1. Delež odkritih nepravilnosti po posameznih aktivnostih. Primerjava uspešnosti preverjalnih metod kaže, da je zaradi neučinkovitega testiranja narastel delež odkritih nepravilnosti v fazi uporabe.

<sup>2</sup> Na internetu jih iščemo z naslednjimi imeni: Problem management system, Bug tracking system, Incident tracking system, Problem tracking system.  
<sup>3</sup> Podrobnejši opis sistema SVPV, ki temelji na IEEE 1044-1993, je na naslovu: <http://saturn.uni-mb.si/č cv>.

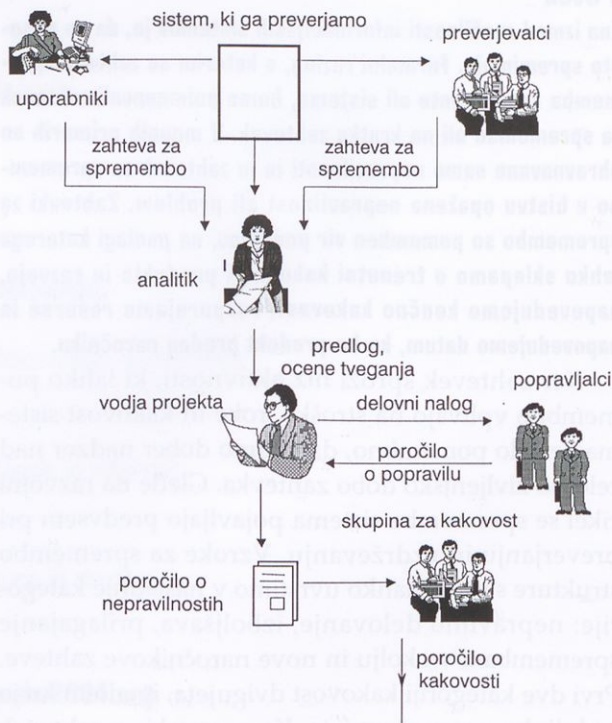
taciji lastnega sistema SVPV<sup>3</sup>. Mnogo praktičnih napač se nahaja v [KANER,1993]. Ker je standard IEEE 1044-1993 eden redkih, ki lahko pomaga organizacijam pri vzpostavitvi sistemov SVPV, se bomo v prispevku nanj pogosto sklicevali.

## 2 Povezava med preverjanjem, popravljanjem in uporabo

V razvoju in kasnejšem vzdrževanju se sistem neprestano spreminja. Namen spremembe sistema lahko uvrstimo v naslednje kategorije: nepravilno delovanje, izboljšava, prilagajanje spremembam v okolju in nove naročnikove zahteve.

Preverjanje, popravljanje in nadzor nad konfiguracijo so mnogokrat popolnoma ločene aktivnosti v življenjskem ciklusu, za katerega skrbi skupina za kakovost (slika 2). Med osebami, ki sodelujejo v teh aktivnostih, poteka izmenjava številnih informacij v obliki pisnih ali ustnih sporočil.

Glede na stopnjo računalniške podpore in organiziranosti preverjanja ter popravljanja obstajajo naslednje stopnje beleženja zahtevkov:



Slika 2: Povezava med preverjanjem, popravljanjem in uporabo

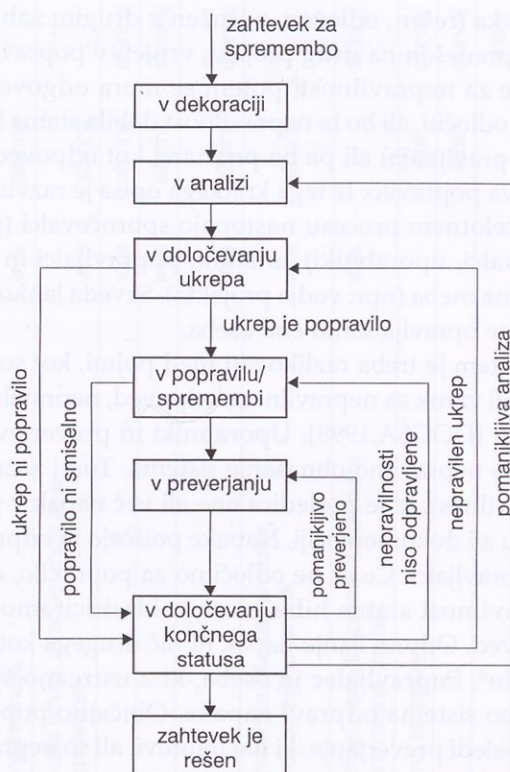


- Kaotični proces. Ni sistematičnega preverjanja. Avtorji sami preverjajo in sproti popravljajo. Če preverjanje izvaja posebna skupina, je komunikacija med preverjevalci in popraviljalci ustna. Ni nobenih formalnih zahtevkov in pisnih poročil o najdenih neustreznostih in o popravilu.
- Papirni proces. Skupina, ki preverja, zapisuje svoje ugotovitve na posebne formularje, ki jih pošiljajo popraviljalcem. Vodja ima delni pregled nad obema skupinama. Ne analizira se učinkovitost popraviljanja in preverjanja, saj analize zahtevajo dodatno delo. Uporabniki sporočajo opažene neustreznosti ustno, po telefonu ali elektronski pošti.
- Avtonomno računalniško podprti procesi. Preverjevalci uporabljajo posebna računalniško podprta orodja, ki omogočajo učinkovito preverjanje. Podobno velja za popraviljalce. Vsak ima svojo podatkovno bazo. Med bazami ni neposredne povezave. Uporabniki sporočajo opažene neustreznosti pisno ali po elektronski pošti. Tudi ti podatki se vnašajo v podatkovno bazo.
- Centraliziran računalniško podprt proces. Sistem SVPV ima centralni repozitorij, ki ga koristijo vsa orodja, ki jih uporabljamo v razvoju. Uporabniki neposredno vpisujejo opažene nepravilnosti. Vodja ima zelo dober nadzor nad potekom popraviljanja in preverjanja. Zasedovati je mogoče status vsake vpisane nepravilnosti. Izvajajo se razne analize, ki kažejo na kakovost preverjanja, popraviljanja in tudi samega produkta. Z rezultati analiz lahko prognoziramo trenutne trende in racionalneje razporejamo resurse. Slabost: velika odvisnost od centra.

Končni (idealni) cilj je računalniško integriran razvoj, ki povezuje vse procese v celoto. Slabost tega pristopa so veliki stroški. V primeru, da že uporabljamo določena razvojna orodja, jih zelo težko vključimo v nov sistem. Potreben je nakup novih orodij in dodatno usposabljanje. Ker gre za zelo radikalne spremembe, povezane z velikim tveganjem, lahko pričakujemo, da se bodo le redki takoj odločali za centraliziran proces.

### 3 Tipični potek procesiranja zahtevka

Ena izmed zelo pomembnih entitet v podatkovnem modelu sistema SVPV je zahtevek za spremembo. Nanaša se na točno določen objekt ali sistem oziroma na njegovo verzijo. Obstaja več postopkov procesiranja zahtevka. Večina ustreza modelu, ki ga prikazuje slika 3. Zahtevek za spremembo, ki je lahko obravna-



Slika 3: Popraviljanje oziroma spreminjanje sistema, prikazano z diagramom prehajanja stanj. Najpogostejša je popolnoma navpična pot.

van kot izboljšava, nepravilnost ali sprememba specifikacij, sproži naslednje štiri aktivnosti:

- sprejem in opis zahtevka za spremembo,
- analizo zahtevka,
- odločitev o ukrepih za odpravo zahtevka,
- odločitev o končnem statusu zahtevka.

Vse te aktivnosti so lahko eksplicitno ločene ali pa so združene in se navzven kažejo kot samo ena aktivnost. Začetna aktivnost je sprejem zahtevka. Oseba, ki zahteva spremembo, navede namen zahtevka (izboljšava, odprava nepravilnosti ali sprememba specifikacij), na kratko opiše problem in ga klasificira glede na izbrani taksonomijski sistem. Nato sledi analiza zahtevka, kjer analitik podrobneje preuči zahtevek in poda predlog o rešitvi ter poišče vzroke, ki so povzročili nastanek zahtevka. Odločiti se mora, ali gre morebiti za napačno uporabo ali se obravnavani zahtevek že obdeluje, kakšen vpliv ima zahtevek na kakovost sistema, oceniti mora razna tveganja (popraviti ali ne popraviti). O predlaganih ukrepih odloča pristojna oseba. Sledi izvajanje tega predloga (npr. popravilo) in nato odločitev o končnem statusu

zahtevka (rešen, odložen, združen z drugim zahtevkom, prenešen na drug projekt, vrnjen v popravilo). Če gre za nepravilnost, potem se mora odgovorna oseba odločiti, ali bo ta nepravilnost dobila status hibe (ni popraviljanja) ali pa bo priznana kot odpoved, ki zahteva popravilo. Iz tega kratkega opisa je razvidno, da v celotnem procesu nastopajo sporočevalci (preverjevalci, uporabniki), analitik, popraviljalci in odgovorna oseba (npr. vodja projekta). Seveda lahko vse te vloge opravlja samo ena oseba.

Pri tem je treba razlikovati med pojmi, kot so napaka ali vzrok za nepravilnost, odpoved, nepravilnost in hiba [DOGŠA,1998]. Uporabniki in preverjevalci opazijo nepravilno obnašanje sistema. Torej zaznajo nepravilnost, ki je posledica ene ali več napak v programu ali dokumentaciji. Napake poiščejo in odpravijo popraviljalci. Če se ne odločimo za popravilo, dobi nepravilnost status hibe, sicer jo klasificiramo kot odpoved. Odpravljanje napak ni nič drugega kot popravilo<sup>4</sup>. Popraviljalec je oseba, ki z ustrezno spremembo sistema odpravi napake. Običajno popraviljanju sledi preverjanje, ki naj ugotovi, ali so nepravilnosti odpravljene.

V vsakem izmed štirih korakov (sprejem, analiza, ukrepi in odločitve o končnem statusu) se izvajajo tri aktivnosti: opisovanje, klasificiranje in ocenjevanje vpliva na kakovost in ocena tveganja. Te aktivnosti se lahko v posameznih korakih ponavljajo oziroma dopolnjujejo; npr. ko nepravilnost odkrijemo, približno določimo domnevni vzrok. Kasneje, ko nepravilnost podrobneje analiziramo, lahko tudi natančneje določimo te vzroke. Ker vsako spreminjanje sistema zahteva tudi ustrezno zapisovanje sprememb, se ves postopek še bolj zaplete, če vzdržujemo več različnih verzij.

Ena izmed najpomembnejših aktivnosti je klasifikacija nepravilnosti in napak. Pri implementaciji sistema SVPV je smiselno, da se odločimo za ustrezen enoten klasifikacijski sistem. Takemu sistemu pravimo tudi taksonomijski sistem. Le če uporabimo enotno klasifikacijo, lahko izvajamo razne primerjave in analize o učinkovitosti testiranja, popraviljanja in kakovosti sistema (glej preglednico 2). Najbolj znani klasifikacijski sistemi so: ortogonalni klasifikacijski sistem<sup>5</sup>, ki ga je razvil IBM [CHILLAREGE,1992], Hewlett-Packardov

sistem [PFLEEEGER,1998] in sistem, ki ga definira standard IEEE 1044 [IEEEstd,1993]. Slednjega bomo v nadaljevanju na kratko opisali.

V mnogih primerih podatkovna aplikacija omogoča spreminjanje klasifikacijskega sistema. To pa ne velja za procesni vidik (slika 3), ki je v večini primerov vgrajen v sistem SVPV.

#### 4 Standard IEEE 1044-1993

Standard IEEE 1044-1993 [IEEEstd, 1993] zelo dobro opisuje potek vodenja popraviljanja in vzdrževanja informacijskih sistemov. Kljub temu, da je poudarek na programski opremi, se lahko z majhno prilagoditvijo uporablja tudi za strojno opremo. Standard opisuje postopek, ki smo ga na kratko povzeli v prejšnjem poglavju. Vodenje popraviljanja in vzdrževanja sistemov je sestavljeno iz štirih zaporednih aktivnosti:

- zaznave nepravilnosti (RR Recognition),
- analize nepravilnosti (IV Investigation),
- ukrepov za odpravo nepravilnosti (AC Action),
- določitve končnega statusa nepravilnosti (DP Disposition).

Glavnina standarda opisuje razne klasifikacijske tabele (preglednica 2), ki so v nekaterih primerih zelo podrobne. Za opis vsakega zahtevka imamo na razpolago 21 atributov, obveznih jih je 11. Za vsako aktivnost je predvidena posebna neobvezna priloga (npr. ekranska slika). Vse priloge so opisane s 113 atributi. Če želimo uporabiti standard v polnem obsegu, je treba izpolniti približno<sup>6</sup> 134 polj. Vrednosti atributov, s katerimi opisujemo zahtevke, črpamo iz klasifikacijskih tabel. Na ta način je možna primerjava med podatki, ki jih dobimo pri različnih projektih in uporabnikih. Z velikostjo klasifikacijske tabele se večja natančnost opisa in hkrati manjša preglednost ter večja napor, ki ga je treba vložiti v opis zahtevka. Najobširnejša je tabela, s katero opisujemo napako. Zaradi preglednosti je sestavljena iz dveh nivojev. Klasifikacijske tabele, ki so označene poudarjeno (domnevna domena vzroka, opis simptoma, dejanska domena vzroka, opis vzroka), so odvisne od vrste sistema. Če gre za strojno ali kombinirano opremo, jih je treba ustrezno spremeniti.

V zadnjih štirih kolonah v preglednici 2 je opisano, kdaj naj določen podatek vpišemo oziroma ažuriramo.

<sup>4</sup> Pogosto se po nepotrebnem uporablja izraz razhroščevanje (angl. debugging).

<sup>5</sup> ODC Orthogonal Defect Classification.

<sup>6</sup> Všteti niso atributi, ki se nanašajo na procesiranje zahtevka.

Kategorije (atribut)	Obvezen	Velikost tabele (število atributov)	Minim. velikost tabele	Prevod	Komentar	Sprejem	Analiza	Sprejeti ukrep	Končni status
Project activity	da	9	9	Preverjevalna aktivnost	Aktivnost, s katero je bila odkrita nepravilnost.	vnos			
Project phase	da	24	6	Faza v projektu	Faza, v kateri se produkt z nepravilnostjo trenutno nahaja.	vnos			
<b>Suspected cause</b>	ne	26	0	<b>Domnevna domena vzroka</b>	Približna domnevna lokacija vzroka (program, strojna oprema)	vnos	ažuriraj		
Repeatability	ne	5	0	Pogostost	Pogostost pojavljanja nepravilnosti.	vnos	ažuriraj		
<b>Symptom</b>	da	18	9	<b>Opis simptoma</b>		vnos	ažuriraj		
Product status	ne	4	0	Uporabnost produkta	Uporabljivost produkta, če ga ne spreminjamo.	vnos	ažuriraj		
Severity	da	5	5	Resnost nepravilnosti	Resnost z v vidika strokovnjaka.	vnos	ažuriraj		
Priority	ne	5	0	Prioriteta popravila		vnos	ažuriraj	ažuriraj	ažuriraj
Customer value	ne	6		Vpliv na zadovoljstvo uporabnika	Kako vpliva odkrita nepravilnost na zadovoljstvo uporabnika?	vnos			
Mission/Safety	ne	5	0	Vpliv na uspešnost projekta		vnos	ažuriraj	ažuriraj	ažuriraj
<b>Actual cause</b>	da	26	6	<b>Dejanska domena vzroka</b>	Približna lokacija vzroka (program, strojna oprema)	vnos		ažuriraj	
<b>Source</b>	da	28	7	<b>Poreklo vzroka (napake)</b>	Vmesni produkt v razvoju (specifikacije, koda), v katerem se nahaja napaka, ki je vzrok za nastanek nepravilnosti.	vnos		ažuriraj	
<b>Type</b>	da	80	9	<b>Opis vzroka (napake)</b>	Podroben opis vzroka, zaradi katerega se je pojavila nepravilnost.	vnos		ažuriraj	
Project schedule	da	4	4	Vpliv na termine projekta		vnos		ažuriraj	ažuriraj
Project costs	da	4	4	Vpliv na stroške projekta		vnos		ažuriraj	ažuriraj
Project risk	ne	4	0	Tveganje v primeru popravila /spremembe produkta		vnos		ažuriraj	ažuriraj
Project quality/reliability	ne	4	0	Vpliv na kakovost/zanesljivost projekta	Ocena spremembe kakovosti produkta, če ga bomo popravili/spremenili.	vnos		ažuriraj	
Societal	ne	4	0	Vpliv na mnenje širše družbe	Kako bi vplivalo popravilo/sprememba produkta na negativno mnenje družbe?	vnos			ažuriraj
Resolution	da	21	4	Ukrep	Ukrepi, ki naj se izvedejo.	vnos		vnos	ažuriraj
Corrective action	da	23	0	Bodoča preventiva		vnos		vnos	ažuriraj
Disposition	da	9	9	Končni status		vnos		vnos	vnos

Preglednica 2: Najpomembnejše klasifikacijske tabele (standard IEEE 1044-1993)

## 5 Sklep

Najosnovnejši cilj sistema SVPV je nadzor nad odpravljanjem nepravilnosti. Pravilno izbran klasifikacijski sistem omogoča analize, s katerimi lahko ugotovljamo uspešnost oziroma slabosti razvoja. Ker zahteva uporaba sistema SVPV dodaten napor sodelujočih, je zelo pomembno, da je le-ta kakovosten v smislu implementacije in vsebine. Le na videz je sistem SVPV zelo preprosta podatkovna aplikacija, ki jo lahko napravimo v accessu v nekaj dneh. Vzdrževanje sistema SVPV obsega popravljanje in širjenje klasifikacijskih tabel ter spreminjanje postopka procesiranja zahtevka (sliko 3). Za slednjega velja, da je izvedljiv samo, če imamo dostop do izvorne kode. To je ena izmed največjih slabosti kupljenih sistemov SVPV. Za konec bomo na kratko opisali najpomembnejše lastnosti, ki jih mora imeti sistem SVPV.

*Ponovljivost* je lastnost zahtevka, ki omogoča popolno rekonstrukcijo vzroka, zaradi katerega je bil podan zahtevek. Za zgled pogledjmo zahtevek, ki obravnava opažene nepravilnosti v delovanju modula. Če je na podlagi informacij, ki so v zahtevku, vedno mogoče povzročiti opisano nepravilno delovanje modula, pravimo, da smo zadostili kriteriju ponovljivosti. Če zahtevek ni ponovljiv, je v splošnem zelo težko odpraviti njegove vzroke. To še posebej velja, kadar ni materialnih dokazov za nepravilno delovanje. Isto velja za ponovljivost klasificiranja. Neka druga oseba mora isti zahtevek klasificirati popolnoma enako.

*Uporabnost* sistema je merilo napora, ki ga moramo vložiti v to, da lahko sistem učinkovito uporabljamo. Problemi se pojavljajo predvsem pri raznih klasifikacijah. Podrobnejše kot so klasifikacijske tabele, težja je klasifikacija. S pregrobo granulacijo lahko izgubimo informacije, ki jih potrebujemo pri analizi uspešnosti razvojnega procesa. Problematični so predvsem pokrelo, domena in opis vzroka. Če želimo zagotoviti ponovljivost klasifikacije, je pri slabo zastavljenem klasifikacijskem sistemu potreben velik napor. V praksi se zato pogosto dogaja, da je klasifikacija izvedena površno. V večini primerov so sistemi SVPV tako zahtevni, da je potrebno posebno izobraževanje. Zelo pomem-

ben je procesni vidik podatkovne aplikacije, katerega je kasneje težko spreminjati.

Ker so zahtevki povezani s stroški, ki nastanejo z odpravljanjem vzrokov, nas zanima, kdo pošilja zahteve. *Avtentičnost* je lastnost, na podlagi katere lahko identificiramo pošiljatelja. V veliki večini primerov firme ne želijo javno objavljati ugotovljene nepravilnosti, saj bi lahko konkurenti te podatke izkoristili. Ker je pri preverjanju in popravljanju potrebna *diskretnost*, morajo biti podatki zaščiteni pred nepooblaščenim branjem oziroma spreminjanjem.

Ker je preverjanje povezano z velikimi stroški, se noben zahtevek ne sme izgubiti. Izgubljen zahtevek pomeni ponovitev preverjanja. To lastnost bomo pomenovali *vestnost*. Lahko se zgodi, da so nekateri zahtevki na prvi pogled nesmiselni oziroma nepomembni. Ignoriranje zahtevka je resna odločitev, za katero se lahko odloči lahko le odgovorna oseba. Običajno je vodja projekta oseba, ki nadzoruje reševanje zahtevkov in tudi nosi *odgovornost* za potek popraviljanja.

Sistem za vodenje popraviljanja in vzdrževanja sistemov bo *učinkovit*, če bomo z njim zmanjšali stroške vzdrževanja.

## 6 Literatura

- [CARD,1998] David N. Card:  
"Learning from our Mistakes with Defect Causal Analysis",  
IEEE Software, januar/februar, 1998, str. 56–63.
- [CHILLAREGE,1992] Ram Chillarege in drugi:  
"Orthogonal defect classification—A concept for in-process  
measurement", IEEE Transaction on Software Engineering,  
VOL 18, številka 11, 1992, str. 943–956.
- [DOGŠA,1998] Tomaž Dogša:  
"Gostota napak in odpovedi – problematično merilo  
kakovosti", Uporabna informatika, šte. 2, letnik VI, 1998,  
str. 20–25.
- [IEEEstd,1993] "IEEE Std 1044-1993:  
IEEE Standard Classification for Software Anomalies", The  
Institute of Electrical and Electronics Engineering, Inc., 1993.
- [KANER,1993] Cem Kaner, Jack Falk, Hung Quoc Nguyen: "Testing  
Computer Software", Van Nostrand Reinhold, 1993.
- [PFLEEGER,1998] S. L. Pfleeger:  
"Software Engineering, Theory and Practice", Prentice Hall, Inc.

Dr. Tomaž Dogša je izredni profesor na Fakulteti za elektrotehniko, računalništvo in informatiko v Mariboru, kjer predava na dodiplomski in dodiplomski stopnji in vodi Center za verifikacijo in validacijo sistemov. Na raziskovalnem področju se ukvarja predvsem z V&V tehnologijo oziroma testirnimi orodji.

# Reševanje problema pomanjkanja unikatnih naslovov internetnega protokola

Stjepan Pervan  
IskraTEL  
pervan@iskratel.si

Martina Kern  
Osnovna šola Staneta Žagarja Kranj  
martinakern@hotmail.com, martina.kern@guest.arnes.si

## Povzetek

Običajno telekomunikacijsko omrežje zamenjuje nova generacija omrežij, ki slonijo na internetnem protokolu. Velik porast spletnih storitev nas je pripeljal do vse pogostejše uporabe interneta. Trenutna verzija internetnega protokola, imenovanovanega IPv4 (internetni protokol verzije 4), predstavlja hrbtnico interneta. Zaradi široke uporabe interneta je postalo jasno, da je neprimeren predvsem zaradi problema premajhnega števila prostih mest za naslavljanje, nezadostne varnosti in slabe kakovosti prenosa. V članku je prikazan problem pomanjkanja unikatnih naslovov internetnega protokola Ipv4, trenutni načini reševanja teh težav in kratek opis novega internetnega protokola IPv6 z vsemi njegovimi prednostmi, kakor tudi vpliv omenjene novosti na gospodarstvo in nadaljnji razvoj informatike v Sloveniji.

## Abstract

### Solving the Problem of the Lack of Unique internet protocol Addresses

The common telecommunication networks have been replaced by the new generation of networks based on the internet protocol. A huge increase in the number of web services has lead us to more and more frequent use of the Internet. The internet protocol we use at present (IPv4 – internet protocol version 4) represents the backbone of the Internet. Increasingly wider use of the Internet shows us that Ipv4 has become inappropriate mainly because of the shortage of available address space, insufficient security and poor transfer quality. This article describes the lack of the unique internet protocol addresses, the ways of solving this problem and a short description of the new internet protocol with all its advantages as well as the impact on Slovenian economy and further development of informatics.

## Uvod

Le redkokdo si je pred kakimi desetimi leti lahko predstavljajal, da bo internet zaživel v takšnem obsegu. Vedno več podjetij in posameznikov ga redno uporablja pri svojem delu in večina podjetij si težko predstavlja, kako bi poslovanje uspešno potekalo brez njega. Njegova uporaba se je v zadnjih nekaj letih tako strahovito razmahnila, da ga danes aktivno uporablja že več kot štiristo milijonov uporabnikov po vsem svetu. Ob tem se nam kar samo od sebe zastavlja vprašanje, kako vsem telekomunikacijskim napravam, kot so "pametni" telefoni, računalniki, mrežne naprave, igralne naprave, sistemi za nadzor avtomobilov, ladij, letal, zagotoviti dovolj unikatnih globalnih naslovov internetnega protokola. Evropska komisija je nedavno sporočila, da bo protokol IPv4 zadostoval predvidoma le še do leta 2005, ko naj bi pričelo primanjkovati IP naslovov. Ko so pred več kot dvajsetimi leti oblikovali internetni protokol IPv4, so predvideli le nekaj več kot

štiri milijarde možnih internetnih naslovov, pa še te so neenaakomerno porazdelili. Omejitve naslovov v različici IPv4 so sicer svetovni problem, a za Evropo toliko hujši, ker bi to omejilo razmah novih generacij brezžičnih tehnologij in komunikacij.

Pri teh oblikah telefonije je Evropa namreč v prednosti pred ostalim svetom, a ta prednost lahko ob najmanjši oviri hitro skopni. Na tem področju se je torej najprej pojavila potreba po pospešeni pripravi nadgradnje internetnega protokola IPv4. Kot odgovor na te potrebe je Internet Engineering Task Force (IETF), odbor, zadolžen za tehnične odločitve za internetni protokol (IP) in internet, v juliju 1992 podal zahtevo po predlogih za novo verzijo internetnega protokola Ipvng (IP Next generation). Prispelo je veliko število predlogov in leta 1994 je bila dosežena ustrezna oblika protokola.

Glavni mejnik predstavlja publikacija Request For Comments (RFC) 1752 "Recommendation for the IP Next Generation", izdana januarja 1995. V njej so opisane zahteve za nov IP, določena je oblika protokola, način naslavljanja, usmerjanja in varnost. Številni drugi RFC dokumenti specificirajo podrobnosti protokola, sedaj uradno imenovanega IPv6. Omogoča dvesto šestinpetdeset sekstilijonov IP naslovov, obenem pa podpira več kot milijardo uporabnikov. Ta veliko enostavnejši protokol zagotavlja tudi večjo varnost, manjše usmerjevalne tabele, podporo različnim tipom storitev in mobilnim uporabnikom.

### Trenutno stanje in zahteve za nov internetni protokol

IP je v 90-ih postal široko sprejet protokol in obenem edini protokol za prenos podatkov, zvoka in video prometa. Glede na vedno večje število uporabnikov interneta je prišlo do problema, kako zagotoviti dovolj unikatnih globalnih naslovov IP za vse naprave, ki pri svojem delovanju uporabljajo povezavo z internetom. Ustvarjalci IPv4 so namreč menili, da bo IP naslovov zadosti in da bodo zadovoljili vsem potrebam, saj imamo lahko sedaj teoretično 4.294.967.296 unikatnih naslovov IP. Vendar je dejansko število manjše zaradi razdeljenosti naslovov v razrede, ki so namenjeni raznim testiranjem in drugim specialnim namenom. Tako lahko v omrežju uporabljamo le tri razrede – A, B in C:

- razred A (125 omrežij) – približno 16 milijonov gostiteljev na omrežje,
- razred B (16382 omrežij) – 65534 gostiteljev na omrežje,
- razred C (2 milijona omrežij) – 254 gostiteljev na omrežje.

Razred D je namenjen za skupinsko naslavljanje, medtem ko je razred E namenjen raziskovalnim namenom. Ti razredi se razlikujejo glede na število omrežij in gostiteljev.

Z eksponentno rastjo globalnega interneta in razvojem novih aplikacij, kot so večpredstavnost, video konference in tretja generacija brezžičnih tehnologij, so se pojavile tudi potrebe po novih karakteristikah internetnega protokola. Pojavljajo se trije pglavitni problemi IPv4:

- premajhen prostor za naslavljanje oz. neučinkovita shema naslavljanja IP naslovov;
- nizka stopnja varnosti;
- neustrezna, neučinkovita kakovost prenosa informacij ter slaba podpora novejšim tehnologijam,

kot so realni čas prenosa podatkov, avdio in video. Očitna rešitev je bila izdelava novega internetnega protokola, ki naj bi omogočal več razpoložljivih naslovov in bi zadostil naslednjim zahtevam [12]:

- podpora več kot milijardi uporabnikov,
- povečanje naslovnega prostora in zmanjšanje usmerjevalnih tabel,
- varnost na nivoju IP,
- podpora različnim tipom storitev,
- boljša podpora za mobilne gostitelje in multimedijo,
- istočasno delovanje IPv6 in IPv4 vsaj še nekaj let,
- mora biti robusten in skalabilen,
- omogočati mora avtomatsko konfiguracijo.

S tem namenom je bil razvit protokol IPv6 ali IPng, vendar bo potrebnih kar nekaj let za njegovo implementacijo, saj zahteva določene modifikacije na celotni internetni infrastrukturi.

### Mehanizmi za zmanjševanje porabe naslovnega prostora IPv4

Velik problem se pojavi pri aplikacijah, ki delujejo v realnem času in prenosih multimedijskega prometa na IP obstoječih omrežjih. Jasno je, da samo dvočkovna komunikacija med dvema IP napravama ne prinaša zaželenega efekta. To dosežemo le z novim protokolom, ki ima dovolj naslovnega prostora.

Veliko zaskrbljenost povzroča tudi unikatno naslavljanje domačih naprav, IP telefonov in drugih naprav. Dokler te naprave dobivajo unikatne naslove od ponudnikov internetnih storitev, ni težav. Problem nastane, ko hočemo imeti efektivno rešitev za domače uporabnike, ki potrebujejo več unikatnih IP naslov, ki naj bi se ne spreminjali. V ta namen so začeli domači uporabniki uporabljati privatno shemo za naslavljanje in posebne naprave, ki omogočajo komunikacijo iz njihovega omrežja v internet (enosmerna komunikacija), kar predstavlja globalni problem. Definirano je bilo precej mehanizmov za zmanjšanje porabe naslovnega prostora IPv4, vendar imajo vsi takšne ali drugačne pomanjkljivosti. Ti mehanizmi so:

- NAT,
- RSIP,
- VPN,
- Proxy.

### NAT (Network Address Translation)

NAT se uporablja v napravah, ki predstavljajo mejo med javnim internetom in zasebnimi (notranjimi) lokalnimi omrežji (slika 1). Notranje omrežje (podjetniško)

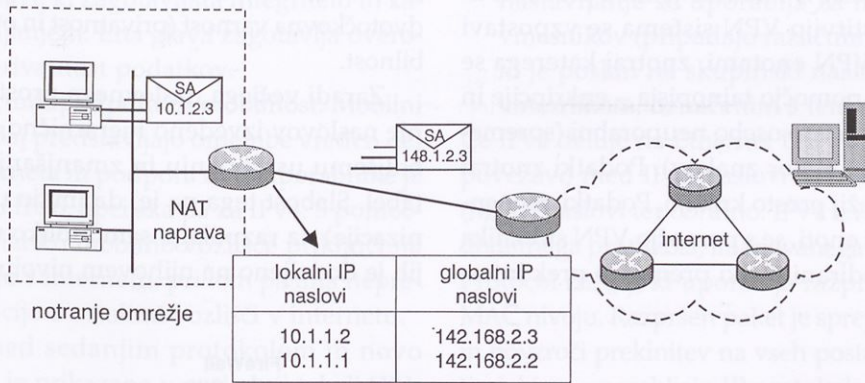
uporablja privatne IP naslove, zunanje (internet) pa globalne naslove, s katerimi notranjim napravam omogoča, da komunicirajo z zunanjim svetom; obratno ni mogoče. Ko IP paketi iz zasebnega omrežja prečkajo prehod, NAT prevede zasebni IP naslov in številko vrat v javni IP naslov in številko vrat ter skrbi, da posamezne seje ostanejo ločene. NAT naprave morajo poleg prevajanja IP naslovov in vrat poznati protokole, da lahko izpisujejo, uporabljajo oz. spreminjajo vse potrebne informacije za vzpostavljanje uspešne komunikacije z ostalimi napravami. [4]

Vendar ima ta mehanizem tudi nekaj slabih strani. NAT mehanizmi namreč onemogočajo dvotočkovno (end-to-end) komunikacijo s postavitvijo NAT naprave med dva komunikacijska dela. Na osnovi tega ne moremo implementirati dvotočkovne varnostne sheme. Podpora za nikoli dokončan multi-medijski protokol zahteva konstantne nadgradnje programskih paketov, kar pa tudi vodi v možno nestabilnost NAT naprav. Prav tako NAT lahko prepreči številne zanimive uporabe računalnikov v omrežjih, kot so igre z več igralci, komunikacije v realnem času

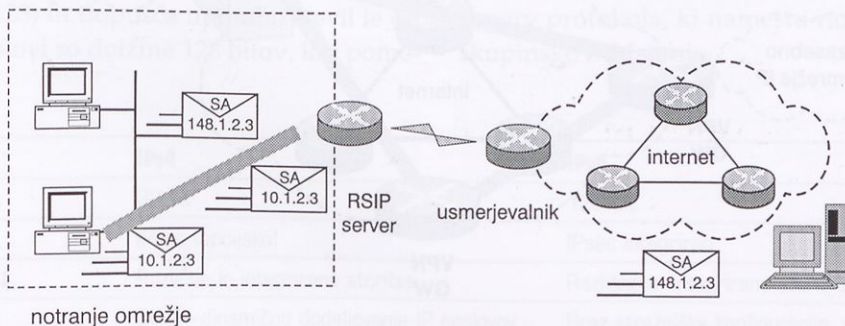
in druge storitve "enak z enakim", ki jih vse več uporabnikov želi uporabljati doma ali v majhnih podjetjih. Te aplikacije ne bodo delovale, dokler se bodo v javnem internetu uporabljali zasebni naslovi hkrati z isto številko vrat. Uporabljati morajo namreč javni naslov in enolično številko vrat za vsako posamezno sejo. [4]

**RSIP (Realm Specific IP)**

Omogoča drugačen pristop k reševanju naslavljanja IP kot NAT. Ena največjih pomanjkljivosti NAT-a je, da IP naslov ne predstavlja končne točke povezave. S pomočjo RSIP-a se ta pomanjkljivost lahko odpravi, tako da omogoča napravam s privatnimi naslovi IP direktno komunikacijo z ostalimi napravami s pomočjo uporabe začasnih globalnih IP naslovov. IP paketi prehajajo iz izvora do cilja brez kakršnih koli sprememb. Ko hoče naprava komunicirati z javnim IP naslovom, se registrira na strežniku RSIP. Strežnik RSIP dodeli unikatnen javni IP naslov in poveže privatni naslov z RSIP gostiteljem na ta globalni naslov. RSIP naprave uporabljajo ta globalni naslov za pošiljanje



Slika 1: Delovanje NAT naprave



Slika 2: Delovanje RSIP

paketov k drugim globalnim ciljem, dokler najem ne preteče, oziroma dokler se ne obnovi. RSIP naprave ne morejo direktno pošiljati paketov z nizom globalnih naslovov, ker morajo najprej poslati pakete RSIP strežnikom. Prav tako se lahko RSIP uporablja za prenašanje prometa med več privatnimi omrežji [5]. RSIP naprava lahko najame nekaj različnih naslovov, da doseže različna ciljna omrežja. RSIP za tuneliranje uporablja različne tunnelske protokole, kot so IP in IP, Generic Routing Encapsulation (GRE) [7] ali Layer 2 Tunneling Protocol (L2TP) [8].

### VPN (Virtual Private Network)

Navidezno zasebno omrežje je sistem, s katerim vzpostavimo varno povezavo za prenos podatkov med dvema ali več točkami, pri katerih za medsebojno povezavo uporabimo internet (slika 3). Najbolj razširjena protokola, ki se uporabljata v navideznih zasebnih omrežjih sta L2TP in IPSec [18]. Podjetja se srečujejo z vse večjo potrebo po medsebojni povezavi svojih poslovnih enot, obenem pa tudi s potrebo po povezavi s strateškimi partnerji. Internet je odprto omrežje, zato je neposreden prenos pomembnih podatkov prek le-tega nezanesljiv. Podatke je moč prestreči in zlorabiti. Z namestitvijo VPN sistema se vzpostavi varen kanal med VPN enotami, znotraj katerega se podatki pošiljajo s pomočjo tajnopisja – enkripcije in so tako za nepooblaščen osebo neuporabni (spremenjeni v neuporaben nabor znakov). Podatki znotraj obeh lokalnih omrežij prosto krožijo. Podatki, namenjeni drugi poslovni enoti, se s pomočjo VPN strežnika kodirajo in se v kodirani obliki prenesejo prek inter-

neta do VPN strežnika druge poslovne enote, kjer se podatki dekodirajo in jih je mogoče ponovno uporabljati znotraj drugega omrežja. Kodiranje poteka v realnem času in ne obremenjuje hitrosti pretoka podatkov. To nam omogoča združitev dveh lokalnih omrežij (LAN) in povezavo le-teh v eno veliko omrežje (WAN).

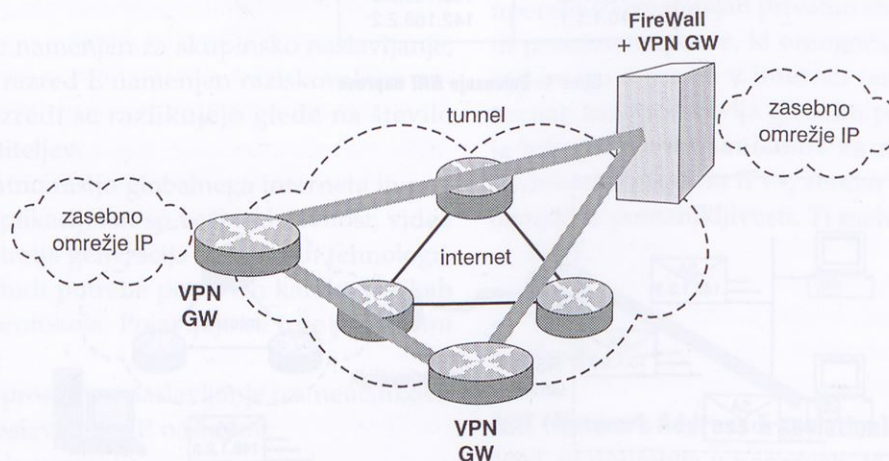
### Proxy (Posredniški strežnik)

Še ena izmed rešitev, ki omogočajo reševanje problema naslavljanja, je uporaba proxy strežnikov. Te naprave delujejo podobno kot NAT naprave, a so veliko bolj enostavne, ker podpirajo samo nekatere protokole (HTTP, SIP, H.323). Naprave se obnašajo kot namestniki za aplikacije, tako da izmenjujejo vse podatke z oddaljenim sistemom v njihovem imenu.

### Globalna rešitev – IPng oz. IPv6

IPng oz. IPv6 je nov protokol, ki predstavlja zameno sedanjega protokola IPv4 in se še vedno razvija. Ima mnogo novosti in izboljšav glede na IPv4, med katerimi so gotovo najbolj opazne večji naslovni prostor, podpora za promet v realnem času (QoS – Quality of Service), avtomatska konfiguracija naslovov, dvotočkovna varnost (privatnost in overovitev) ter mobilnost.

Zaradi večjega naslovnega prostora je dodeljevanje naslovov izvedeno hierarhično, kar pripomore k boljšemu usmerjanju in zmanjšanju usmerjevalnih tabel. Slabost tega pa je, da imajo uporabniki (organizacije) na razpolago samo toliko naslovov, kolikor jih je dodeljeno na njihovem nivoju hierarhije. Poleg



Slika 3: Navidezno privatno omrežje



usmerjenih in skupinskih naslovov so vpeljani »anycast« naslovi, s pomočjo teh je naslovljen en usmerjevalnik (npr. eden iz skupine usmerjevalnikov, ki se nahajajo v organizaciji).

Podpora za promet, ki zahteva posebno obravnavo na svoji poti do cilja (QoS), je izvedena s pomočjo dveh polj v glavi IPv6 paketa. Na podlagi teh dveh polj lahko usmerjevalniki dodelijo zahtevano QoS določeni obliki prometa (aplikaciji), vendar uporaba teh polj še ni standardizirana.

Izredno pomembna je avtomatska konfiguracija naslovov, saj si lahko le težko zamislimo administratorje omrežja, ki bi morali ročno konfigurirati 128-bitne naslove na posameznih računalnikih. IPv6 definira avtomatsko konfiguracijo s pomočjo usmerjevalnika(ov) in s pomočjo strežnika DHCP (Dynamic Host Configuration Protocol). Prav tako je pomembna ta funkcionalnost v primeru ponovnega oštevilčenja vseh računalnikov v organizaciji, npr. zaradi zamenjave ponudnika internetnih storitev.

Novi protokol na IP nivoju lahko zagotavlja varnost za vse aplikacije, ne samo za tiste, ki imajo vgrajene varnostne mehanizme. V ta namen IPv6 uporablja dve dodatni glavi, ki zagotavljata integriteto in zaščito pred ponavljanjem. Ena glava zagotavlja overovitev, druga pa privatnost podatkov.

Pomembna je tudi podpora za mobilnost. Mobilni računalniki že sedaj predstavljajo omembe vreden del uporabnikov interneta in podpora za te uporabnike je zahtevana s strani IPv6, opcijaska pa za IPv4. S pomočjo te podpore se lahko mobilno vozlišče priključi na katerokoli omrežje v internetu, pri tem pa ima nepretrgano komunikacijo z ostalimi vozlišči v internetu.

Primerjava med sedanjim protokolom in novo verzijo protokola je prikazana v naslednji tabeli [11]:

#### Primerjava naslavljanja v IPv4 in naslavljanja v IPv6

IPv4 uporablja fiksno 32-bitno polje za naslavljanje (npr. 123.123.123.123) in dopušča uporabo števil le od 0 do 255. IPv6 naslovi so dolžine 128 bitov, kar pome-

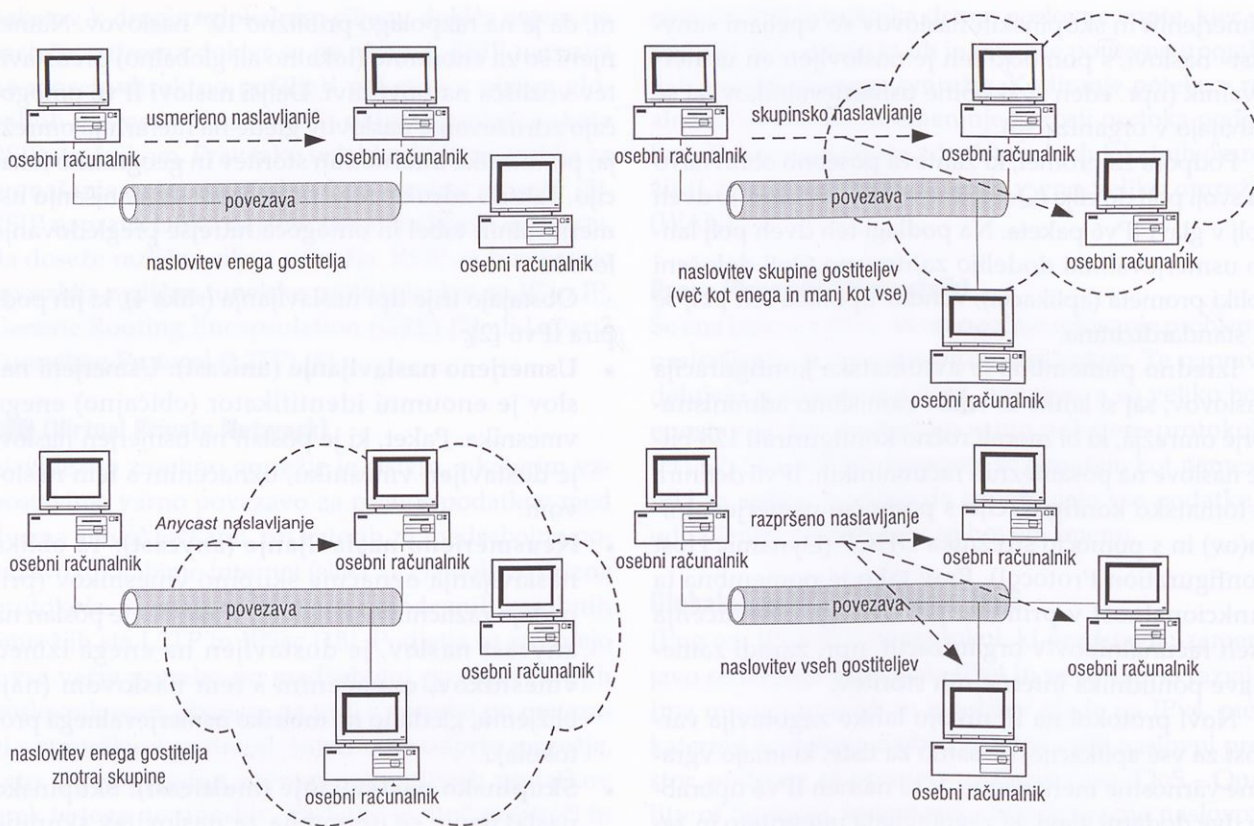
ni, da je na razpolago približno  $10^{38}$  naslovov. Name-njeni so za enoumno (lokalno ali globalno) predstavitev vozlišča na povezavi. Daljši naslovi IPv6 omogočajo združevanje naslovov glede na hierarhijo omrežja, ponudnika internetnih storitev in geografsko lokacijo. Takšno združevanje pripomore k zmanjšanju usmerjevalnih tabel in omogoča hitrejšo pregledovanje le-teh.

Obstajajo trije tipi naslavljanja (slika 4), ki jih podpira IPv6 [2]:

- **Usmerjeno naslavljanje (unicast):** Usmerjeni naslov je enoumni identifikator (običajno) enega vmesnika. Paket, ki je poslan na usmerjen naslov, je dostavljen vmesniku, označenim s tem naslovom.
- **Neusmerjeno naslavljanje (anycast):** Ta oblika naslavljanja označuje skupino vmesnikov (pripadajo različnim vozliščem). Paket, ki je poslan na anycast naslov, je dostavljen na enega izmed vmesnikov, označenim s tem naslovom (najbližjemu, gledano na metriko usmerjevalnega protokola).
- **Skupinsko naslavljanje (multicast):** Skupinsko naslavljanje se uporablja za naslovitev skupine vmesnikov (pripadajo različnim vozliščem). Paket, ki je poslan na skupinski naslov, je poslan vsem vmesnikom, označenim s tem naslovom.

Če IPv4 deluje na Ethernet LAN-u, pogosto potrebuje povezavo med IPv4 naslovi in Media Access Control (MAC) naslovi ter obratno. IPv4 izvaja to funkcijo prek dodatnega protokola, imenovanega Address Resolution Protocol (ARP), ki uporablja razpršeno oddajanje na MAC nivoju. Razpršen paket je sprejet na vseh postajah in povzroči prekinitve na vseh postajah, vključno s tistimi, ki ne uporabljajo IP protokola. Ta neučinkovitost je popravljena v IPv6, kjer ni razpršenih naslovov, njihovo funkcijo so prevzeli skupinski naslovi. V ta namen se na LAN-u uporabljajo metode Neighbour Discovery protokola, ki namesto razpršenega uporablja skupinsko oddajanje.

Protokol	IPv4	IPv6
Prostor za naslavljanje	32 bit	128 bit
Varnost	IPsec (opcijsko)	IPsec integriran
Kakovost storitev (QoS)	Različne in integrirane storitve	Različne in integrirane storitve
Autokonfiguracija	DHCP–dinamično dodeljevanje IP naslovov	Brez strežniška konfiguracija, rekonfiguracija, DHCP
Mobilnost	Mobilni IP naslovi	Mobilni IP naslovi z direktnim usmerjanjem



Slika 4: Oblike naslavljanja

IPv6 naslovi so predstavljeni z 32 heksadecimalnimi besedami, ki so porazdeljene v osem skupin. Te skupine so med seboj ločene z znakom ":".

Obstajajo 3 oblike predstavitve IPv6 naslovov [2]:

- Priporočena oblika je:  
 $x:x:x:x:x:x:x$ ,  
 kjer je 'x' heksadecimalna vrednost osmih 16-bitnih delov naslova.

**Primer:**

ABCD:1298:7CD4:3210:12DC:BA32:A651:C2A0  
 1A80:0:0:0:C:A800:230C:A17C

- V primeru, da je v IPv6 naslovu veliko število ničel, je vpeljana posebna sintaksa, ki lahko kompresira te ničle. Uporaba znaka "::" označuje več skupin po 16 bitov s samimi ničlami. Lahko pa se uporablja samo enkrat v naslovu.

**Primer:**

1A80::C:A800:230C:A17C

- Prav tako je na razpolago oblika, ki je bolj primer-  
 na v mešanih okoljih IPv4 in IPv6. Predstavljena je

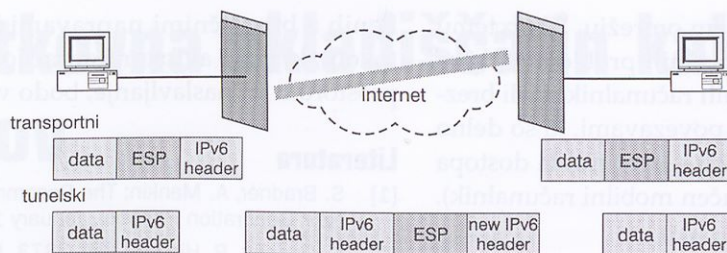
v obliki  $x:x:x:x:x:d.d.d.d$ , kjer 'x' predstavlja heksadecimalno vrednost šestih najvišjih 16-bitnih delov naslova, ž'd' so decimalne vrednosti štirih spodnjih 8-bitnih delov naslova (standardna predstavitev naslovov IPv4).

**Primer:**

1A80::C:A800:193.2.49.201  
 ::193.2.49.252

**Varnost**

V sedanji verziji IP (IPv4) je uporaba šifrirnih in overovitvenih metod izvedena v glavnem prek specifičnih aplikacij, npr. elektronske pošte (Privacy Enhanced Mail, PGP), omrežnega upravljanja (SNMPv3 varnost), dostopa preko WWW (Secure HTTP, Secure Socket Layer) in drugih. Zaradi tega je velik del paketa, ki je poslan prek omrežne povezave, viden v "prosti" obliki in lahko uporabljen za analizo. Z implementacijo varnosti na IP nivoju se lahko zagotovi varno povezovanje ne le za tiste aplikacije, ki imajo vgrajene varnostne



Slika 5: Primer varnosti v IPv6

mehanizme, ampak tudi za mnogo tistih, ki nimajo vgrajenih teh možnosti. IPv6 lahko na tem področju zagotovi overovitev, privatnost in zaščito pred ponavljanjem paketov s pomočjo dveh dodatnih glav. V splošnem Authentication header [19] priskrbi overovitev, Encapsulation Security Payload header [9] pa privatnost oz. šifriranje podatkov (slika 5).

**Mobilnost**

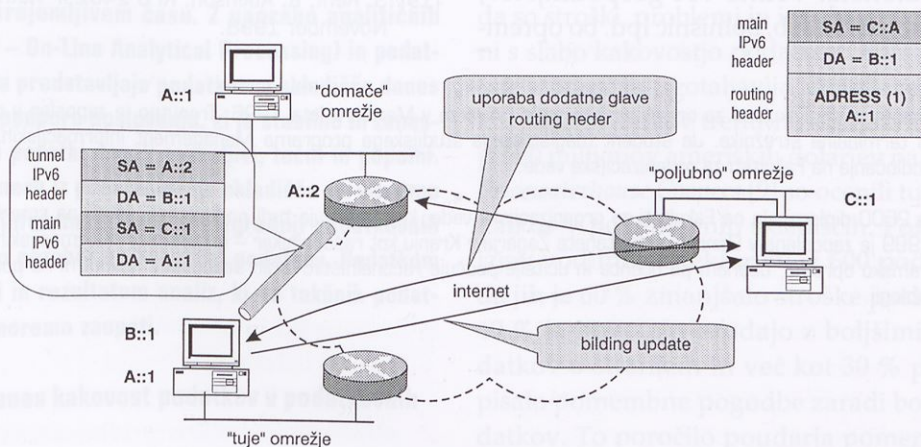
Pod pojmom mobilnost se razume premik vozlišča iz "domačega" omrežja na "tuje". Ne sme se zamenjevati z oddaljenim dostopom, ki je namenjen povezovanju uslužbencev, poslovnih partnerjev, serviserjev itd. iz dislociranih lokacij prek klicnih linij (analognih ali ISDN). Podpora mobilnosti v IPv6 (Mobile IPv6) je še posebej pomembna, saj bodo (so) postali mobilni računalniki pomemben del uporabnikov interneta.

Za IP naslove je značilno, da enoumno predstavljajo točko priključitve vozlišča v internet. Posledica tega je, da se morajo vozlišča nahajati na omrežju, ki je označen z njenim IP naslovom, če želi sprejemati IP pakete. Takšna situacija pa ni najbolj ustrezna za mo-

bilno vozlišče, ki neprestano menja točko povezljivosti v internet in mora zaradi tega neprestano spreminjati svoj IP naslov. Rezultat tega je, da za podporo mobilnosti IPv6 vpelje poleg domačega IP naslova še tuji naslov, ki je bil pridobljen na tujem omrežju s pomočjo tehnik avtomatske konfiguracije naslovov [10].

Problem komunikacije z mobilnim vozliščem v IPv6 se nanaša na problem upravljanja odvisnosti med domačim in tujim naslovom. Mobilno vozlišče, ki je povezano na tuje omrežje, mora določiti na svojem domačem omrežju usmerjevalnik, ki ga bo predstavljal v njegovem imenu, ob njegovi odsotnosti. Zato mu mora mobilno vozlišče sporočiti njegov trenutni naslov, da bo lahko preusmeril paket, ki mu je namenjen, na njegov tuji (trenutni) naslov. Posledica tega je, da bo lahko vsako mobilno vozlišče naslovljeno z njegovim domačim naslovom ne glede na njegovo trenutno točko povezljivosti v internet. Primer komunikacije je prikazan na sliki 6.

Podpora mobilnosti v IPv6 rešuje problem transparentnega usmerjanja paketov do in od mobilnega



Slika 6: Primer komunikacije z mobilnim vozliščem

vozlišča, ko se nahaja na tujem omrežju. Kljub temu pa ne poskuša rešiti vseh splošnih problemov, ki se nanašajo na uporabo mobilnih računalnikov ali brezžičnih omrežij (ukvarjanje s povezavami, ki so delno dosegljive – brezžična omrežja – in kontrola dostopa na povezavah, kjer je priključen mobilni računalnik).

### Sklep

Kljub vsem prednostim, ki jih prinaša IPv6, se pojavlja vprašanje, kdaj bodo (bomo) uporabniki interneta pripravljene začeti uporabljati nov protokol, saj zamenjava IPv4 z IPv6 ne pomeni zamenjave samo protokolnega sklada, ampak tudi drugih stvari. Seveda pa ne smemo pozabiti, da mora biti omogočeno istočasno delovanje IPv4 in IPv6 vsaj še nekaj let, če ne desetletij. Spreminjanje omrežne programske opreme je podobno spreminjanju verzije operacijskega sistema; ta korak lahko prinese nekatere nekompatibilnosti in povzroči potrebo po posodobitvi tako strojne, kakor tudi programske opreme. Po nekaterih predvidevanjih se bo to zgodilo v času med leti 2005 in 2015.

Čeprav novi protokol IPv6 ponuja veliko več, kot le zadosten prostor za naslavljanje, je to najpomembnejši dejavnik, ki spodbuja njegovo čim hitrejšo implementacijo. Največji porast novih naslovov je pričakovati na področju brezžične telefonije. Kljub temu, da v nekaterih delih sveta tehnologija te vrste še ni v takšnem razcvetu, se povpraševanje po njej veča iz dneva v dan, kar seveda zahteva večji prostor za naslavljanje in ostale prednosti, ki jih nudi internetni protokol IPv6.

Prihajajo vedno nove in nove tehnologije, ki bodo v bližnji prihodnosti potrebovale unikatni naslov IP. Edina praktična rešitev bi bila uporaba novega internetnega protokola. Vedno več gospodinjev, pisarn, tovarn, hotelskih sob, bolnišnic ipd. bo oprem-

ljenih z brezžičnimi napravami in potrebe po protokolu, ki nudi avtokonfiguracijo z ustreznim velikim prostorom za naslavljanje, bodo vedno bolj očitne.

### Literatura

- [1] S. Bradner, A. Mankin: The Recommendation for the IP Next Generation Protocol, January 1995.
- [2] S. Deering, R. Hinden: RFC 2373, IP Version 6 Addressing Architecture, July 1998.
- [3] D. C. Plummer: RFC 826, An Ethernet Address Resolution Protocol, November 1982.
- [4] K. Egevang, P. Francis, "The IP Network Address Translator (NAT)", RFC 1631, May 1994.
- [5] M. Borella, J. Lo, D. Grabelsky, G. Montenegro, "Realm-Specific IP: Framework", RFC 3102, October 2001.
- [6] J. Bound, C. Perkins: draft, Dynamic Host Configuration Protocol for IPv6 (DHCPv6), June 1998.
- [7] S. Hanks, T. Li, D. Farinacci, "Generic Routing Encapsulation (GRE)", RFC 1701, October 1994.
- [8] W. Townsley, A. Valencia, A. Rubens, G. Pall, G. Zorn, B. Palter, "Layer Two Tunneling Protocol -L2TP", RFC 2661, August 1999.
- [9] S. Kent, R. Atkinson: RFC 2406 IP Encapsulating Security Payload (ESP), November 1998.
- [10] D. B. Johnson, C. Perkins: draft, Mobility Support in IPv6, November 1998.
- [11] S. Deering, R. Hinden: RFC 2460, Internet Protocol, Version 6 (IPv) Specification, December 1998.
- [12] T. Aljaž: Internet protokol verzija 6, Zbornik Vitel, maj 1999.
- [13] <http://www.ipv6.org>
- [14] <http://6bone.net>
- [15] <http://6ren.net>
- [16] T. Narten, E. Nordmark, W. Simpson: RFC 2461, Neighbour Discovery for IPv6 (IPv6), December 1998.
- [17] S. Thomson, T. Narten: RFC 2462, IPv6 Stateless Address Autoconfiguration, December 1998.
- [18] S. Kent, R. Atkinson: RFC 2401, Security Architecture for the Internet Protocol, November 1998.
- [19] S. Kent, R. Atkinson: RFC 2402 IP Authentication Header, November 1998.

Stjepan Pervan je diplomiral na Fakulteti za organizacijske vede Univerze v Mariboru leta 1996. Trenutno je zaposlen v podjetju IskraTEL kot skrbnik za Windows terminalne strežnike. Je študent magistrskega študijskega programa Management informacijskih sistemov, na smeri Sistemi za podporo odločanja na Fakulteti za organizacijske vede.

Martina Kern je leta 2000 diplomirala na Fakulteti za organizacijske vede, kjer nadaljuje tudi podiplomski študij na smeri Sistemi za podporo odločanju. Od leta 1999 je zaposlena v Osnovni šoli Staneta Žagarja v Kranju kot računalnikar – organizator informacijskih dejavnosti. Skrbi za vso strojno in programsko opremo, obenem pa učence in učitelje poučuje računalništvo in jih seznanja z novostmi na področju informacijsko-komunikacijskih tehnologij.

# Podatkovna skladišča in kakovost podatkov

Krista Rizman Žalik

## Povzetek

Podatkovna skladišča predstavljajo pomembno informacijsko podporo poslovanju in poslovnim odločitvam, saj hranijo podatke o poslovanju. Poslovne odločitve so natančne le, če so podatki točni, investicija v podatkovno skladišče pa bo poplačana samo, če so podatki zanesljivi. Kakovost informacij iz podatkovnega skladišča predstavlja resno tveganje, ki ga je treba obvladati ob načrtovanju in razvoju podatkovnega skladišča. V prispevku so predstavljene in analizirane metode za merjenje kakovosti podatkov podatkovnega skladišča in metode za zagotavljanje in povečanje kakovosti podatkov. Podana je kratka primerjava med njimi, iz nje pa izhajajo smernice za izbiro metode za zagotavljanje in izboljšanje kakovosti podatkov podatkovnega skladišča.

## Abstract

### Data Warehouses and Data Quality

Data warehouses collect data and provide an important information support to produce information for business decision making. It is successful only if complete and accurate data are applied. The investment will be returned only if data are reliable. The information quality provided by data warehouses is a serious risk, which should be taken into account during the process of data warehouse design and development. In this paper, methods for data quality measurement and methods for providing and increasing the data quality are analyzed. A short comparison of analyzed methods is given. Based on the comparison, directions for choosing a method to provide and improve data quality are given.

## 1 Uvod

**Podjetja danes hranijo in sproti obdelujejo transakcije – poslovne dogodke v poslovnem procesu – v sistemih za sprotno obdelavo, ki jih označujemo tudi s kratico OLTP (angl. On-Line Transactional Processing). Za povečanje uspešnosti poslovanja pa gradijo podatkovna skladišča (angl. Data Warehouses), ki dajejo celovit pogled na podatke posameznega podjetja. Omogočajo izdelavo potrebnih analiz, opazovanj trendov in predvidevanj posameznih kazalcev poslovanja. Na voljo so analitikom in upravljalcem, ki potrebujejo informacije za odločanje v sprejemljivem času. Z uporabo analitičnih orodij (angl. OLAP – On-Line Analytical Processing) in podatkovnega rudarjenja predstavljajo podatkovna skladišča danes vrh informacijske podpore poslovanju, ki je stabilno in zanesljivo le, če so tudi podatki dovolj zanesljivi, točni in popolni. V nasprotnem primeru iz podatkovnega skladišča ne moremo dobiti zanesljivih informacij. Zato pri izgradnji podatkovnih skladišč ne smemo pozabiti na kakovost podatkov. Podatkom neznane kakovosti in rezultatom analiz, ki na takšnih podatkih temeljijo, ne moremo zaupati.**

### 1.1 Kakšna je danes kakovost podatkov v podatkovnih skladiščih?

Kakovost podatkov v obstoječih podatkovnih skladiščih je pogosto slaba zaradi napak pri vnosu, raz-

ličnih sprememb podatkov in struktur, ki so nastajale skozi čas, napačnih podatkov iz spletnih aplikacij ali iz zunanjih virov, ali pa zaradi združevanja dobrih podatkov z zastarelimi in nezmožnostjo ločevanja med njimi.

V poročilu leta 2002 [20], ki ga je izvedel *Data Warehousing Institute*, je skoraj polovica (44 %) vprašanih poročala, da je kakovost podatkov "slabša, kot si kdorkoli lahko predstavlja". 40 % anketiranih je priznalo, da so stroški, problemi in izgube neposredno povezani s slabo kakovostjo podatkov. *Data Warehouse Institute* v poročilu ugotavlja, da težave povezane s kakovostjo podatkov trenutno stanje ameriška podjetja 600 milijonov ameriških dolarjev na leto. V poročilu *Pricewaterhouse Coopers* [9] so ocenili tudi kakovost podatkov v podatkovnih skladiščih. Poročilo z analizo izjav vodilnih direktorjev iz 600 podjetij ugotavlja, da jih je 60 % zmanjšalo stroške poslovanja, več kot 40 % je povečalo prodajo z boljšimi analizami podatkov o strankah in več kot 30 % podjetij je podpisalo pomembne pogodbe zaradi boljše analize podatkov. To poročilo poudarja pomen kakovosti podatkov in povzema, da imajo podjetja, ki upravljajo svoje podatke kot strateški vir in investirajo v

njihovo kakovost, večji ugled in so bolj dobičkonosna kot tista, ki tega ne počnejo.

## 1.2 Prenova podatkovnih skladišč

Podatkovna skladišča danes hranijo podatke, ki največkrat odgovarjajo na splošna vprašanja: kateri izdelki prinesejo največji/najmanjši dobiček, kakšne izdelke in storitve bodo uporabniki želeli v prihodnje, kateri so sedaj in so bili v bližnji preteklosti najbolj prodajani izdelki. Značilni podatki, ki se zbirajo v podatkovnih skladiščih, so: enolična oznaka izdelka, izdelek, trgovina, cena, popust, količina, vrednost, datum. S temi ključnimi podatki lahko analiziramo prodajo po izdelkih, času, skladiščih in področjih. Ti običajni podatki podatkovnih skladišč pa ne povedo nič o kupcu, kaj vse je kupil in koliko, katere izdelke je kupil hkrati in v kakšnem časovnem zaporedju. Podjetja danes dodajajo informacije o svojih strankah v podatkovna skladišča, kar jim prinese konkurenčno prednost. 70–80 % podatkovnih skladišč svetu je namenjenih obravnavi strank. Podjetja težijo k pridobivanju novih strank - kupcev. Ker pa število le-teh ni neomejeno, so cilji podjetij tudi obdržati svoje stranke čim dlje, prodati jim čim več stvari in poslovati z njimi na učinkovitejši način. Da lahko te težnje udeležijo in povečujejo prodajo ter dobiček, si morajo ustvariti jasno sliko obnašanja svojih strank, pravila nakupov in analizirati njihove navade. Podatki o strankah, ki jih dodajajo v podatkovna skladišča, so manj formalizirani, strukturirani in zato tudi manj kakovostni v primerjavi s podatki, ki so že v podatkovnih skladiščih. Z dopolnjevanjem podatkovnih skladišč postane problem kakovosti podatkov za podjetja še večji.

## 2. Kaj je podatkovno skladišče in kaj kakovost podatkov?

Immon definira podatkovno skladišče kot subjektivno orientirano, integrirano, stanovitno, časovno raznoliko zbirko podatkov, ki podpirajo poslovno odločanje [11]. Skladišče ni funkcionalno orientirano ampak so v njem podatki subjektivni. V podatkovnih skladiščih CRM (angl. Customer Relationship Management) je subjekt stranka. Podatki so integrirani in poenoteni. Podatki so lahko elementarni ali že integrirani. Podatkovno skladišče hrani podatke stare tudi več let, kar omogoča analize trendov in oblikovanje napovedi gibanja posameznih kazalcev v prihodnje. Kimball [13,14] vidi podatkovno skladišče kot področne

shrambe, kjer hranimo podrobne transakcijske podatke. V virih, kjer so primerjali obe definiciji, ne vidijo bistvene razlike med definicijami in poudarjajo, da se podatkovno skladiščenje razvija in dopolnjuje [1,7].

Obstaja več videnj kakovosti podatkov:

- *Current Analysis* [2] pravi, da je kakovost podatkov preprosto zrcalo točnosti podatkov organizacije. Dobra kakovost podatkov pomeni, da so podatki organizacije točni, popolni, konsistentni, pravočasni, enoviti in pravilni. Bolj kakovostni so podatki, bolj jasno predstavljajo natančen, usklajen vidik podjetja skozi podsisteme, organizacijske enote in vrste poslovanja.
- Kakovost podatkov je stanje popolnosti, veljavnosti, konsistentnosti, pravočasnosti in točnosti, kar dela podatke primerne in uporabne za specifične analize [10].
- Kakovost podatkov je često definirana kot proces urejanja informacij, tako da so posamezni zapisi natančni in točni, ažurni in konsistentno predstavljeni. Kakovost podatkov je težko izmeriti in je dovolj kritičen dejavnik za neuspeh projekta ali neizpolnitev strateškega cilja [3,5,6]. Podatkovna plast organizacije je kritični element, ker je zelo enostavno zanemariti kakovost podatkov ali narediti preoptimistične ocene. Imeti kakovostne podatke za analizo je velika konkurenčna prednost.

## 3. Kako merimo kakovost podatkov?

Že iz širine zgornjih definicij o kakovosti podatkov je jasno, da bomo poskušali le ovrednotiti kakovost podatkov, ne bomo pa je mogli natančno izmeriti. Meritev kakovosti podatkov ni delo za perfekcionista, ampak za ljudi, ki poznajo pomen meritve. Ogleдали si bomo dve metodi za merjenje kakovosti podatkov. Prva, opisana v viru [15], zelo poudarja pomen množice osnovnih pravil za ocenitev kakovosti vhodnih podatkov, ki so pomembni za uspešnost gradnje podatkovnega skladišča glede na vsebino in pravočasnost.

Nekaj primerov pravil:

- samo izdelki dodani v zadnjem mesecu nimajo natančnih informacij o prodaji,
- vsi izdelki imajo enolične identifikatorje v tabeli prodaj izdelkov,
- ni presledkov pred imeni kupcev,
- ni posebnih znakov pri imenih kupcev,
- vsi identifikatorji izdelkov so enolični,
- vse prodaje so med 0 in 100.000.000,00 SIT itn.

Ko so pravila definirana in so določeni podatki, moremo oceniti kakovost podatkov. Ocenitev bo predstavljala stanje kakovosti podatkov za to pravilo. Celotna ocena bo povprečje ocen izraženo v točkah. Samo povprečje ni tako učinkovito kot sistem, ki da večjo težo pomembnejšim pravilom. Vrednost celotne ocene (število točk) definira kakovost podatkov. Avtor iz empiričnih izkušenj predlaga ocenitev, ki da naslednje rezultate prikazane v tabeli 1 [15].

Ocena kakovosti	
99–100	izjemna kakovost podatkov
96–98	dobra kakovost podatkov
90–95	povprečna kakovost podatkov
81–89	podpovprečna kakovost podatkov Takšna kakovost negativno vpliva na poslovanje.
65–80	slaba kakovost podatkov Sistem bo največkrat dal zelo slabe rezultate.
0–64	zelo slaba kakovost podatkov Podatki so neuporabni in jih treba izboljšati.

Tabela 1: Ocenitve kakovosti podatkov

Druga metoda, opisana v viru [10], opravi meritev kakovosti podatkov z opazovanjem vseh lastnosti kakovosti podatkov. V tabeli 2 so primeri meril, katerih kršenje nam v odstotkih predstavlja neakvost podatkov. Tabela prikazuje množico skupnih lastnosti za merjenje kakovosti podatkov in navaja primere meril za določanje zahtev za kakovost.

Latnosti kakovosti podatkov	Opis	Primer merila
Natančnost	Natančnost je razmerje ujemanja med množico podatkov in pravih vrednosti.	Merilo natančnosti je odstotek pravih vrednosti.
Popolnost	Popolnost je razmerje, ki opisuje število atributov, ki imajo določene vrednosti.	Merilo popolnosti je odstotek atributov, ki imajo vrednosti.
Doslednost	Doslednost je ujemanje ali logična skladnost med podatki.	Merilo doslednosti je odstotek ujemanja vrednosti, ki zadovoljujejo pogoje ujemanja ali skladne vrednosti.
Primerljivost	Primerljivost je ujemanje ali logična usklajenost, ki dovoljuje korelacije in primerjave z ostalimi podobnimi podatki.	Merilo primerljivosti je odstotek referenčne integritete.
Pravočasnost	Pravočasnost je dostopnost do podatka ali več podatkov, ki je zagotovljena v zahtevanem ali določenem času.	Merilo pravočasnosti je odstotek podatkov zagotovljenih v predpisanem času (na primer: v eni uri ali enem dnevu).
Enoličnost	Enoličnost je pojav, ki opisuje, da se vrednost atributa pojavi le enkrat.	Merilo enoličnosti je odstotek zapisov, ki ne ustrezajo principu enoličnosti.
Veljavnost	Veljavnost je lastnost, ki opisuje ustreznost in spremenljivost podatkov in zmanjšuje verjetnost napak.	Merilo veljavnosti je odstotek vrednosti, ki sodijo v pričakovane zaloge dovoljenih vrednosti.

Tabela 2: Množica lastnosti kakovosti podatkov

### 3.1 Primerjava metod merjenja kakovosti

Obe obravnavani metodi merjenja kakovosti podatkov temeljita na postavljanju pravil in nato na ocenitvi odstotka primerov podatkov, ki kršijo postavljena pravila in predstavljajo odstotek neakvostnih podatkov.

Druga metoda oceni kakovost podatkov bolj natančno, z vseh vidikov definicij kvalitete podatkov podatkovnega skladišča. Najde neakvostne podatke, ki kršijo pravila dobre kakovosti podatkov, in zahteva njihovo zmanjšanje. Ni pa merila, ki bi definiralo, kaj pomenijo odstotki neakvosti podatkov. Ocena ukrepanja je v rokah razvijalca podatkovnega skladišča.

Prva metoda daje empirično merilo za opisno oceno kakovosti podatkov iz povprečja odstotkov neujemanja pravil. Metoda poudarja določitev množice osnovnih pravil za ocenitve kakovosti vhodnih podatkov, ki so pomembni za uspešnost in dobičkonosnost gradnje podatkovnega skladišča glede na vsebinske in časovne omejitve.

Meritev kakovosti podatkov lahko izvedemo z orodji (na primer Data Quality Inspector [17]), ki omogočajo definiranje pravil in izbiro podatkovnih množic. Nato analizirajo podatke in poiščejo kršitve pravil ter posredujejo neakvostne zapise odgovornim za njihovo kakovost.

## 4 Možnosti povečanja kakovosti podatkov

Obstajajo tri možnosti povečanja kakovosti podatkov:

1. Odložiti aktivnost povečanja kakovosti podatkov na kasneje.

2. Vgraditi pravila, ki bodo preverjala podatke v aplikacijah, ki zbirajo podatke.

3. Zgraditi podatkovno skladišče in čistiti podatke. Odločitev za eno izmed možnosti za povečanje kakovosti podatkov je odvisna od:

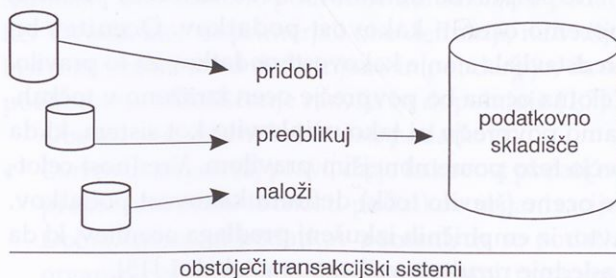
- stopnje kakovosti podatkov,
- stanja obstoječih transakcijskih sistemov (ali so že potrebni prenove ali pa so to novejši sistemi) in
- potreb in želja po poslovanju s kakovostnimi podatki.

Prva možnost je nasprotje reka »Kar lahko storiš danes, ne odlašaj na jutri« in se v praksi največkrat slabo obnese.

Druga možnost je nujno potrebna, če so podatki zelo slabi, tako rekoč neuporabni. Vgradnja pravil v aplikacijah, ki zbirajo podatke, pa ne poveča kakovosti starih podatkov. Le-ti pa so v podatkovnih skladiščih še kako potrebni za izvajanje analiz in primerjanj v daljšem časovnem obdobju in za napovedovanje gibanj posameznih količin v prihodnje. To možnost izvedejo v podjetjih, kjer so transakcijski sistemi zastareli in so potrebni prenove.

Tretja možnost je izboljšanje kakovosti podatkov v podatkovnem skladišču. Nalagati in hraniti nekakovostne podatke in jih nato kasneje prečistiti, je veliko večji strošek, kot pa prečistiti podatke in jih čiste hraniti v podatkovnem skladišču.

Mesto za zagotavljanje kakovosti podatkov v okolju podatkovnega skladišča je sam razvoj le-tega. Podatkovno skladišče s še tako izpopolnjenim načrtovanjem, a s podatki, ki niso kakovostni, je malo vredno. Zgrajeno podatkovno skladišče za upravljanje podatkov podatke tudi očisti, kar je pomembna naloga, ki je bila poudarjena že ob definiciji podatkovnega skladišča. Immon, oče podatkovnih skladišč, pravi, da je naloga koraka ETL (angl. Extract Transform Load, slov. pridobivanje, preoblikovanje in nalaganje podatkov) naložiti v podatkovno skladišče integrirane in prečiščene podatke (slika 1). Usklajenost procesa ETL in kakovost podatkov daje možnost za enostavnejšo upravljanje kompleksnih podatkovnih integracij. Z uvedbo kakovosti podatkov v proces ETL sta zagotovljena kakovost podatkov in točnost. Najtežavnejše v procesu ETL je pridobivanje podatkov iz različnih virov, preoblikovanje podatkov v nove formate in nalaganje podatkov v podatkovna skladišča. Cilj procesa ETL mora biti tudi zajemanje čistih in točnih podatkov.



Slika 1: Proces ETL

Za izboljšanje kakovosti podatkov v procesu ETL uporabljamo naslednje tehnike: čiščenje podatkov, dopolnjevanje in ujemanje ter usklajevanje.

Čiščenje podatkov je odkrivanje in popravljanje nekakovostnih podatkovnih elementov in podatkovnih struktur.

S tehniko dopolnjevanja povečujemo obseg informacij, ki jih lahko dobimo iz podatkov. Na primer, naslov lahko dopolnimo z vljudnostnimi nazivi, ki so odvisni od spola ali pa z geolokacijskimi podatki o pošti in mestu.

Z ujemanjem ugotovimo množico ponavljajočih se ujemanj zapisov. S postopkom združevanja takšne zapise uskladimo in združimo v en zapis. Poznamo tri metode ujemanja in združevanja: odpravo podvojevanj, združevanje v družine in povezovanje zapisov. Vsaka izmed njih je uporabna v določenem primeru. Odprava podvojevanj z ugotavljanjem ujemanja in združevanjem, odpravi podvojene zapise. Združevanje v družine združi zapise, ki imajo vsaj en enak atribut. Ta atribut je ključ za združevanje. Tipičen primer je združevanje podatkov o strankah v skupna gospodinjstva po skupnem naslovu. Povezovanje zapisov je splošnejši primer združevanja v družine. Zapise povezujemo z drugim namenom in ne po skupnem naslovu, ampak na primer po organizacijskem principu.

Poznamo mehko, popolno in verjetnostno ujemanje. Mehko ujemanje temelji na bolj ali manj natančnih pravilih za ujemanje in na območjih podatkov, ki niso natančno definirana. Popolno ujemanje daje enako težo različnim atributom zapisa. Verjetnostno ujemanje izkorišča statistično verjetnost, da ujemanje določenega elementa zapisa z določeno verjetnostjo zagotavlja, da sta zapisa ista.



Čiščenje, dopolnjevanje, ujemanje in združevanje izvajamo zaporedno. Izvedemo samo tiste tehnike, ki so primerne in ustrezne našim problemom in podatkom.

Graditelji podatkovnih skladišč, kot je na primer Oracle Warehouse Builder [19], so orodja, ki nudijo podporo razvoju in avtomatizirajo obravnavane tehnike povečanja kakovosti podatkovnih skladišč. Poleg razvojnih orodij obstajajo tudi razvojne metode, ki poudarjajo povečanje kakovosti [4,10,12,15,18]. Obravnavali in primerjali bomo:

- pristop k povečanju kakovosti podatkov zasnovan na tveganjih,
- pristop k povečanju kakovosti podatkov in ROI,
- metodo za razvoj podatkovnih skladišč DWM.

## 5 Pristop k povečanju kakovosti podatkov zasnovan na tveganjih

Ta vključuje štiri ključne aktivnosti za zagotavljanje kakovosti podatkov v razvojni cikel [10]:

1. Določitev pričakovane kakovosti podatkov in metrike za merjenje.
2. Identificiranje tveganja v kakovosti podatkov in predvidevanje, kdaj in kateri podatki iz podatkovnega skladišča ne bodo uspeli zadovoljiti pričakovanj.

3. Zmanjšanje tveganja – določitev akcije za zmanjšanje vsakega večjega tveganja.
4. Opazovanje in ocenitev rezultatov.

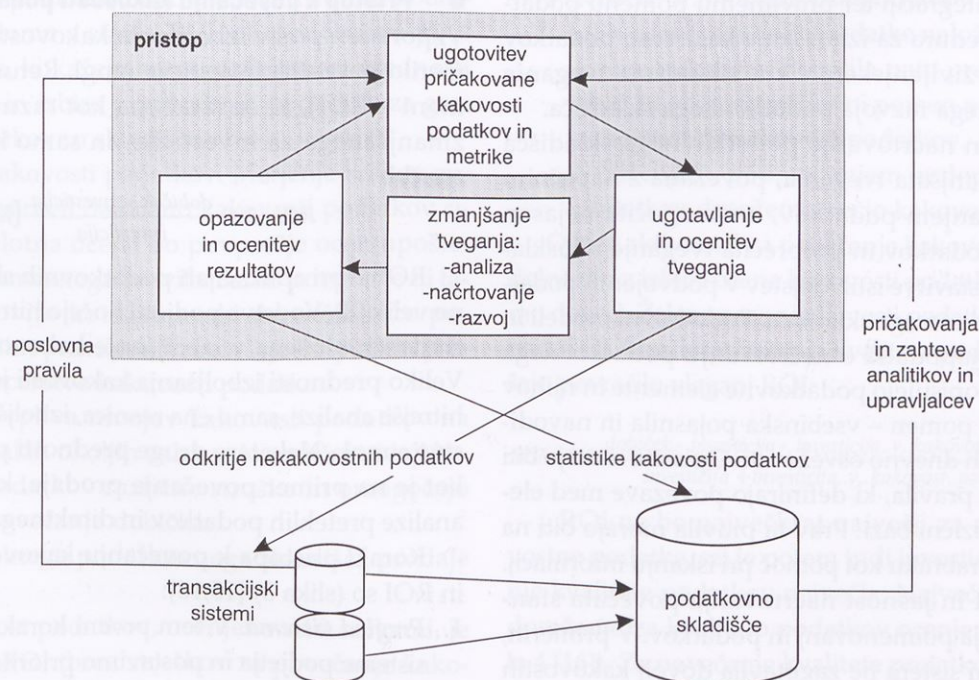
### 5.1 Določitev pričakovane kakovosti podatkov in metrike za merjenje

Prvi korak pri zagotavljanju kakovosti podatkov so definicije kakovosti podatkov, ki bodo pomagale uresničiti poslovne cilje. V primeru poslovnega cilja povečanja trga, so pričakovanja za kakovost podatkov osredotočena predvsem na zbiranje podatkov o prodaji strankam in kakovosti agregiranih podatkov.

Skupne lastnosti kakovosti podatkov po definirani metriki (tabela 2) določimo predvsem za podatke, ki povzročajo večje tveganje pri doseganju zastavljenega poslovnega cilja. Tveganje lahko povežemo s ciljem podatkovnega skladišča. Na primer, za doseganje poslovnega cilja optimizacije dobave so zelo pomembni podatki o razdeljevanju izdelkov.

### 5.2 Identificiranje tveganja v kakovosti podatkov

Je postopek predvidevanja, zakaj ne bi podatki iz podatkovnega skladišča uspeli zadovoljiti pričakovanj. Tveganje je lahko zunanji dogodek ali poznano stanje, ki povzroči nekakovostne podatke. Izpad računalnika je tveganje za pravočasnost podatkov. Na



Slika 2: Zagotavljanje kakovosti podatkov po pristopu, zasnovanem na tveganjih

primer, podatki o velikosti dobav izdelkov so obvezni podatki, toda v primeru izpada sistema kar za nekaj zapisov teh podatkov ne bo. Veliko tveganje je tudi napačna uporaba podatkov. Uporabniki podatkovnega skladišča v prvem koraku preverijo podatke, ki jih dobijo iz podatkovnega skladišča glede na podatke virov in pomagajo oblikovati ustrezna navodila za pravilno uporabo podatkov. Seveda pa v procesu gradnje sami preverimo podatke in ocenimo njihovo kakovost v skladu z merili, postavljenimi v prejšnjem koraku. Za to lahko uporabimo orodja, kot je na primer Data Quality Inspector [17]. Če ta preverjanja pokažejo, da je kakovost podatkov zajeta v podatkovno skladišče tako majhna, da predstavlja tveganje za doseg pričakovanih rezultatov podatkovnega skladišča, potem moramo oceniti napor za zmanjšanje tveganj in izvesti aktivnosti, ki zagotovijo večjo kakovost vhodnih podatkov.

### 5.3 Zmanjšanje tveganja

Za vsako tveganje definiramo kako in v katerem koraku razvoja ga bomo odpravili.

Procedura za izboljšanje kakovosti podatkov v različnih korakih razvoja podatkovnega skladišča je naslednja:

- V koraku definiranja zahtev je poudarek na metapodatkih, ki so pomembni za podporo analizi domen in integraciji ter pravilnemu pomenu podatkov. Procedura za izboljšanje kakovosti podatkov zgodaj v življenjskem ciklu zmanjšuje tveganje neuspešnega razvoja podatkovnega skladišča.
- Analiza in načrtovanje podatkovnega skladišča lahko zmanjšata tveganja, povezana z napačnim razumevanjem podatkov, onemogočita nejasen pomen podatkov in preprečita tveganje neusklajene predstavitve istih dejstev v podvojenih podatkih podatkovnega skladišča. Podatkovni modeli in ostali metapodatki o načrtovanju podatkovnega skladišča opisujejo podatkovne elemente in njihov natančen pomen – vsebinska pojasnila in navodila, kako jih dnevno osveževati. Opisana morajo biti poslovna pravila, ki definirajo povezave med elementi v fizični bazi. Prav ta pravila morajo biti na voljo uporabniku kot pomoč pri iskanju informacij.
- Izraznost in jasnost načrtovanja povečata standardizacijo poimenovanj in podatkov. V primerih, ko izvorni sistem ne zagotavlja dovolj kakovostnih virov, je smiselno začeti projekt za izboljšanje izvorne sistema, tako da bo zmanjšano tveganje

zaradi neakovostih podatkov. Iz zahtev za kakovost podatkov razvijemo množico pravil in metrike kakovosti, načrtamo in razvijemo preverjanja za pravila kakovosti podatkov za ugotavljanje in poročanje o napačnih podatkih. Onemogočimo vstop neakovostnih podatkov v podatkovno skladišče.

- Razvoj lahko zmanjša tveganje z vključevanjem popravljanj in merjenj kakovosti podatkov v proces ETL. S preverjanjem ugotovimo preveri format in tip podatkov, usklajenost vrednosti z domenami, usklajenost z drugimi povezanimi podatki in usklajenost z metapodatki. Ugotavljanje pravilnosti preverja točnost podatkov, torej natančen opis realnih objektov. Potrjevanje pravilnosti pokaže primernost uporabe podatkov za različne uporabe in zmanjša tveganje napačne uporabe podatkov.

### 5.4 Vzdrževanje podatkovnega skladišča – opazovanje, pregledovanje in ocena kakovosti podatkov

Kakovost vhodnih podatkov v podatkovno skladišče je potrebno nadzirati. Večkratna merjenja kakovosti vhodnih podatkov nam kažejo, ali se res približujemo zastavljeni kakovosti podatkov. Vzdrževanje pa predstavlja tudi ocenitev novih uporab podatkov in ugotavljanje vzrokov slabe kakovosti podatkov in njihovo odpravljanje.

### 6 Pristop k povečanju kakovosti podatkov in ROI

Vzporeden pojav izboljšanja kakovosti podatkov je merilo povračila investicije (angl. Return On Investment – ROI), ki se izračuna kot razmerje dobička zmanjšanega za investicijo, in samo investicijo po enačbi:

$$ROI = 100 \frac{\text{dobiček} - \text{investicija}}{\text{investicija}} [\%]$$

ROI je pri aplikacijah podatkovnih skladišč običajno velik [8]. Vodstva podjetij hočejo hitro povračilo za sredstva, vložena v izboljšanje kakovosti podatkov. Veliko prednosti izboljšanja kakovosti je nemerljivih: hitreje analize, samo ena resnica, izboljšano zadovoljstvo strank. Nekatere druge prednosti pa so merljive, kot je na primer povečanje prodaje, ki je posledica analize preteklih podatkov in direktnega oglaševanja.

Koraki pristopa k povečanju kakovosti podatkov in ROI so (slika 3) [15,16]:

1. **Pregled sistema.** V tem prvem koraku pregledamo sisteme podjetja in postavimo prioritete glede uporabe in potrebe po kakovosti podatkov. Ti sistemi temeljijo na točnosti podatkov in jih lahko najdemo

med sistemom OLTP, odločitvenimi sistemi in podatkovnimi skladišči.

2. Definiranje pravil. Pravil za kakovost podatkov ne dobimo z določitvijo, kako naj bodo podatki videti, ampak z oceno škode sistema, ki jo povzročijo nekakovosti podatki. Meritev vseh pravil daje oceno kakovosti podatkov. Pravilom, ki opisujejo zahteve za kakovost podatkov, ki, če so nekakovostni, povzročijo največ škode, damo največji pomen in težo. Nabor elementov kakovosti, ki jih lahko zagotavljajo pravila, je:

- Referenčna integriteta, ki se nanaša na integriteto sklicevanja med podatki v različnih tabelah. Kot primer, identifikatorji izdelkov v tabeli prodaje izdelkov morajo biti natančneje opisani v tabeli izdelkov.
- Enoličnost nekaterih podatkov, kot so identifikatorji. Primer kršitve pravila je, če isti identifikator opisuje dva ali več izdelkov.
- Kardinalnost povezav določa količinska razmerja povezav. Izdelek z enim enoličnim identifikatorjem lahko prodamo večkrat ali pa ga še nismo nikoli prodali, če gre za nov izdelek.
- Smiselne vrednosti domen.
- Oblikovne pravilnosti, kot npr. nobenega presledka pred imeni, imena brez posebnih znakov, uporaba velikih in malih črk in ne samo velikih.

3. Označitev podatkov. Ugotavljanje lastnosti podatkov izvedemo s stavki SQL ali podobnimi povpraševanji, ki kažejo na porazdeljenost vrednosti podatkov in izbiro pomembnih podatkov za gradeno podatkovno skladišče.

4. Meritev kakovosti podatkov. Merjenje kršitev posameznih pravil da oceno kakovosti podatkov sistema. Celotna ocena bo povprečje ocen upoštevanja posameznega pravila. Samo povprečje ni tako učinkovito kot sistem, ki da večjo težo pomembnejšim pravilom. Tabela 2 je empirična preglednica ocenitve kakovosti podatkov.

5. Ocenitev vpliva nivojev kakovosti podatkov na ROI. V enakih okoljih zahtevajo različni nivoji kakovosti podatkov različne stroške in omogočajo različno povečanje dobička.

Primer iz realnega podatkovnega skladišča prodaje kaže, [16]:

- Za kakovost podatkov, ocenjeno z 90 točkami, je projektni ROI (brez stroškov za povečanje kakovosti podatkov) 175 %.

- Za kakovost podatkov, ocenjeno s 85 točkami, je projektni ROI (brez stroškov za povečanje kakovosti podatkov) 101 %.

- Za kakovost podatkov, ocenjeno z 80 točkami, je projektni ROI (brez stroškov za povečanje kakovosti podatkov) 64 %.

Tako vidimo, da je investicija v podatkovno skladišče (brez upoštevanja stroškov povečanja kakovosti podatkov) tem bolje poplačana, čim bolj kakovostne podatke obdelujemo. ROI (brez upoštevanja stroškov povečanja kakovosti podatkov) se nelinearno povečuje in je večji od linearnega prirastka za najbolj kakovostne podatke.

6. Povečanje kakovosti podatkov: Velja, da je ceneje povečati kakovost podatkov v virih kot pa v podatkovnih skladiščih. Če imamo na primer polje, ki ima veliko analitično vrednost in je polovica vrednosti nevnešenih in neuporabnih, potem nobena transformacijska strategija ne more pridobiti zanesljivih informacij iz teh podatkov. V primerih zastarelih sistemov povečanje v virih ni možno zaradi kompleksnosti in stare tehnologije. Kakovost podatkov povečamo z naslednjo množico aktivnosti v procesu ETL:

- Podatke, ki zelo kršijo pravila kakovosti podatkov, naložimo v vmesne tabele, kjer jih lahko ročno pregledujemo in popravljamo.
- Kršitve pravil kakovosti podatkov ugotovimo, o njih poročamo, toda podatke naložimo v podatkovno skladišče. Poročila nato pregledajo analitiki, ki poznajo poslovni pomen podatkov in so zadolženi za kakovost podatkov.
- S čiščenjem, popravljanjem in dopolnjevanjem podatkov dosežemo večjo kakovost podatkov.

Obseg aktivnosti za povečanje kakovosti podatkov določimo z izbiro ocene kakovosti podatkov, ki jo želimo doseči. Zelena ocena kakovosti podatkov je povezana z ROI. Stroški povečanja kvalitete podatkov zmanjšajo povračilo vlaganj ROI:

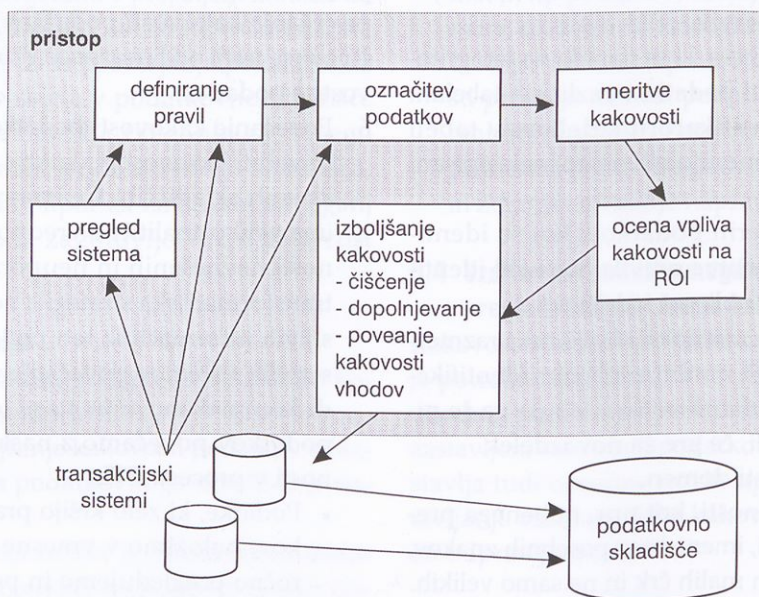
$$ROI = 100 \frac{\text{dobiček} - \text{investicija} - \text{investicija} \cdot \text{v. kakovost} \cdot \text{podatkov}}{\text{investicija} + \text{investicija} \cdot \text{v. kakovost} \cdot \text{podatkov}} - [\%]$$

ROI ne bo največkrat največji za najbolj kakovostne podatke, saj je potem tudi investicija v povečanje kvalitete podatkov največja. Največji ROI (83 %) dosežemo za kvaliteto podatkov ocenjeno z 90 (tabela 4 [16]). Za povečanje kvalitete podatkov do 99 % se ROI zmanjša na 72 %.

Ocena kvalitete podatkov	Direktno trženje (A)	Uspešnost trženja (B)	Povprečen dobiček (\$) (C)	Dobiček (\$) (D=A*B*C)	Investicija* (I)	ROI* (=D-I/I)	Stroški povečanja kvalitete (I1)	ROI (=D-I-I1/I+I1)
99	115.000	0,0425	250	1221875	410000	198 %	410000	72 %
90	110.000	0,04	250	1100000	400000	175 %	400000	83 %
85	100.000	0,03	250	787500	390000	102 %	390000	69 %

\* brez stroškov za povečanje kakovosti podatkov

Tabela 4: Stroški in koristi povečanja kvalitete podatkov



Slika 3: Koraki pristopa k povečanju kakovosti podatkov in ROI

## 7 Razvojna metoda DWM in povečanje kakovosti podatkov

Metodologije za razvoj podatkovnih skladišč natančno opisujejo vse aktivnosti in procese v razvoju podatkovnega skladišča ter tudi bolj ali manj uspešno obravnavajo zagotavljanje kakovosti podatkov. Oglejmo si aktivnosti, ki jih poudarja metoda DWM (angl. Data Warehouse Method) [18] za zagotavljanje in povečanje kakovosti podatkov.

V prvem koraku razvoja podatkovnega skladišča (strategiji) je potrebno postaviti tudi strategijo kakovosti podatkov: pristop za sprotno integriteto podatkov v podatkovnem skladišču vključno z upravljanjem podatkov, upravljanjem z napakami in izjemami, čiščenjem podatkov, opazovanjem in nadzorovanjem podatkov. Metoda predlaga izdelavo lastništva podatkov in procese odprave odstopanj podatkov ter definiranje standardov za podatke.

Kakovost podatkov in integriteta morata biti obravnavani že v strategiji. Vpliv kakovosti podatkov na rezultirajoče podatkovno skladišče je zelo pomemben, saj kakovost podatkov določa zaupanje uporabnikov v podatkovno skladišče. Pri tem skrbimo za kakovost podatkov v vseh korakih razvoja.

- Pri definicijah določimo obseg zagotavljanja kakovosti podatkov ter izdelamo načrt. Definiramo naloge, vire in časovne okvirje za zagotavljanje kakovosti podatkov.
- Pri analizi izdelamo plane in postopke za povečanje kakovosti podatkov. Izdelamo natančne zahteve za kakovost podatkov vključno z obravnavanjem napak in izjem, čiščenjem podatkov in opazovanjem ter nadzorovanjem podatkov. Ustvarimo procedure za sprotno reševanje neakovostnih podatkov (uporabe podatkov, standardov poimenovanja, formati, definicije podatkov, lastništva, popravljanje

napak). Metapodatki opisujejo tudi kakovost podatkov.

- V koraku načrtovanja načrtamo module, ki bodo izvedli zahtevane aktivnosti za zagotavljanje kakovosti podatkov, module za čiščenje, za obravnavo napak in izjem ter module za opazovanje in nadzorovanje podatkov.

Tabela 3 kaže prisotnost aktivnosti zagotavljanja kakovosti podatkov v vseh korakih razvoja podatkovnega skladišča in odstotek, potreben za proces zagotavljanja kakovosti podatkov v posameznem koraku razvoja [11].

FAZE	PROCESI Zbiranje zahtev	Kakovost podatkov	...	Skupaj %
Strategija	...	2,2 %	...	100 %
Definicije	...	12,1 %	...	100 %
Analiza	...	6,5 %	...	100 %
Načrtovanje	...	7,9 %	...	100 %
Izgradnja	...	0,6 %	...	100 %
Uvedba	...	...	...	100 %

Tabela 3: Kakovost podatkov v DWM je ena izmed aktivnosti, ki pokriva večino korakov načrtovanja

## 8 Primerjava metod

Identificirali bomo posamezne elemente metod in pristopov ter izvedli primerjalno analizo med njimi. Za določitev učinkovitega pristopa pri zagotavljanju kakovosti podatkov podatkovnega skladišča, ki bo učinkovito izrabljaj vire in nudil najboljše možne analize iz obstoječih podatkov, je potrebno poznavanje lastnosti posameznih pristopov. Primerjali smo:

- izhodišča,
- način merjenja kakovosti,
- razvojni cikel,
- druge parametre (pomen metapodatkov, potrebna pripravljenost v večja vlaganja v podatkovna skladišča).

### 8.1 Izhodišča

Metode za razvoj podatkovnih skladišč in tudi obravnavana razvojna metoda DWM, obravnavajo kakovost podatkov kot pomembno in jo vključujejo v vse korake razvoja podatkovnega skladišča. Zagotavljanje kakovosti podatkov obravnavajo kot pogoj za uspešno izvedbo projekta.

Pristop k povečanju kakovosti podatkov zasnovanem na tveganjih gleda na proces upravljanja podat-

kov z vidika odprave tveganj za učinkovito zagotavljanje informacij iz podatkovnega skladišča.

V pristopu k povečanju kakovosti podatkov in ROI je izhodišče izboljšati kakovost podatkov tako, da bo učinek podatkovnega skladišča in ROI največji.

### 8.2 Način merjenja kakovosti

Oba primerjana pristopa predlagata za merjenje definirane pravil in meritev odstopanj podatkov od definiranih pravil. V pristopu k povečanju kakovosti podatkov in ROI določimo pravila za kakovost podatkov na osnovi škode za učinkovito delovanje sistema; večja je le-ta, pomembnejše je pravilo. Pravil za kakovost podatkov ne dobimo z določitvijo, kakšni naj bodo podatki, ampak z oceno vpliva nekakovostnih podatkov na uspešnost delovanja sistema. Pravilom, ki opisujejo zahteve za kakovost podatkov, ki, če so nekakovostni, povzročijo največjo škodo, damo največji pomen in težo.

V pristopu k povečanju kakovosti podatkov zasnovanem na tveganjih definiramo pravila glede na skupne lastnosti kakovosti podatkov po definirani metriki (tabela 1) predvsem za podatke, ki povzročajo večje tveganje.

### 8.3 Razvojni cikel

Ugotovimo lahko, da pa se predstavljeni razvojni cikli obravnavanih metod bistveno ne razlikujejo med seboj (tabela 4). Pristop za povečanje kakovosti podatkov zasnovan na tveganjih, veliko bolj natančno opisuje in poudarja, kako v posameznem koraku razvoja pazimo na kakovost, kot ostale obravnavane metode. To je v skladu z izhodiščem pristopa. Pristop k zagotavljanju kakovosti in ROI pa daje večji poudarek ocenitvi vpliva nivojev kakovosti in stroškom izboljšav kakovosti podatkov, saj izhaja iz drugega izhodišča (tabela 4).

### 8.4 Drugi parametri

Pristop k povečanju kakovosti podatkov zasnovanem na tveganjih daje večji pomen metapodatkom kot pristop k povečanju kakovosti podatkov in ROI, saj natančno opisani metapodatki zmanjšujejo tveganja razvoja in napačnega razumevanja in uporabe dobljenih podatkov iz podatkovnega skladišča. Tudi razvojna metoda DWM daje velik poudarek metapodatkom, saj niso pomembni le za kakovost ampak tudi za uspešen razvoj.

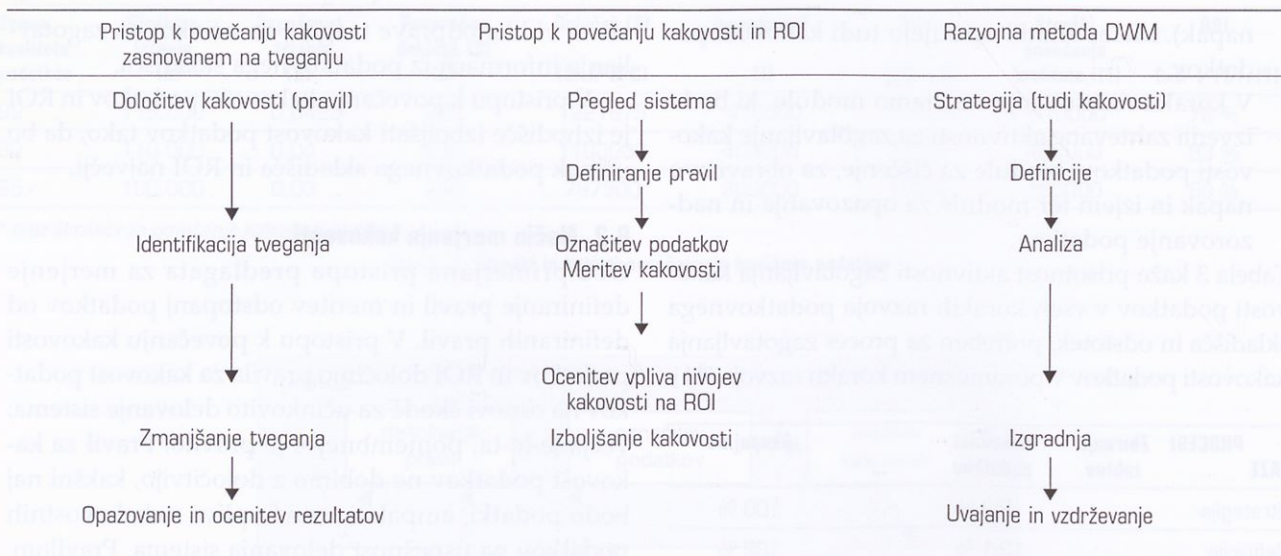


Tabela 4: Razvojni cikli

Za pristop zasnovan na tveganjih je potrebna pripravljenost podjetja v večja vlaganja v podatkovna skladišča, saj je bolj celovit in dosleden in sam ne poudarja dobičkonosnosti investicije v podatkovno skladišče, tako kot pristop k povečanju kakovosti podatkov in ROI.

### 8.5 Izbira pristopa povečanja kakovosti podatkov

Pomembna razlika med pristopi je v pojmovanju vloge podatkovnega skladišča. Pri prvem obravnavanem pristopu k zagotavljanju kakovosti podatkov zasnovanem na tveganjih, mora biti podatkovno skladišče brezhiben in izjemno kakovosten vir ter zelo zanesljiva informacijska podpora za poslovno odločanje. Pri drugem obravnavanem pristopu, kjer gre za povečanje kakovosti podatkov in ROI, pa je podatkovno skladišče kakovosten vir za podporo odločanju in orodje za povečanje dobička in se mora investicija vanj tudi čim bolje poplačati. Prvi pristop je bolj celovit, drugi pa je usmerjen bolj na dobičkonosnost. Prvi pristop zmanjša tveganja, drugi pa poveča dobiček.

Ne glede na izbrani pristop pa moramo upoštevati, da so podatkovna skladišča z neakovostnimi podatki le hitra rešitev [9,20], ki jo je treba kmalu dograditi. Projekt razvoja podatkovnih skladišč mora biti izpeljan s spremenljivimi stroški in mora zadovoljiti zahteve uporabnikov ter nuditi dovolj kakovostne podatke.

Pristop k povečanju kakovosti podatkov, zasnovanem na tveganjih, izberemo, ko obstaja:

- zahteva po zelo zanesljivi in brezhibni podpori odločanju,
- želja po zelo kakovostni podpori odločanju,
- z viri dobro podkrepljen projekt.

Pristop k povečanju kakovosti podatkov in ROI izberemo, ko:

- ni zahtevano podatkovno skladišče z zelo kakovostnimi podatki,
- je zahtevan čim večji izplen (povračilo investicije) izdelanega podatkovnega skladišča.

Razvojno metodo DWM uporabimo, ko:

- problem kakovosti ni posebej pereč, ni zahtevana izjemno velika kakovost podatkov,
- kakovost lahko obravnavamo enokovredno z drugimi razvojnimi procesi in izzivi v standardnem zaporedju aktivnosti razvojnega cikla podatkovnega skladišča.

## 9 Sklep

Izbira metode za merjenje kakovosti podatkov in metode za zagotavljanje in povečanje kakovosti podatkov je pomembna za uspešno izdelavo podatkovnega skladišča.

Za zagotavljanje kakovostnega podatkovnega skladišča je zelo pomemben korak izbire metode merjenja, oziroma ocenitve kakovosti. Pomembno je določiti pravila, ki jih morajo kakovostni podatki zagotavljati. Orodja (kot je Oracle Data Quality Inspector [17]) olajšajo merjenje. Za gradnjo celovitega kakovostnega

podatkovnega skladišča je uporabna obravnavana metoda, ki temelji na tveganjih in nudi metrike za merjenje vseh vidikov kakovosti. Toda realnost je daleč od idelanege sveta in največkrat je pač potrebno meriti in izboljšati tiste lastnosti kakovosti tistih podatkov, ki so potrebne za uspešnost projekta. Kakovost podatkov, ki je manjša kot 90 %, negativno vpliva na poslovanje. Velik korak pa storimo že s tem, če se zavedamo pomena kakovosti podatkov in potrebe po merjenju kakovosti, saj brez merjenja ni možno izvesti izboljšanja in nadzora kakovosti podatkov.

Splošna najboljša metoda za zagotavljanje in povečanje kakovosti podatkov ne obstaja, saj ima vsaka svoje lastnosti. Zato se moramo odločiti za tisto, ki najbolj ustreza dejavnikom v projektu razvoja podatkovnega skladišča. Za brezhibno in zelo zanesljivo podatkovno skladišče uporabimo pristop k povečanju kakovosti podatkov zasnovanem na tveganju. Za projekte razvoja podatkovnih skladišč, ki zahtevajo čim večji ROI, izberemo pristop k povečanju kakovosti podatkov in ROI. Za projekte, kjer lahko kakovost obravnavamo enakovredno z ostalimi razvojnimi izzivi uporabimo metodo DWM.

Ne glede na izbran pristop in razvojna orodja za gradnjo podatkovnih skladišč mora biti projekt izpeljan s sprejemljivimi stroški, zadovoljiti trenutne potrebe uporabnikov, zgraditi dovolj prilagodljivo podatkovno skladišče ter nuditi dovolj kakovostne podatke.

## Literatura

- [1] Barbusinski, B., Howard, S. Kelley, C. (2002): How would you characterize the difference between Bill Inmon's philosophy on data warehousing and Richard Kimball's?, DM Review.
- [2] Current Analysis, (2001): Data Quality Product Assessment.
- [3] Dubois., L. (2002): Business Intelligence: The Dirty (and Costly) Little Secret of Bad Data, BI Report.
- [4] English, L. P. (1999): Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits, New York: John Wiley & Sons.
- [5] English, L. P. (2002): The Essentials of Information Quality Management, DM Review.
- [6] Hacknez, D. (2003): Data Warehouse Delivery: Data Quality Fear, DM Review.
- [7] Gallas, S. (1999): Kimball Vs. Immon", DM Review.
- [8] Groh, T. (2004): Beyond ROI ... Justifying a Business Intelligence Initiative, DM Review.
- [9] Global Data Management Survey (2001): PriceWaterhouseCooper, <http://pwcglobal.com>.
- [10] Hufford, D. (1996): Data Warehouse Quality: Special Feature from January 1996.
- [11] Immon, W. H. (1996): Building the data warehouse, Wiley, New York.
- [12] Kachur, R. J. (2000): The Data Warehouse Management Handbook, Prentice Hall, 2000.
- [13] Kimball, R. (1996): The Data Warehouse Toolkit: Practical Techniques for Building Dimenzional Data Warehouses, John Wiley & Sons, New York.
- [14] Kimball, R. (1999): The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing and Deploying Data Warehouses, Wiley, New York.
- [15] McKnight, W. (2003): Building Business Intelligence: Overall Approach to Data Quality ROI, DM Review.
- [16] McKnight, W.: Overall Approach to Data Quality ROI, White Paper, <http://www.mcknight-associates.com/>.
- [17] Oracle Data Quality Inspector: <http://www.oracle.com/consulting/offerings/platform/index.html?dqi.html>.
- [18] Oracle Method (1998): Oracle Data Warehouse Method Handbook, Oracle.
- [19] Oracle9i Warehouse Builder 9.2, Integrated Data Quality, <http://www.oracle.com>.
- [20] TDWI Report Series, (2002): Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data.

Krista Rizman Žalik ima več kot petnajstletne izkušnje pri analizi podatkov, podatkovnem modeliranju informacijskih sistemov, upravljanju metapodatkov in načrtovanju in razvoju podatkovnih skladišč. Delala je kot projektantka, razvijalka in svetovalka na različnih projektih razvoja informacijskih sistemov in rešitev. V zadnjih petih letih se je posvetila svetovanju, projektiranju in razvoju podatkovnih skladišč.

**SRC SI**  
sistemske integracije



**MARAND**  
Napredna računalniška hiša

# System za upravljanje delovnih procesov

Miroslav Ribič, Andrej Kovačič, Marjan Lončarič

## Povzetek

Sistem za upravljanje delovnih procesov (WFMS) nudi podporo le podrobno opredeljenim poslovnim procesom, zato je njegova uporaba pogojena s prenovo poslovnih procesov, ki postaja nujna za konkurenčnost katerekoli organizacije. WFMS mora organizaciji zagotoviti učinkovito definiranje, upravljanje in izvajanje tako organizacijskih kot medorganizacijskih delovnih procesov. Za upravljavca procesa je pomembno enostavno koordiniranje delovnih procesov in enostavno dodeljevanje nalog, hkrati pa mora WFMS omogočiti, da se proces po potrebi z uporabo drugih nastavitev in različnih programskih rešitev učinkovito prilagodi.

## Abstract

### Workflow Management System

A workflow management system (WFMS) is mainly used for business processes that can be effectively and efficiently specified in detail. As business process reengineering becomes a norm for competitiveness, any future WFMS must be able to respond to organizational changes on an ongoing basis. From the organization's point of view, WFMS must be able to define, manage and execute Inter-Organizational and Intra-organizational workflows. From the supervisor's point of view, workflow management must be able to coordinate easily various workflows, assign the worker's work contents and change the workflow using data and software from different sources.

## 1 Opredelitev delovnega procesa

Upravljanje delovnih procesov pomeni optimizacijo nalog v smislu posredovanja dokumentov in informacij, ki v procesu nastajajo, njihovim izvajalcem, da ti nemoteno opravljajo delo in tako dosegajo poslovne cilje organizacije. Lahko je organizirano tradicionalno, lahko pa je delno ali popolnoma avtomatizirano. Tradicionalno zasnovan delovni proces temelji na uporabi papirnih dokumentov. Tak pristop je v praksi vedno manj pogost, saj se z informatizacijo poslovanja papirne dokumente nadomesti z elektronskimi. Sodoben delovni proces je računalniško podprt, saj je le tako mogoča optimizacija in/ali avtomatizacija nalog. Delovni proces torej opredelimo kot računalniško predstavitev poslovnega procesa, s katerim zagotovimo delno ali popolno avtomatizacijo poslovnega procesa (Hollingsworth, 1995). Opredeljen je z vsemi informacijami o poslovnem procesu, ki so potrebne za njegovo izvršitev. Sem sodijo informacije o sprožilnih in zaključnih pogojih procesa, sestavnih aktivnostih in pravilih za krmiljenje procesa, sklici na izvajalce posameznih aktivnosti ter, ker gre za računalniško predstavitev poslovnega procesa, tudi sklici na programske rešitve.

## 2 Sistem za upravljanje delovnih procesov

Delovni proces je neposredno povezan s prenovo poslovnih procesov, ki se ukvarja z analiziranjem, ocenjevanjem in modeliranjem ključnih poslovnih

procesov v organizaciji, s čimer neposredno vpliva na njihovo operativno implementacijo. Delovni proces igra pri tem vlogo vmesnega člana med poslovnim procesom in programskimi rešitvami, ki podpirajo njegovo izvajanje. Tako zagotovimo neodvisnost logike poslovnih procesov od programskih rešitev, s čimer postane spreminjanje poslovnih procesov v prihodnosti bolj enostavno. Računalniško podprto izvajanje poslovnega procesa terja torej preslikavo poslovnega procesa v delovni proces. Ta pri tem podeduje vse aktivnosti in pravila poslovnega procesa, vanj je treba prek organizacijskega načrta vgraditi še izvajalce posameznih aktivnosti ter sklice na programske rešitve in parametre, ki so potrebni za njihovo uporabo. Pri tem si pomagamo z računalniško podporo za upravljanje delovnih procesov, ki ji pravimo sistem za upravljanje procesov (angl. Workflow Management System, v nadaljevanju: WFMS). WFMS skrbi za avtomatizirano izvajanje delovnih procesov tako, da v pravilnem zaporedju aktivira ustrezne človeške in informacijske vire. Lahko ga opredelimo kot sistem za modeliranje in krmiljenje delovnih procesov (Keller, Teufel, 2000).

Temeljni namen sistema WFMS je torej posredovati pravo nalogo oz. aktivnost pravemu izvajalcu ob pravem času, kar vpliva na bistveno hitrejšo izvajanje



poslovnih procesov. Tovrstni sistem temelji na uporabi celovitih poslovno-informacijskih rešitev (v nadaljevanju: ERP), ki podpirajo upravljanje notranje vrednostne verige, hkrati pa se vključujejo v zunanjo vrednostno verigo. Sestavljen je iz večjega števila medsebojno povezanih in odvisnih osnovnih modulov. Glede na njihov nastanek in dosednji razvoj jih lahko razvrstimo na orodja, ki v večji meri pokrivajo modeliranje, ter orodja, ki so namenjena predvsem učinkovitemu izvajanju (krmiljenju) delovnih procesov. Ločnica med tema dvema značilnostma je vse manj izrazita, saj si proizvajalci orodij prizadevajo priti na tržišče s kar se da celovito ponudbo. Generično arhitekturo sistema za upravljanje delovnih procesov in medsebojne odnose med posameznimi moduli prikazuje slika 1.

Posamezne module sistema za upravljanje delovnih procesov lahko opredelimo z naslednjimi vsebinami: modeliranje in prenova poslovnih procesov, skrbništvo procesov, razvoj programskih rešitev za krmiljenje delovnih procesov, organizacijski načrt, izvajalno okolje in krmiljenje delovnega procesa.

## 2.1 Modeliranje in prenova poslovnih procesov

Prenova poslovnih procesov je temeljito preverjanje in spreminjanje procesov ter pripadajočih aktivnosti z namenom, da bi dosegli pozitivne rezultate na področjih, kot so zniževanje stroškov, povečanje kakovosti proizvodov in storitev, skrajševanje časovnih

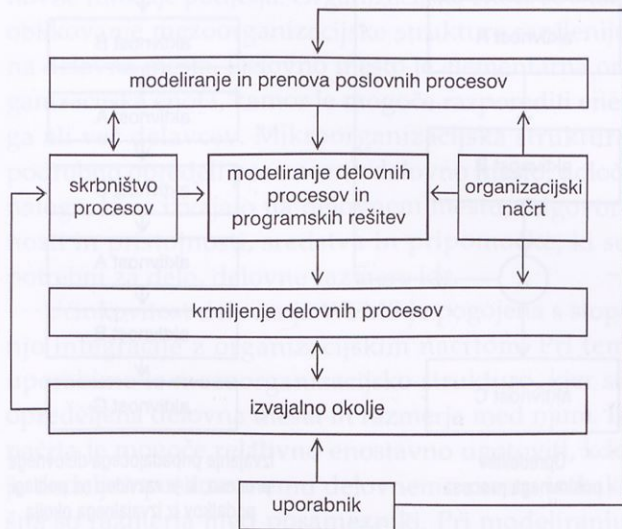
ciklov in podobno (Kovačič, 1998). Na tem področju je velik korak naprej storilo mednarodno združenje BPMI (angl. Business Process Modeling Initiative), ki je v obliki jezika BPML (angl. Business Process Modeling Language) podalo predlog za standardiziran zapis poslovnih procesov. Ker je BPML formaliziran s shemami XML, je kot nalašč za modeliranje organizacijskih in medorganizacijskih delovnih procesov ter podporo elektronskemu poslovanju.

BPML loči tri tipe aktivnosti, in sicer procese, kompleksne in enostavne aktivnosti (Assaf, 2003). *Enostavne aktivnosti* so elementarne in jih vgradimo v kompleksne aktivnosti. Omogočajo izvajanje najbolj osnovnih nalog, kot so izvedbe operacij, izvajanje časovnih zank, dodeljevanje vrednosti lastnostim aktivnosti in odzivanje na spremembe lastnosti v obliki proženja dogodkov, signalov in napak, prek katerih se vršijo klici drugih aktivnosti.

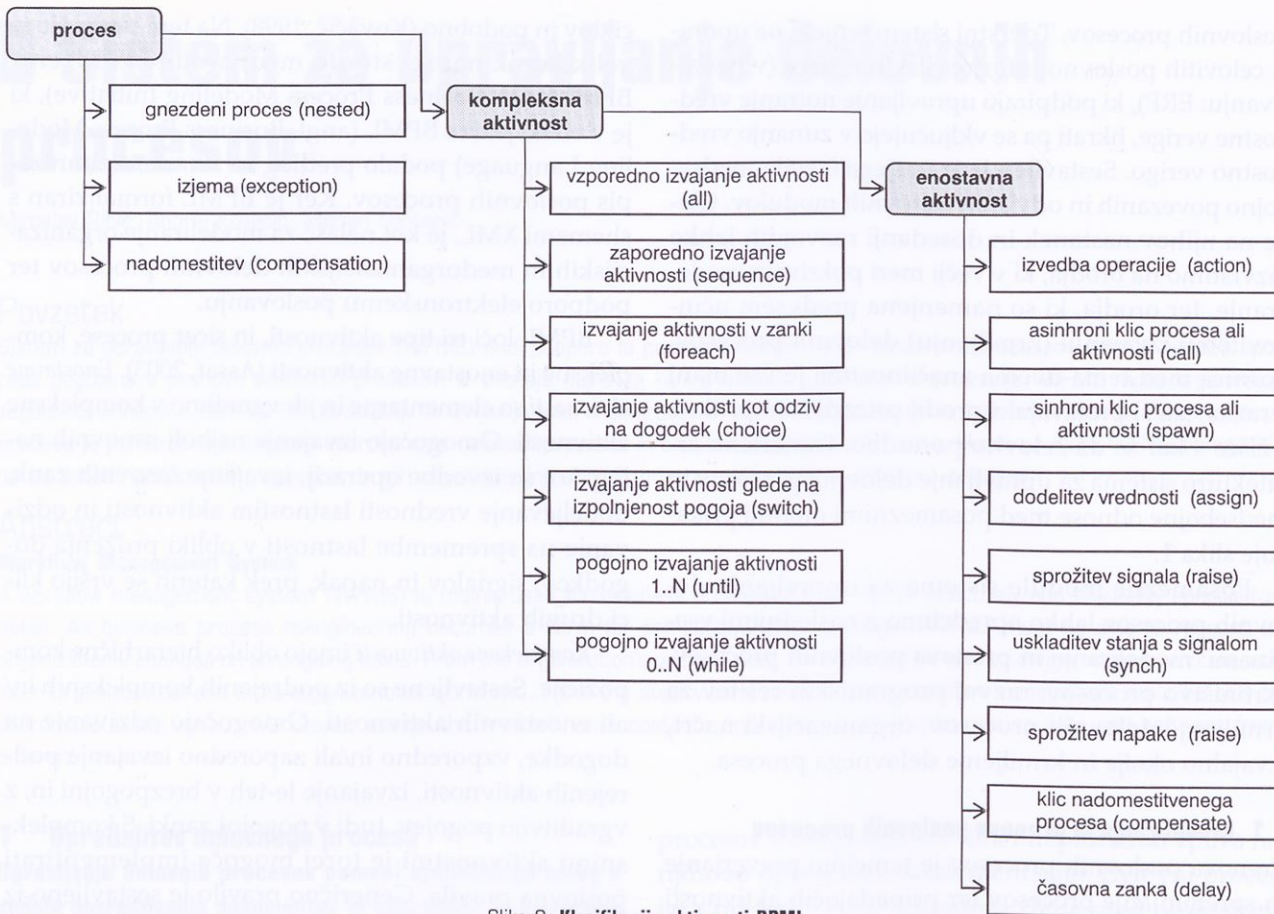
*Kompleksne aktivnosti* imajo obliko hierarhične kompozicije. Sestavljene so iz podrejenih kompleksnih in/ali enostavnih aktivnosti. Omogočajo odzivanje na dogodke, vzporedno in/ali zaporedno izvajanje podrejenih aktivnosti, izvajanje le-teh v brezpogojni in, z vgraditvijo pogojev, tudi v pogojni zanki. S kompleksnimi aktivnostmi je torej mogoče implementirati poslovna pravila. Generično pravilo je sestavljeno iz dogodka, ki ga sproži, kontrolnega mehanizma, v katerega so vgrajeni pogoji, in aktivnosti, ki se prožijo skladno z kontrolnim mehanizmom.

*Proces* je kompleksna aktivnost, ki se izvaja v specifičnem kontekstu. To pomeni, da mora biti ob njegovem izvajanju v vsakem trenutku zagotovljena možnost, da se v primeru izjemnega dogodka ali napake, vzpostavi prvotno stanje. V primeru dogodkov, ko se skozi izvajanje aktivnosti pojavijo izjemni, vendar ne napačni pogoji, se sproži poseben proces za reševanje izjem, v primeru napačnih pogojev pa se sproži nadomestitveni proces, ki vzpostavi prvotno stanje, to je stanje, kakršno je bilo pred začetkom izvajanja procesa.

Za modeliranje poslovnih procesov uporabljamo posebna orodja, ki nudijo poleg zapisa procesnih modelov v obliki BPML tudi metodološka izhodišča za ugotavljanje, zbiranje in analiziranje podatkov o izvajanju poslovnega procesa. Zajema tudi dodatke, namenjene prenovi poslovnih procesov oziroma grafični predstavitvi in ureditvi (poenostavitvi, racionalizaciji in standardizaciji) procesov pred njihovo nadaljnjo informatizacijo. Omogoča modeliranje procesov ter preverjanje njihove celovitosti in povezljivosti, simulacijo izvajanja



Slika 1: Generična arhitektura sistema WFMS



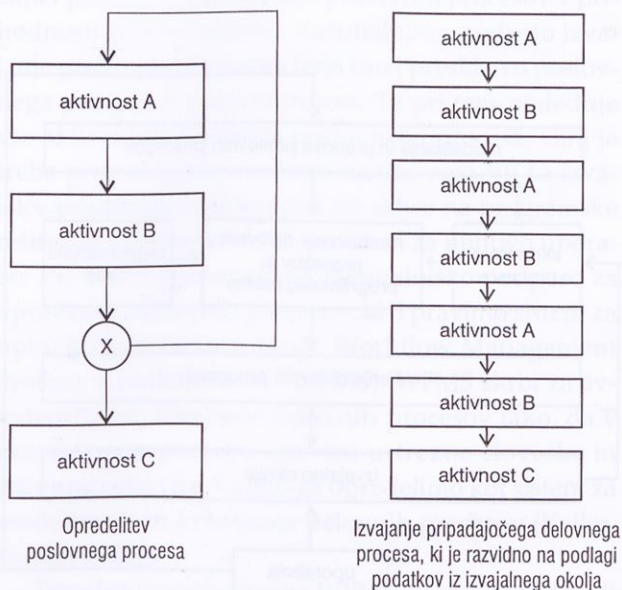
Slika 2: Klasifikacija aktivnosti BPML

procesov oziroma posameznih aktivnosti v povezavi s potrebnimi podatki ter analiziranje ustreznih variantnih rešitev izvajanja procesov v smislu dejavnikov, kot so: stroški (ABC analiza), kritične poti in generiranje novih modelov delovnih procesov, nastalih ob izbiri najustreznejše rešitve. Na tržišču jih srečamo pod nazivi, kot so: ActionWorkflow, Business Modelling Tool, ProcessWise, ARIS, Bonapart, Optima!, Income ...

### 2.2 Skrbništvo procesov

Skrbništvo procesov nadzira izvajanje procesov, uporabo virov, časa in stroškov. Podatki, ki jih nadziramo, izhajajo iz poslovnih dokumentov, ki nastajajo v izvajalnem okolju sistema WFMS.

Pri tem se največkrat uporabljajo podatki o izvajalcu procesa, ki jih kombiniramo s tako imenovani časovnimi žigi (angl. time-stamp), to so podatki, ki časovno določajo nastanek in spremembe poslovnih dokumentov in njihovih poprejšnjih verzij. Analiza



Slika 3: Nadzor nad izvajanjem delovnih procesov

omenjenih podatkov pokaže, kako učinkovito se izvajajo delovni procesi.

Slika 3 prikazuje primer opredelitve poslovnega procesa na levi in njegovega izvajanja skozi pripadajoči delovni proces na desni strani. Iz opredelitve procesa je razvidno, da je treba ob neuspešno opravljeni aktivnosti B ponovno začeti z izvajanjem aktivnosti A. Skrbnik procesa na podlagi analize izvajanja pripadajočega delovnega procesa ugotovi, da se je pred izvajanjem aktivnosti C kar trikrat izvedla zanka z aktivnostma A in B, kar kaže na priložnost za optimizacijo poslovnega in/ali delovnega procesa. Tako se je treba s prenovo poslovnega in/ali delovnega procesa čimbolj približati optimalnemu izvajanju, t. j. »Aktivnost A« → »Aktivnost B« → »Aktivnost C«.

### 2.3 Organizacijski načrt

Lipovec opredeli organizacijo kot sestav medsebojnih razmerij med ljudmi, ki zagotavljajo obstoj, družbeno-ekonomske in druge značilnosti podjetja in smotrno uresničevanje cilja podjetja. Iz dane opredelitve sledi, da organiziranje pomeni vzpostavljanje organizacijskih struktur, iz katerih izhajajo razmerja med ljudmi (Lipovec, 1987).

Načrt organizacijske strukture se praviloma opredeli na treh nivojih, in sicer makro-, mezo- in mikro-nivoju. Na oblikovanje makroorganizacijske strukture vpliva tehnična delitev dela in situacijski dejavniki, kot so tehnologija, konkurenca, kupci, velikost podjetja, kadri idr. Rezultat je organizacijska struktura v obliki organizacijskih enot, iz katerih so razvidne osnovne funkcije podjetja. Organizacijske enote se skozi oblikovanje mezoorganizacijske strukture razčlenijo na delovna mesta. Delovno mesto je elementarna organizacijska enota, kamor je mogoče razporediti enega ali več delavcev. Mikroorganizacijska struktura podrobno opredeli posamezno delovno mesto, določi naloge, ki se izvajajo na delovnem mestu, odgovornosti in pristojnosti, sredstva in pripomočke, ki so potrebni za delo, delovne razmere idr.

Učinkovitost delovanja WFMS je pogojena s stopnjo integracije z organizacijskim načrtom. Pri tem uporabimo le mezoorganizacijsko strukturo, kjer so opredeljena delovna mesta in razmerja med njimi. Iz načrta je mogoče relativno enostavno ugotoviti, kdo je dodeljen posameznemu delovnemu mestu in kakšna so razmerja med posamezniki. Pri modeliranju procesa se tako sklicujemo na delovno mesto, nalogo pa WFMS dodeli posamezniku, ki delovno mesto

zaseda. Tako ostaja delovni proces ob prerazporeditvah ter prihodih in odhodih posameznikov nespremenjen. Potrebe po njegovi spremembi se pojavijo šele, ko se spremeni organizacijski načrt.

Učinkovito delovanje sistema WFMS zahteva torej tesno integracijo s sistemom za upravljanje s človeškimi viri (angl. HRMS – Human Resource Management System), kjer se vzdržujejo podatki o zaposlenih in njihovi umeščenosti v organizacijski načrt. Tovrstna integracija omogoča enostavno dodeljevanje vlog posameznikom v delovnem procesu.

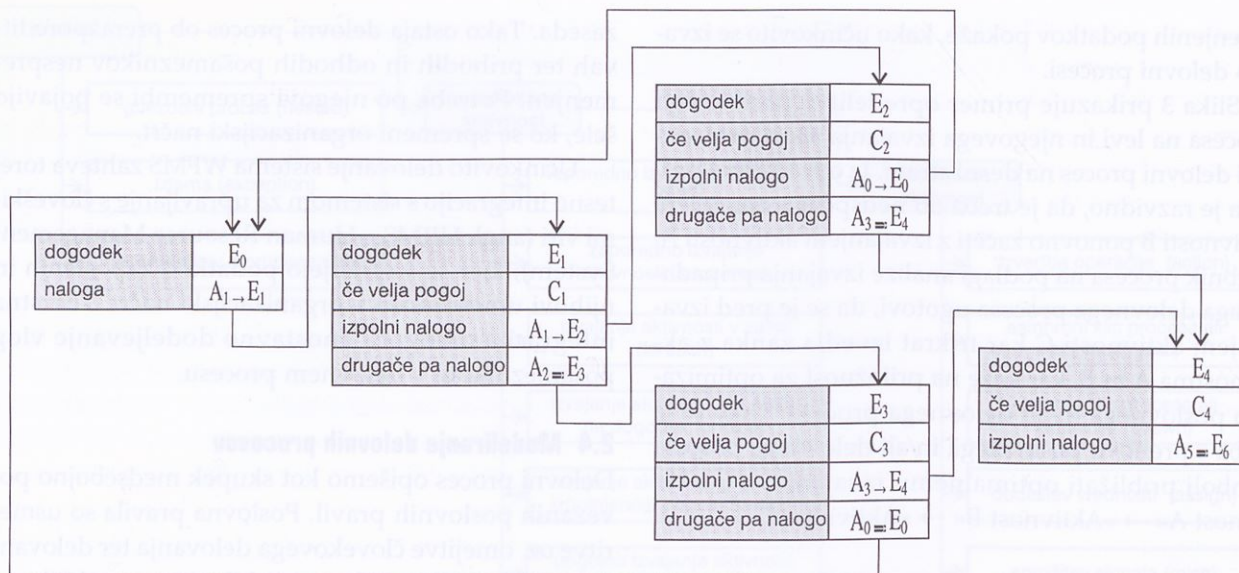
### 2.4 Modeliranje delovnih procesov

Delovni proces opišemo kot skupek medsebojno povezanih poslovnih pravil. Poslovna pravila so usmeritve oz. omejitve človekovega delovanja ter delovanja proizvodnih in informacijskih sistemov. Njihova vsebina mora jasno izražati poslovne zahteve. Vanje je treba vgraditi znanje, pridobljeno iz lastnih izkušenj, najboljšo prakso drugih organizacij ter omejitve, ki izhajajo iz zakonodaje in drugih standardov kakovosti (Endl, Knolmayer, Pharer, 1998).

Modeliranje delovnih procesov je torej oblikovanje poslovnih pravil in njihovo povezovanje. Poslovna pravila morajo biti na nivoju WFMS jasno izražena, in sicer v obliki deklarativnih stavkov, ki imajo enotno strukturo. V stavek oz. pravilo je treba vgraditi logične gradnike, ki so lahko dogodki, kontrolni mehanizmi ali naloge. Z dogodkom se opredeli, kdaj se poslovno pravilo izvede, kontrolni mehanizem določi, kaj vse mora biti pred tem preverjeno, naloga pa, kaj vse mora biti narejeno. Skladno s tem lahko poslovna pravila in sam proces zapišemo v obliki notacije ECAA (angl. Event, Condition, Action, Alternative action).

Iz slike 4 je razvidno, da je modeliranje delovnih procesov preprosto povezovanje poslovnih pravil. Pravila se povezujejo v delovni proces prek dogodkov, ki se zgodijo, ko je neka naloga uspešno zaključena, ali kako drugače ustavljena. Delovni proces se vedno izvaja na zahtevo, saj bi v nasprotnem primeru prišlo do nesmiselne porabe virov. Izvajanje se torej sproži, ko se zgodi ustrezen zunanji dogodek, to je dogodek, na katerega sam proces ne more neposredno vplivati.

Sprožitev pomeni začetek izvajanja prve naloge. Samo izvajanje je treba dokumentirati, saj se s tem dokazujejo rezultati in učinkovitost izvajanja nalog. Tako nastajajo ali pa se spreminjajo poslovni dokumenti. Ob zaključku naloge se torej pripravi nov dokument ali pa



Slika 4: Z notacijo ECAA zapisana in v delovni proces povezana poslovna pravila

se z novo vrednostjo statusnega atributa popravi obstoječi. Kreiranje novega dokumenta oz. popravek obstoječega je pravzaprav nov dogodek, ob katerem WFMS sproži naslednje poslovno pravilo. Njegov kontrolni mehanizem preveri izpolnjevanje pogojev, pri čemer se vedno sklicuje na poprejšnje dokumente. Kontrolni mehanizem tako določi izvajanje naslednje naloge, in sicer glede na podatke, ki iz omenjenih dokumentov izvirajo. Tako različni statusi dokumenta pomenijo izvajanje naslednje naloge z drugačnimi vhodnimi parametri, v primeru izjem ali napak pa se v nadaljevanju izvede popolnoma druga naloga. Vsaka naloga se sklicuje na metodo implementiranega poslovnega objekta. Pri tem velja, da morajo biti podatki, ki nastanejo ob dogodku, zadostni za klic ustrezne metode. Z vidika WFMS poznamo:

- *Avtomatizirane naloge*, ki se izvajajo v ozadju. To pomeni, da se ob določenem dogodku izvede klic ustrezne metode poslovnega objekta, ki izvede nalogo brez prekinitev. Izvedba naloge je odvisna le od njene kompleksnosti in količine sistemskih virov, ki so na razpolago.
- *Interaktivne naloge*, ki se izvajajo skozi interakcijo z uporabnikom. Pri tovrstnih nalogah je treba poleg sklica na metodo poslovnega objekta opredeliti tudi izvajalca naloge. Za izvajalca se ponavadi opredeli organizacijsko enoto oz. delovno mesto iz organizacijskega načrta. WFMS v okviru krmilje-

nja delovnega procesa iz razporeditev na opredeljeno delovno mesto dodeli nalogo najprimernejšemu izvajalcu.

Tako avtomatizirane kot interaktivne naloge so implementirane v obliki programskih rešitev, ki morajo biti dovolj fleksibilne, da jih lahko sistem WFMS uporabi pri izvajanju delovnega procesa. Izkušnje zadnjih let kažejo na uporabo objektne pristopa k modeliranju programskih rešitev.

## 2.5 Modeliranje programskih rešitev

V vsako poslovno programsko rešitev so vgrajena poslovna pravila, ki opredeljujejo poslovne procese. Zaradi nenehnih sprememb, ki se dogajajo na trgu, je treba poslovna pravila nenehno spreminjati. Zato je treba zagotoviti dovolj fleksibilno arhitekturo programskih rešitev, da je upravljanje s poslovnimi pravili in procesi enostavnejše. Pravi pristop k razvoju tovrstnih sistemov je koncept objektne modeliranja in objektne razvoja informacijskih sistemov. Koncept temelji na razumevanju in modeliranju realnega sveta v obliki poslovnih objektov. Objekt lahko razumemo kot preslikavo oz. poenostavitev entitete iz realnega sveta v model, ki vsebuje njene lastnosti in vedenjske vzorce. Medtem ko posamezna stanja objekta opišemo z lastnostmi, se vedenjske vzorce vgradi v objekt v obliki metod in dogodkov. Temeljni elementi objekta so torej:

- *Lastnosti*, ki so tiste bistvene karakteristike, ki označujejo entiteto iz realnega sveta. Vsaka lastnost ima svoje ime in vrednost. Njen temeljni namen je, da se na podlagi primerjave vrednosti istoimenskih lastnosti lahko identificira in med seboj razlikuje istovrstne objekte.
- *Metode* so storitve, ki jih objekt opravlja. Metoda je v tehničnem smislu funkcija, ki vhodne parametre spremeni v izhodne. Njena vsebina je praviloma skrita, medtem ko mora biti njen vmesnik z vhodnimi in izhodnimi parametri znan. Objekt skozi metode izraža svoje vedenje. Pri tem uporablja vrednosti svojih lastnosti kot vhodne parametre. Skozi izvajanje metod se opredelijo vrednosti izhodnih parametrov, ki so nove vrednosti lastnosti objekta. V samo vsebino metod so lahko vgrajeni tudi klici metod drugih objektov. Tako lahko objekt skozi izvajanje svojih metod spreminja vrednosti tako svojih kot tudi lastnosti drugih objektov.
- *Dogodki* so po svoji obliki podobni metodam, razlika je le v tem, da imajo v svojem vmesniku samo vhodne parametre. Dogodki so vedno vgrajeni v metode. To pomeni, da metoda izvede klic dogodka, ko se skozi njeno izvajanje ustrezno spremeni vrednost lastnosti obravnavanega ali drugih objektov. Ob klicu dogodka metoda napolni vrednosti vhodnih parametrov in tako sproži dogodek. Objekt se torej skozi dogodka odziva na spremembe lastnosti svojih ali drugih objektov.

Skozi modeliranje objektov je treba zagotoviti, da so interni elementi objekta, to so lastnosti, metode in dogodki, med seboj čim bolj povezani. V dobro oblikovanem objektnem modelu je poskrbljeno, da so za klic metode vsi potrebni podatki pripravljeni že v obliki lastnosti objekta. S tem se prepreči klice metod drugih objektov, da se pridobijo ustrezni podatki za klic obravnavane metode. Dober objektni model ima torej za klice metod oblikovane enostavne vmesnike. To pomeni, da je stopnja neodvisnosti med objekti različnega tipa visoka, kar se odraža v enostavni uporabi objektov, ki je nujna za implementacijo WFMS.

Pri modeliranju objektnih modelov se je za zelo primerne izkazal jezik UML. Namenjen je vizualizaciji, modeliranju, gradnji in dokumentiranju informacijskih sistemov, ki so objektno orientirani. Podpira torej celoten cikel razvoja programske opreme, še posebej pa poudarja izkušnje in potrebe uporabnika. Jezik je enostaven in splošno uporaben, zato ima veli-

ko izrazno moč. Ker ima jasno definirano strukturo, ga podpirajo mnoga CASE orodja. Z uporabo metodologije UML predstavimo vsak sistem s petimi različnimi pogledi, ki so različne preslikave organizacije in strukture sistema (Sturm, 1999):

- *Pogled na uporabniške zahteve* (angl. Use case view), ki ga predstavimo skozi diagrame primerov uporabe (angl. Use case diagram), diagrame zaporedja (angl. Sequence diagram) in diagrame sodelovanja (angl. Collaboration diagram). Z njimi opišemo obnašanje sistema skozi oči uporabnikov.
- *Načrtovalski pogled* (angl. Design view), ki skozi razredni (angl. Class diagram) in objektni diagram (angl. Object diagram) opiše vse funkcijske zahteve sistema.
- *Procesni pogled* (angl. Process view), s katerim se na podlagi diagramov aktivnosti (angl. Activity diagram) in diagramov načrtovanja stanj (angl. Statechart diagram) zagotovijo mehanizmi hkratnosti in sinhronizacije delovnih procesov.
- *Pogled na komponente sistema* (angl. Component View), ki obravnava razdelitev informacijskega sistema na komponente in datoteke, ki so potrebne za njegovo fizično implementacijo.
- *Pogled na razvrstitev* (angl. Deployment View), s katerim predstavimo zahtevano namestitev strojne opreme, ki je potrebna za optimalno delovanje sistema.

CASE orodja, ki podpirajo metodologijo UML, na podlagi omenjenih diagramov s pomočjo generatorjev programske kode oblikujejo objektno orientirane programske rešitve. Take rešitve so, če so pravilno načrtovane, dovolj fleksibilne za modeliranje delovnih procesov in uporabo s strani sistema WFMS. Vendar je na tem mestu treba poudariti, da je kakovost tovrstnih programskih rešitev izrazito odvisna od kakovosti in podrobnosti modeliranja diagramov.

## 2.6 Izvajalno okolje

WFMS izvaja delovni proces skozi proženje programskih rešitev, in sicer v obliki avtomatskih aktivnosti v ozadju ter skozi dodeljevanje interaktivnih nalog, to je nalog, ki se izvajajo skozi interakcijo z uporabnikom. WFMS torej skrbi za pravilno in pravočasno proženje nalog, njihovo dejansko izvedbo pa podpirajo programske rešitve, na katere se sklicuje delovni proces, ki ga WFMS izvaja. Skozi izvajalno okolje se spreminjajo obstoječi ter nastajajo novi poslovni dokumenti, ki dokazujejo poslovne dogodke. Izvajalno

okolje predstavlja torej dejansko sliko izvajanja delovnega procesa skozi očala uporabnika, ki ustvarja in spreminja poslovne dokumente, ter skrbnika procesa, ki na podlagi podatkov, ki izhajajo iz dokumentov, nenehno išče priložnosti za izboljšanje poslovnih procesov.

### 2.7 Krmilnik delovnega procesa

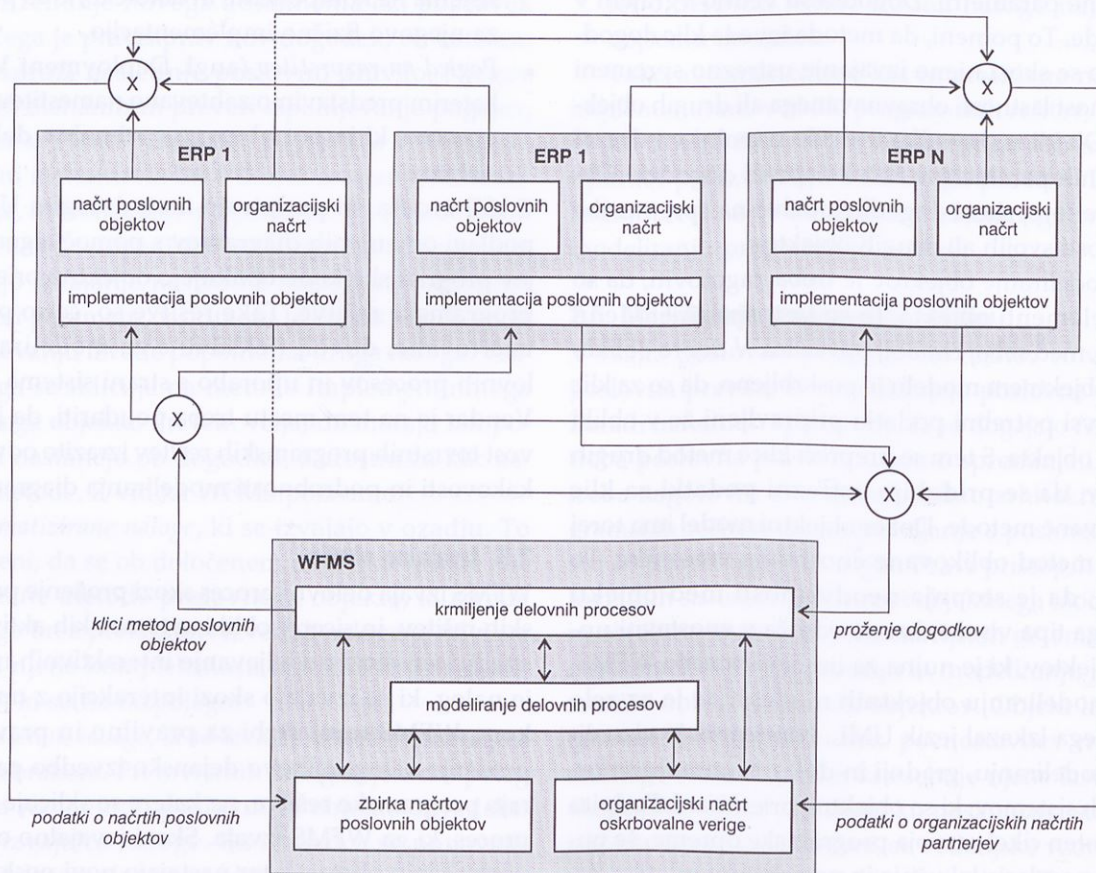
Krmilnik (angl. WorkFlow Management Engine) proži aktivnosti v izvajalnem okolju in nadzoruje izvajanje delovnega procesa. Posamezni proizvajalci navadno ponujajo krmilnike delovnega procesa kot del njihovega izvajalnega okolja. Najdemo jih pod nazivi, kot so ActionWorkflow, CSE, EpisWorkflow, FlowMark, Oracle Workflow, Staffware, SAP Workflow idr.

## 3 Arhitektura sistema za upravljanje medorganizacijskih procesov

WFMS lahko uporabimo tudi kot sistem, ki prek upravljanja poslovnih objektov različnih sistemov ERP usmerja medorganizacijske procese. Če pozna poslovne objekte enega sistema ERP, gre za upravljanje

procesa znotraj organizacije, v nasprotnem primeru pa za upravljanje medorganizacijskih procesov. Pri tem opravlja dve temeljni funkciji, in sicer modeliranje in krmiljenje procesov. Obe funkciji se izvajata na podlagi zbirke načrtov poslovnih objektov ter organizacijskih načrtov. Zbirka načrtov poslovnih objektov je pravzaprav metapodatkovna zbirka, ki jo WFMS uporablja tako pri modeliranju kot pri koordinaciji izvajanja procesov. V zbirki so poleg seznama poslovnih objektov, ki pripadajo sistemu(om) ERP, ki jih WFMS upravlja, tudi informacije o uporabi le-teh, torej nabor lastnosti posameznega objekta, njegovih metod in dogodkov vključno z natančnimi načrti njihovih vmesnikov (SAP Library – SAP Business Workflow).

Na podlagi organizacijskega načrta se dodeljujejo naloge. Sestavni del organizacijskega načrta je načrt organizacijske strukture in razporeditve posameznikov na delovna mesta, ki so elementarne organizacijske enote. Medtem ko se načrt organizacijske strukture uporablja pri modeliranju procesov, se podatki o razporeditvah na delovna mesta, vključno s komunikacijskimi podatki, uporabijo za krmiljenje procesov.



Slika 5: Upravljanje medorganizacijskih delovnih procesov

### 3.1 Modeliranje delovnih procesov

Poslovna pravila in njihove logične komponente preslikamo v podatkovni model sistema WFMS, ki je lahko podoben idejnemu načrtu iz slike 6. Tako zagotovimo zbiranje podatkov o organizacijskem načrtu, poslovnih objektih in procesnih modelih, po katerih se izvajajo delovni procesi. Podatkovni model WFMS je torej sestavljen iz treh logičnih sklopov: organizacijski načrt, načrt poslovnih objektov in načrt poslovnih pravil oz. procesnih modelov.

*Organizacijski načrt*, prek katerega WFMS dodeli naloge njihovim izvajalcem. Z uvedbo entitete »Organizacija« se razširi delovanje WFMS na upravljanje medorganizacijskih procesov. Tako je vsaka »Organizacijska enota« opredeljena tudi s šifro organizacije, prav tako pa se s pomočjo entitet »Organizacijska struktura« ne gradijo razmerja samo med organizacijskimi enotami ene organizacije, ampak tudi razmerja med organizacijami. Tako se pridobi nabor vseh potencialnih izvajalcev nalog v oskrbovalni verigi. WFMS jim dodeli naloge prek kontrolnih mehanizmov, ki so vgrajeni v poslovna pravila. Tako sta skozi entiteto »Kontrolni mehanizem« opredeljena organizacijska enota, ki je izvajalec naloge, in sama naloga, ki mora biti izvedena, če so pogoji, na katere se kontrolni mehanizem sklicuje, izpolnjeni. Ker organizacijska enota kot organizacijski pojem ne more biti neposredni izvajalec naloge, je treba organizacijski načrt dopolniti z entitetama »Razporeditve« in »Delavec«. Skozi entiteto »Razporeditve« so na delovna mesta, ki so elementarne organizacijske enote, razporejeni delavci. WFMS torej prek entitet »Razporeditve« in »Delavec« pridobi potrebne komunikacijske podatke, npr. uporabniško ime ali naslov elektronskega predala, ki jih uporabi, da nalogo dodeli konkretnemu izvajalcu.

*Načrt poslovnih objektov* je definiran z več entitetami. Temeljna entiteta je »Poslovni objekt«, v kateri se vsak objekt opredeli s šifro organizacije, ki ji pripada, in načinom proženja. Vsak izmed objektov ima v ustreznih entitetah opredeljene lastnosti, metode in dogodke. Za vsako metodo oz. dogodek se prek entitete »Parameter« določijo parametri vmesnika. Vsakemu načrtu parametra je treba določiti podatkovni tip, smer, torej vhodni ali izhodni parameter, ter obveznost uporabe, ki je lahko obvezna ali opcijska.

*Načrt poslovnih pravil* predstavlja procesne modele, saj so pravila, ki so med seboj povezana prek dogodkov, temeljni elementi procesnega modela. Poslovno pravilo se torej opredeli s pomočjo entitete »Poslovno

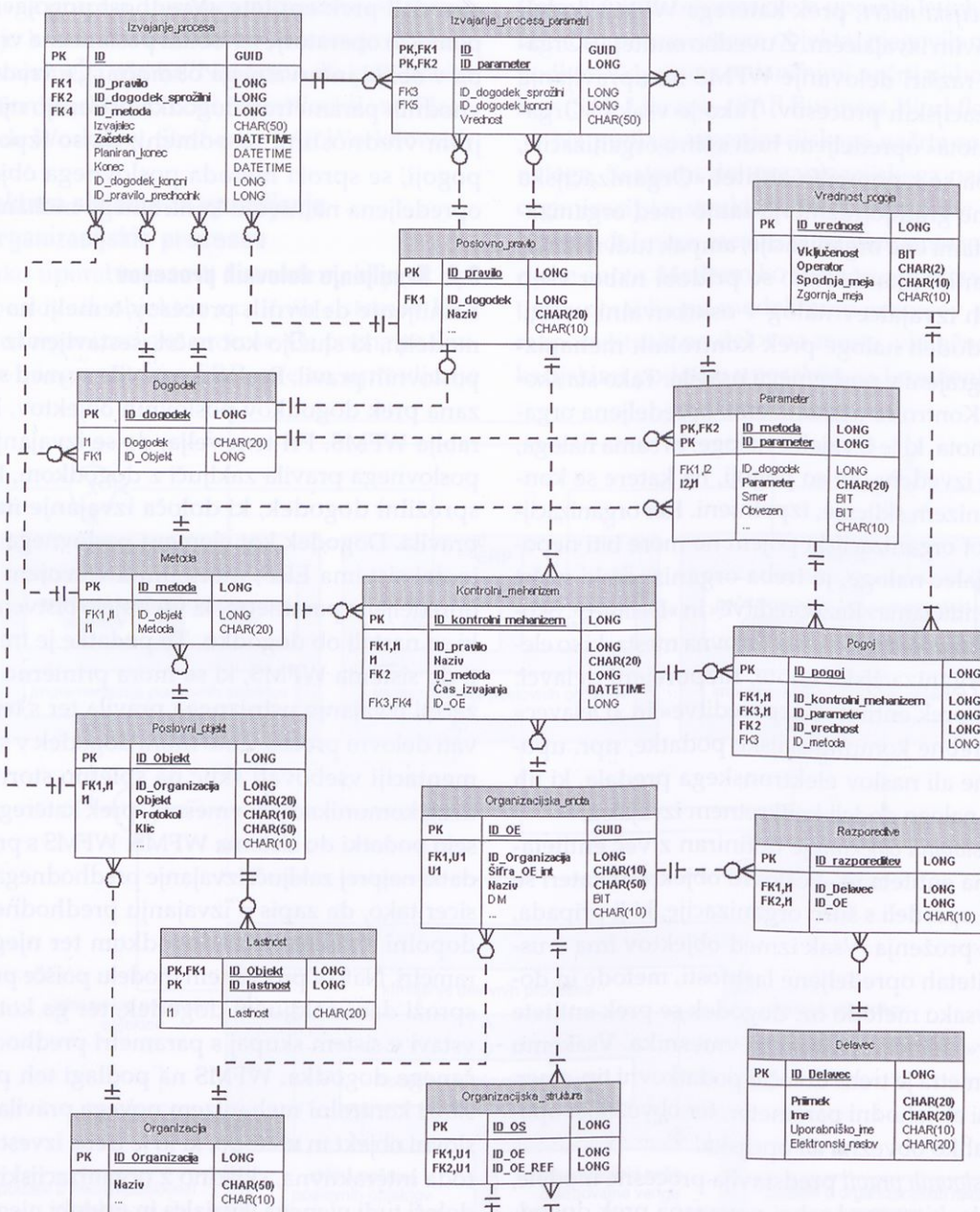
pravilo«, kjer se določi sprožilni dogodek in nabor kontrolnih mehanizmov, ki obravnavajo podatke, ki ob dogodku nastanejo. Vsak kontrolni mehanizem vsebuje poleg sklica na nalogo, to je metodo »Poslovnega objekta«, in izvajalca, to je šifra organizacijske enote, tudi sklic na enega ali več pogojev, ki morajo biti izpolnjeni, da se naloga dodeli izvajalcu. Vsak pogoj se sklicuje na vhodni parameter sprožilnega dogodka ter na ciljno vrednost parametra, ki jo je treba doseči, da je pogoj izpolnjen. Ciljne vrednosti se opredeli prek entitete »Vrednost pogoja«, in sicer s pomočjo operatorjev v obliki posamezne vrednosti ali pa v obliki intervalnega območja. Če vrednosti vseh vhodnih parametrov dogodka ustrezajo njihovim ciljnim vrednostim, to pomeni, da so izpolnjeni vsi pogoji, se sproži metoda poslovnega objekta, ki je opredeljena na nivoju kontrolnega mehanizma.

### 3.2 Krmiljenje delovnih procesov

Krmiljenje delovnih procesov temelji na procesnih modelih, ki služijo kot načrt, sestavljen iz zaporedja poslovnih pravil. Poslovna pravila so med seboj povezana prek dogodkov poslovnih objektov, ki jih uporablja WFMS. Pri tem velja, da se izvajanje vsakega poslovnega pravila zaključí z dogodkom, ki je hkrati sprožilni dogodek, ki določa izvajanje naslednjega pravila. Dogodek kot element poslovnega objekta, ki je del sistema ERP, mora imeti v svojem vmesniku opredeljene parametre, ki vsebujejo bistvene podatke, ki so nastali ob dogodku. Te podatke je treba prenesti do sistema WFMS, ki se mora primerno odzvati in začeti izvajanje ustreznega pravila ter s tem nadaljevati delovni proces. Zato mora dogodek v svoji implementaciji vsebovati sklic na spletno storitev ali kak drug komunikacijski vmesnik, prek katerega se prenesejo podatki do sistema WFMS. WFMS s prejetimi podatki najprej zaključí izvajanje predhodnega pravila, in sicer tako, da zapis o izvajanju predhodnega pravila dopolni z zaključnim dogodkom ter njegovimi parametri. Nato v procesnem modelu poišče pravilo, ki ga sproži dani zaključni dogodek, ter ga kot nov zapis vstavi v sistem skupaj s parametri predhodno zaključenega dogodka. WFMS na podlagi teh parametrov skozi kontrolni mehanizem novega pravila določi poslovni objekt in metodo, ki jo je treba izvesti. Če je metoda interaktivna, skladno z organizacijskim načrtom določi tudi njenega izvajalca in pridobi njegove komunikacijske podatke. WFMS nato na podlagi podatkov o načinu klica poslovnega objekta pošlje parametre

ustreznemu vmesniku do sistema ERP, npr. spletni storitvi. Spletna storitev uporabi prejete parametre za dejanski klic metode poslovnega objekta ter tako sproži njeno izvajanje. Parameter, ki je pri tem vedno prisoten, je identifikacijska številka pravila (v nadaljevanju: ID pravila), ki se trenutno izvaja v sistemu WFMS. Ko je metoda izvršena, se sproži dogodek, ki ima v svoji implementaciji vsebovan sklic na vmesnik

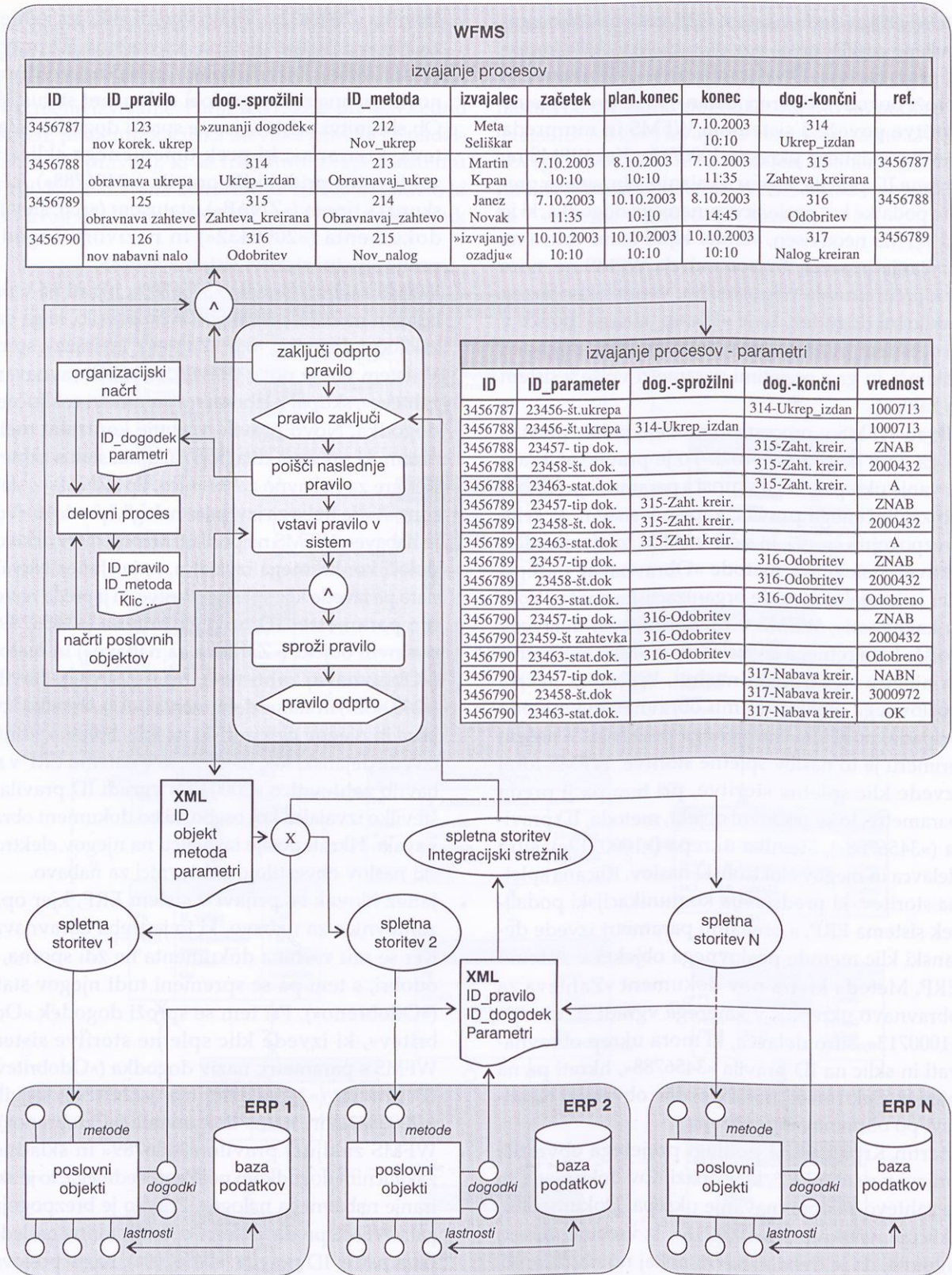
do WFMS, npr. spletno storitev. Metoda tako prek dogodka in spletne storitve posreduje sistemu WFMS ID pravila, ki je v izvajanju, ter ostale podatke, ki so nastali ob proženju dogodka. WFMS na podlagi prejetega ID zaključi izvajanje pravila in začne izvajanje novega, s tem pa je krog sklenjen. Idejni načrt arhitekture in delovanja sistema WFMS lahko s pomočjo slike 7 opišemo v naslednjih korakih:



Slika 6: Idejni načrt podatkovnega modela WFMS



- Meta Seliškar doda v sistem ERP, kot posledico ugotovljene neskladnosti, nov dokument v obliki korektivnega ukrepa. Ob shranjevanju dokumenta se sproži dogodek »Ukrep\_izdan«, ki se prek spletne storitve poveže s sistemom WFMS in mu preda številko izdanega ukrepa »1000713«. Ker WFMS ne prejme ID pravila, ki je v izvajanju, obravnava prejete podatke kot posledico zunanjega dogodka, ki je od WFMS neodvisen, vendar kljub temu vpliva na izvajanje procesa. To pomeni, da WFMS ne more zaključiti pravila neposredno, vendar ker pozna zaključni dogodek, to je »Ukrep\_izdan«, poišče v procesnem modelu pravilo, ki ga dani dogodek zaključí, in ga s prejetimi parametri vpiše v sistem in zaključí.
- Hkrati WFMS v procesnem modelu poišče pravilo, ki ga dani dogodek sproži. To je pravilo »Obravnavanje ukrepa«, ki ga skupaj s parametri predhodno zaključenega pravila, zapiše v sistem. Pravilo brezpogojno sproži, in sicer v obliki izvajanja interaktivne naloge oz. metode »Obravnavaj\_ukrep«. Ker je izvajalec metode organizacijska enota »Vodja kakovosti«, WFMS v razporeditvah delavcev poišče konkretnega izvajalca, to je »Martin Krpan«, in njegov elektronski naslov. WFMS iz načrta poslovnega objekta, ki mu obravnavana metoda pripada, poišče način njenega proženja. V našem primeru je to naslov spletne storitve. WFMS torej izvede klic spletne storitve, pri tem pa ji preda parametre, to so poslovni objekt, metoda, ID pravila (»3456788«), številka ukrepa (»1000713«), šifra delavca in njegov elektronski naslov. Klicana spletna storitev, ki predstavlja komunikacijski podaljšek sistema ERP, s prejetimi parametri izvede dejanski klic metode poslovnega objekta v sistemu ERP. Metoda kreira nov dokument »Zahteva za obravnavo ukrepa«, v katerega vgradi št. ukrepa »1000713«, šifro delavca, ki mora ukrep obravnavati in sklic na ID pravila »3456788«, hkrati pa na njegov elektronski naslov pošlje obvestilo o zahtevi po obravnavanju ukrepa.
- Martin Krpan se na podlagi prejetega obvestila prijavi v sistem ERP, kjer opazi nov dokument, to je zahtevo za obravnavanje ukrepa. Dokument se sklicuje na ukrep št. »1000713«. Iz vsebine ukrepa ugotovi, da je treba nabaviti nekaj novih računalnikov, zato pripravi nov dokument tipa »ZNAB«, to je zahtevnica za nabavo, ki ima zaporedno številko »2000432«. Vanjo vgradi sklic na zahtevo za obravnavanje ukrepa, ki vsebuje ID pravila (»3456788«). Zaradi velike nabavne vrednosti novih računalnikov pridobi dokument status »A«. Ob shranitvi zahtevnice se sproži dogodek »Zahteva\_kreirana«, ki prek omenjenega sklica na zahtevnico pridobi ID pravila (»3456788«), ki ga skupaj s tipom (»ZNAB«), statusom (»A«), številko dokumenta (»2000432«) in nazivom dogodka pošlje spletni storitvi sistema WFMS.
- WFMS zaključí pravilo »3456788«, hkrati pa v procesnem modelu poišče poslovno pravilo, ki ga dani zaključni dogodek, to je »Zahteva\_kreirana«, sproži. V sistem vnese novo pravilo, to je »Obravnavanje zahteve«, skupaj z izhodnimi parametri zaključnega dogodka. Novo pravilo vsebuje kontrolni mehanizem, ki preverja dva pogoja, tip in status zahteve. Ker gre za nabavno zahtevnico, tip »ZNAB« s statusom »A«, je treba za izvajalca naloge opredeliti »Vodjo nabave«. WFMS na podlagi razporeditev delavcev določi konkretnega izvajalca, to je »Janez Novak«, nato pa izvede klic spletne storitve, ki ji preda naslednje parametre: ID novega pravila (»3456789«), poslovni objekt (»Zahteva za nabavo«) in metodo (»Obravnavaj zahtevo«), tip (»ZNAB«), številko (»2000432«) in status dokumenta (»A«), številko izvajalca in njegov elektronski naslov. Spletna storitev izvede dejanski klic metode, ki v sistemu ERP v nabavno zahtevnico »2000432« vgradi ID pravila in številko izvajalca kot osebo, ki bo dokument obravnavala. Hkrati pošlje izvajalcu na njegov elektronski naslov obvestilo o zahtevnici za nabavo.
- Janez Novak se prijavi v sistem ERP, kjer opazi zahtevnico za nabavo, ki jo je treba obravnavati. Ker se mu vsebina dokumenta ne zdi sporna, ga odobri, s tem pa se spremeni tudi njegov status (»Odobreno«). Pri tem se sproži dogodek »Odobritev«, ki izvede klic spletne storitve sistema WFMS s parametri: naziv dogodka (»Odobritev«), ID pravila (»3456789«), tip (»ZNAB«), številko (»2000432«) in status dokumenta (»Odobreno«).
- WFMS zaključí pravilo »3456789« in skladno z zaključnim dogodkom poišče naslednjega, to je »kreiranje nabavnega naloga«. Pravilo je brezpogojno, zato WFMS preda ustrezni spletni storitvi naslednje parametre: ID pravila (»3456790«), naziv poslovnega objekta (»nabavni nalog«), naziv metode (»nov nalog«), tip (»ZNAB«), številko (»2000432«) in status



Slika 7: Krmiljenje delovnih procesov s sistemom WFMS

dokumenta (»Odobreno«). Spletna storitev z danimi parametri izvede klic metode »nabavni\_nalogov\_nalog«, ki ni interaktivna. Metoda torej avtomatično na podlagi podatkov iz odobrene zahtevnice kreira nov dokument, to je nabavni nalog, ki mu dodatno vgradi ID pravila (»3456790«). Tako nastane dokument tipa »NABN« s številko »3000972« in statusom »OK«. Ob tem se sproži dogodek »Nalog kreiran«. Ta izvede klic spletne storitve sistema WFMS, ki ji preda naslednje parametre: ID pravila (»3456790«), naziv dogodka (»nalog kreiran«), tip (»NABN«), številko (»3000972«) in status dokumenta (»OK«). WFMS na podlagi prejetih parametrov zaključi pravilo »3456790«.

Skozi WFMS je mogoče obravnavani delovni proces nadaljevati kot avtomatiziran medorganizacijski proces, in sicer tako, da WFMS posreduje kreirano nabavno naročilo ustreznemu prodajnemu referentu dobavitelja.

#### 4 Uporabna vrednost sistema za upravljanje delovnih procesov

WFMS omogoča tako modeliranje delovnih procesov kot krmiljenje njihovega izvajanja, prav tako pa nudi možnost integracije različnih programskih rešitev, da se zagotovi nemoteno izvajanje delovnih procesov. Uporabiti ga je smotno tudi za modeliranje in izvajanje medorganizacijskih procesov, kjer je nerealno pričakovati tolikšno stopnjo enotnosti, da bi vsi partnerji uporabljali popolnoma centraliziran celovit informacijski sistem oskrbovalne verige. Zato je smiselno uporabiti WFMS kot sistem, ki prek skupne upora-

be poslovnih objektov avtomatizira ali vsaj pospeši izvajanje medorganizacijskih procesov.

Uporabna vrednost sistema WFMS je torej velika, če je le arhitektura obstoječih sistemov dovolj fleksibilna. V idealnih okoliščinah predstavlja WFMS jedro vsake celovite poslovno-informacijske rešitve, v praksi pa so njegove prednosti le redko v celoti izkoriščene. Organizacija, ki vpelje tovrsten sistem, redko izkoristi vse njegove funkcionalnosti. Tako mnoge poslovne procese še vedno podpirajo stari, neprilagodljivi informacijski sistemi, ki niso povezani z WFMS.

#### 5 Viri in literatura

- [1] ASSAF, Arkin: Business Process Modeling Language – BPMI Proposed Recommendation 2003, (URL: <http://www.bmpi.org>), 7. 11. 2003.
- [2] ENDL, Rainer, KNOLMAYER, Gerhard, PHARER, Marcel: Modeling Processes and Workflows by Business Rules. Bern: University of Bern, 1998. 10 str.
- [3] HOLLINGSWORTH, David: The Workflow Reference Model. Hampshire: WFMC, 1995. 55 str. (URL: <http://www.wfmc.org>), 5. 11. 2003.
- [4] KELLER Gerhard, TEUFEL Thomas: SAP R/3 Process-Oriented Implementation. Harlow: Addison-Wesley Longman Limited, 2000. 845 str.
- [5] KOVAČIČ, Andrej: Informatizacija poslovanja. Ljubljana: Ekonomska fakulteta, 1998. 214 str.
- [6] LIPOVEC, Filip: Razvita teorija organizacije. Maribor: Založba Obzorja, 1987.
- [7] ROZMAN, Rudi, KOVAČ, Jure, KOLETNIK, Franc: Management. Ljubljana: Gospodarski vestnik, 1993. 310 str.
- [8] SAP Library – SAP Business Workflow. Walldorf: SAP AG.
- [9] STURM Jake: VB6 UML Design and Development. Wrox Press Ltd, 1999.

Mag. Miroslav Ribič je po končani Srednji šoli za računalništvo nadaljeval izobraževanje na Ekonomski fakulteti v Ljubljani, kjer se je usmeril v študij informatike. Med študijem se je ukvarjal tudi z izgradnjo informacijskih sistemov, ki podpirajo upravljanje s človeškimi viri, aktivno pa je sodeloval pri informatizaciji skladov dela. Dodiplomski študij je sklenil z delom Informacijski sistem spremljanja in usmerjanja presežnih delavcev v RS. Kot imetnik Microsoftovih licenc MCP in MCS D se aktivno ukvarja s preučevanjem internetnih tehnologij. S tega področja je tudi uspešno zagovarjal magistrsko delo Implementacija elektronskega poslovanja med podjetji. Trenutno je zaposlen v podjetju IDS Scheer, kjer sodeluje pri uvajanju rešitev podjetja SAP.

Mag. Marjan Lončarič je diplomiral leta 1981 na Visoki šoli za organizacijo dela v Kranju. Naziv magistra znanosti je obranil na Fakulteti za organizacijske vede v Kranju leta 1991 z nalogo Dinamični model spremljanja in usmerjanja razvoja kadrov s posebnim ozirom na presežke zaposlenih v lesni industriji. Na FOV je bil v obdobju od 1985 do 1990 tudi redno zaposlen kot samostojni svetovalec in predstojnik konzultantskega centra, kar ga je v nadaljevanju vodilo v ustanovitev lastnega razvojnega, raziskovalnega in svetovalnega podjetja. Med drugim je kot ekspertni svetovalec vladne projektne skupine snoval in udeleževal projekt Skladi dela v Republiki Sloveniji. Sedanje glavno področje njegovega dela je razvoj in upravljanje z organizacijskimi in informacijski sistemi, še zlasti na področju upravljanja s človeškimi viri.

Dr. Andrej Kovačič je v zadnjih desetih letih delal kot projektant, razvijalec in svetovalec pri projektih strateške prenovne in informatizacije poslovanja. Je izredni profesor s področja poslovne informatike na Ekonomski fakulteti in Fakulteti za upravo ter predstojnik Inštituta za poslovno informatiko pri EF v Ljubljani. Več let je bil predsednik programskega odbora Dnevo slovenske informatike v Portorožu, je član izvršnega odbora Slovenskega društva INFORMATIKA, odgovorni urednik revije Uporabna informatika, svetovalec in veščak s področja vodenja in upravljanja podjetij (PHARE, Zveza ekonomistov) in pooblaščen revizor informacijskih sistemov.

# ▣ Natančnost uvrščanja slovenskih besedil

Marko Hölbl, Petra Grm, Boštjan Brumen, Tatjana Welzer, Izidor Golob  
Fakulteta za elektrotehniko, računalništvo in informatiko,  
Univerza v Mariboru, Smetanova 17, 2000 Maribor  
{marko.hölbl, petra.grm, brumen, welzer, izidor.golob}@uni-mb.si

## Izvleček

V članku predstavljamo meritve natančnosti strojnega uvrščanja slovenskih besedil. Sorodne raziskave na tem področju so bile izvedene predvsem na angleško govorečem področju in na angleških besedilih, prav tako jih je bilo moč zaslediti tudi na ostalih govornih področjih, na primer na nemškem in kitajskem. Na slovenskem govornem področju pa podobnih raziskav ni bilo možno zaslediti. Namen raziskave je ugotavljanje učinkovitosti tujih orodij pri uvrščanju slovenskih besedil. V ta namen smo izbrali IBM-ovo orodje za uvrščanje besedil, imenovano "IBM Intelligent Miner for Text" [7]. Raziskava je potekala v 3 fazah. Najprej smo zbrali množico 270 slovenskih besedil iz dnevnega časopisja in jih razvrstili v 3 kategorije: politika, gospodarstvo in šport. Množico smo razdelili na učni del na testni del. Nato je sledila izgradnja uvrščevalnega modela s pomočjo učne množice. V tretji fazi smo testirali natančnost izgrajenega modela na podlagi testnega dela. Rezultati uvrščanja slovenskih besedil so se izkazali za presenetljivo dobre (natančnost nad 87 %), kljub temu, da smo uporabili orodje, ki ni namenjeno slovenskem govornem območju.

## Abstract

### Accuracy of Categorization of a Slovene Text

The article presents the topic of automatic text categorization of a Slovene text. Related articles mostly deal with the categorization of an English text; some address other languages like German or Chinese as well. A study of Slovene text categorization has not been carried out yet. We used a tool called "IBM Intelligent Miner for Text" [7] for this task. The research was conducted in three phases. In the first phase we collected 270 articles from daily newspapers and divided them into three categories: economics, politics and sports. We split the articles into a learning-group and a test-group. We created a categorization model in the second phase on the basis of the learning-group. Then in the third phase we tested the categorization model on the test-group. In spite of the fact that we used a foreign tool, which is not designed for the Slovene language environment, the accuracy of the tool was very good (average 87 %).

## 1 UVOD

**Z razvojem sodobne informacijske družbe in s tem povezanim ogromnim naborom javno dostopnega pisnega gradiva v elektronski obliki se poraja potreba po samodejnem uvrščanju besedil po kategorijah. To omogoča hitrejšo iskanje zelenih informacij in hkrati prinaša večjo preglednost nad podatki. Prav zato je raziskovanje področja samodejnega uvrščanja besedil danes zelo aktualno.**

Zato smo naredili raziskavo natančnosti uvrščanja slovenskih besedil. Ker pa slovenskega orodja ni na voljo, smo nalogo opravili s tujim. Izbrali smo IBM-ovo orodje za uvrščanje besedil "IBM Intelligent Miner for Text" [7]. Orodje nam omogoča uvrščanje besedil v kategorije s pomočjo poprejšnjega nadzorovanega učenja.

Članek je razdeljen na šest razdelkov. V razdelku 2 so navedene sorodne raziskave s tega področja, ki

pa jih tudi v angleškem prostoru ni prav veliko. Razdelek 3 obravnava opis problema in metodologijo; 4. razdelek govori o postopku izvedbe uvrščanja besedil v kategorije. V razdelku 5 so podani dobljeni rezultati in obrazložitve le-teh; zadnji, 6. razdelek je zaključek.

## 2 SORODNE RAZISKAVE

Prve javno objavljene raziskave s področja avtomatskega uvrščanja besedil segajo v leto 1990. V tem času se je namreč povečala potreba po samodejnem uvrščanju tekstov.

Največ raziskav s tega področja temelji, kot pričakovano, na angleških besedilih. Zasledili smo tudi raziskave z nemško govorečega področja, medtem ko je uvrščanje besedil v drugih jezikih manj raziskano.

Tiskovna agencija Reuters je eno prvih podjetij, ki se je začelo aktivno ukvarjati s tem problemom. Že leta 1990 je uporabljalo ekspertni sistem Construe [2] za avtomatsko uvrščanje besedil in z njim doseglo kar dobre rezultate. Učenje (izgradnja modela) je potekalo s pomočjo ročno vnesenih pravil, ki so jih pripravili domenski eksperti v podjetju. Razporejanje v kategorije je temeljilo na algoritmu k-tega najbližjega soseda (angl. k-Nearest Neighbor - kNN) [3]. Njihov sistem je na množici 750 testnih primerkov dosegel 90 % natančnost [2].

Prav tako smo zasledili članke, ki opisujejo primerjavo različnih algoritmov samodejnega uvrščanja tekstov. Članek [6] navaja različne algoritme oz. postopke, ki temeljijo na linearni algebri, teoriji verjetnosti in ostalih postopkih. Najbolj uporabljani med njimi so:

- Sistem k-tega najbližjega soseda (angl. k-Nearest Neighbor - kNN)
- Podporni vektorski stroji (angl. Support Vector Machines - SVM)
- Linearna aproksimacija s pomočjo najmanjših kvadratov (angl. Linear Least Squares Fit - LLSF)
- Nevronske mreže (angl. Neural Network - NNet)
- Metoda naivnega Bayesa (angl. Naive Bayes - NB)

Tudi ta članek obravnava samo primerjavo natančnosti posameznih algoritmov na angleških besedilih.

V članku [1] je bila izvedena raziskava na angleških in nemških besedilih. Tudi na nemških besedilih se je sistem Construe izkazal kot zelo perspektiven, čeprav je bila njegova uspešnost manjša kot pri uvrščanju angleških tekstov. To smo pričakovali, saj gre za angleško orodje. Sistem je uporabljal slovar, ki ga je zgradil v fazi učenja, hkrati s pravili, ki so jih določili uporabniki. Pri tej raziskavi je bila izvedena tudi primerjava ekspertnega sistema s samoučečim se postopkom, temelječim na odločitvenih drevesih in sistemu z Bayesovo verjetnostjo. Tudi tu se je za najučinkovitejšega izkazal Construe sistem.

Najbližje naši raziskavi je članek [4], kjer je skupina kitajskih raziskovalcev preučevala natančnost uvrščanja kitajskih besedil. Za uvrščanje so uporabljali naslednje metode:

- Sistem k-tega najbližjega soseda (angl. k - Nearest Neighbor - kNN)
- Podporni vektorski stroji (angl. Support Vector Machines - SVM)
- Adaptivna resonančna asociativna mreža (angl. Adaptive Resonance Associative Map - ARAM)

Postopki so se izkazali pri predpostavki, da je bila učna množica dovolj velika, kot relativno dobri, vendar avtorji ne navajajo konkretnega podatka o doseženi natančnosti.

### 3 OPIS PROBLEMA IN METODOLOGIJA

Samodejno uvrščanje se s pridom uporablja pri internetnih iskalnikih, za urejanje elektronske pošte in uvrščanje raznih člankov [8].

Cilj naše raziskave je bil ugotoviti natančnost uvrščanja slovenskih časopisnih člankov s pomočjo tuje orodja.

Na začetku raziskave smo bili skeptični glede natančnosti uvrščanja slovenskih besedil, saj smo uporabljali ameriško programsko orodje [7], ki ni namenjeno slovenskemu govornemu področju.

Za to orodje smo se odločili zaradi treh poglavitnih razlogov. Prvi razlog je bil ta, da slovenskega orodja za uvrščanje slovenskih besedil ni. Drugi pomemben razlog je dejstvo, da je orodje brezplačno za akademske ustanove. Tretji razlog: za že izdelano programsko orodje smo se odločili tudi zato, ker večina vsakdanjih uporabnikov ni dobro podkovana v poznavanju različnih algoritmov uvrščanja, ampak želijo le usluge programa. Zato smo nalogo uvrščanja opravili tako, da smo v orodje vnesli vhodne podatke in opazovali izhodne podatke. Princip, ki je znan kot princip črne škatle, je bil v našem vidiku zelo primeren, saj smo lahko z njim ugotovili dejansko učinkovitost oz. natančnost orodja, ki ga lahko uporablja tudi o podrobnostih nepoučeni uporabnik.

V naslednjem poglavju opisujemo izvedbo eksperimenta.

### 4 OPIS POSTOPKA

#### 4.1 Priprava

Najprej smo se lotili iskanja primernih besedil za uvrščanje. Odločili smo se, da bomo za raziskavo uporabili članke iz dnevnega časopisja. Članke smo črpali iz internetnih strani časopisnih hiš in letne zgoščenke časopisa Večer [15], saj so le-ti lažje dosegljivi. Izbrali smo internetne strani časopisov Delo [10], Večer [11], Dnevnik [12], Morel [14] in Primorski dnevnik [13]. Poiskali smo 270 člankov, ki smo jih razvrstili v tri kategorije. To so bili politika, gospodarstvo in šport. Skupna dolžina vseh člankov je znašala 1.027,3 kilobajtov, povprečna dolžina pa 3,8. Skupno je bilo v člankih 161.400, v povprečju pa 597,8 besed. V kategoriji šport

so bili v povprečju članki dolgi 474,0 besed oz. 2,9 kilobajtov, v kategoriji politika 796,3 besed oz. 5,1 kilobajtov in v kategoriji gospodarstvo 523,2 besed oz. 3,4 kilobajtov. Iz navedenega je razvidno, da so bili članki precej kratki (tabela 1).

	Povpr. dolžina (besed)	Povpr. dolžina (kb)
Šport	474,0	2,9
Politika	796,3	5,1
Gospodarstvo	523,2	3,4

Tabela 1: Povprečne dolžine člankov

	Število besed	$\sigma$
Šport	201,8	1,2
Politika	655,3	4,0
Gospodarstvo	250,7	1,6

Tabela 2: Standardni odklon

Za prvi dve kategoriji smo se odločili iz predpostavke, da je razpoznavanje tako podobnih področij, kot sta gospodarstvo in politika, še posebno zahtevno. Na drugi strani se je kategorija šport povsem razlikovala od prvih dveh omenjenih.

Članki in pripadajoče kategorije so na voljo na internetni strani <http://lpt.uni-mb.si/public/ev/raz-slo-besedil.htm>

Članke smo razvrstili v dve množici [9]:

- 2/3 člankov vsake kategorije je predstavljalo učno množico, na kateri se je orodje učilo,
- 1/3 člankov je predstavljalo testno množico člankov, ki smo jih razvrstili in z njimi preverili natančnost naučenega modela.

Razporejanje posameznih člankov v ti dve množici je bilo ključno.

Zaradi zahtev IBM-ovega orodja smo vse članke spremenili v tekstovne datoteke. Pri tem je bilo treba ustvariti še datoteko, ki je vsebovala seznam vseh člankov za posamezno kategorijo učne in testne množice.

S tem smo zaključili pripravo vhodnih podatkov.

#### 4.2 Izvedba meritve natančnosti

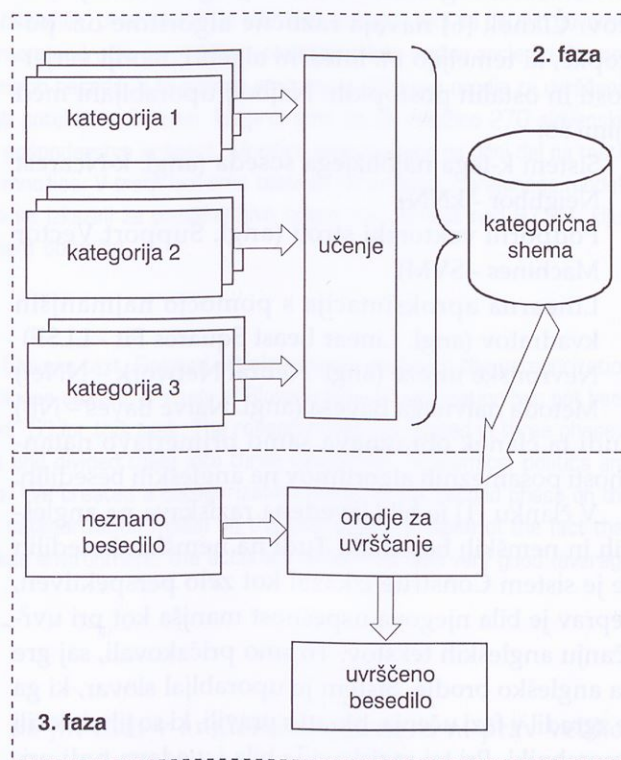
Meritve natančnosti uvrščanja besedil so bile izvedene v dveh korakih:

1. učenje na učni množici (izgradnja modela),
2. testiranje oz. uvrščanje besedil na testni množici.

S kazalnimi datotekami učne množice smo izvedli učenje. Po opravljenem učenju na vseh treh učnih množicah – za vsako kategorijo ena učna množica – smo dobili kategorično shemo. Ta shema je dejansko slovar, ki hrani pomembne besedne statistike za vsako kategorijo. Te statistike so bile potem uporabljene za uvrščanje testnih primerov.

Podrobnosti principov delovanja algoritma, na katerem temelji učenje oz. izgradnja sheme, zaradi komercialne narave orodja ni poznana.

Omenjeni postopek prikazuje naslednja shema (slika 1):



Slika 1: Prikaz postopka učenja in uvrščanja

Po opravljenem učenju je sledila druga faza – testiranje oz. uvrščanje testne množice besedil.

Podana je bila kazalna datoteka za testne primerke posamezne kategorije. Na podlagi kategorične sheme je orodje uvrstilo vse testne datoteke. Rezultat tega procesa je bila izhodna datoteka; primer takšne datoteke je prikazan na sliki 2. Vsebuje posamezni testni članek, za katerega so navedene točke ujemanja z določeno kategorijo. Članek pripada tisti kategoriji, pri kateri je dosegel največ točk.

Document Identifier: g02-d.TXT	
Category	Score
P1-trn	153.652
G1-trn	140.937
S1-trn	108.618

Slika 2: Primer izhodne datoteke

## 5 REZULTATI

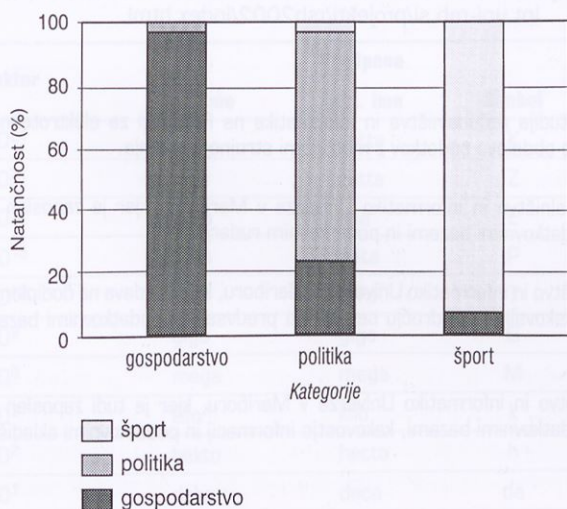
Iz tabele 3 je razvidno, da je testna množica za vsako kategorijo vsebovala 30 člankov. Iz kategorije gospodarstvo je orodje pravilno razvrstilo 29 člankov, enega pa je napačno razvrstilo v razred politika.

Uvrščen kot \ Dejanski	Dejanski		
	Gospodarstvo	Politika	Šport
Gospodarstvo	29	7	2
Politika	1	22	0
Šport	0	1	28

Tabela 3: Razvrstitev člankov v kategorije

Pri politiki je bila natančnost uvrščanja nižja, saj je bilo le 22 od 30 člankov pravilno uvrščenih. V drugi dve kategoriji, gospodarstvo in šport, je bilo uvrščenih skupno osem člankov s področja politike, in sicer sedem v kategorijo gospodarstvo ter eden v kategorijo šport.

Članki s športno tematiko so bili uvrščeni bolj kot politični članki. Od 30 testnih primerkov je bilo pravilno umeščenih 28, dva pa sta bila uvrščena kot gospodarstvo.



Slika 3: Učinkovitost uvrščanja po temah

Največje težave pri uvrščanju so se pojavile pri politiki, saj se politika meša z gospodarstvom, kot na primer v članku številka g47-d.txt [16].

Problem izvira iz dejstva, da je politika tematika, ki je v dnevnem časopisju nastopala na različnih področjih, med drugim tudi pri gospodarstvu. Velikokrat sta se gospodarstvo in politika prepletala. Kljub težavnemu ločevanju med tema dvema kategorijama se je orodje izkazalo kot uspešno – 73,3 %.

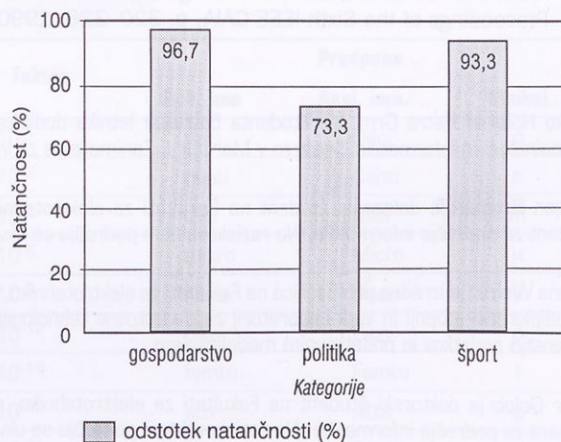
Osnova razpoznavanja je bila kategorična shema, ki temelji na že prej omenjenih besednih statistikah za vsako posamezno kategorijo. Problem se je pojavil, ker so lahko članki različnih kategorij vsebovali enake besede oz. besedne zveze. Tem bolj je bila tematika kategorij različna, tem manjša je bila verjetnost, da nastopajo v člankih različnih kategorij iste besede.

V nadaljevanju podajamo tabelo (tabela 4) in graf natančnosti (slika 4) v odstotkih za posamezno kategorijo.

Kategorija	Odstotek natančnosti [%]
Gospodarstvo	96,7
Politika	73,3
Šport	93,3
Skupaj	87,8

Tabela 4: Natančnost uvrščanja

Kot je razvidno iz grafa (slika 4), je natančnost uvrščanja dobra, v povprečju 87,8 %. Tako je bilo orodje uspešno v vseh kategorijah, ki so služile kot testne in učne množice za uvrščanje.



Slika 4: Natančnosti uvrščanja

## 6 SKLEP

Namen raziskave je bil ugotoviti natančnost uvrščanja slovenskih besedil. Takšna raziskava še ni bila izvedena na slovenskem govornem področju. Raziskovali smo s pomočjo ameriškega orodja, saj v Sloveniji trenutno ni na voljo ustrezne programske opreme. Kljub temu, da smo razpoznavanje izvedli s pomočjo tuje orodja, ki ne podpira slovenskega jezika, smo prišli do spoznanja, da lahko natančno uvrščamo slovenska besedila s pomočjo omenjenega orodja. Dano orodje uspešno ločuje besedila, ki so si po tematiki podobna. Če bi želeli uvrščati tekste, ki bi se po tematiki zelo razlikovali, bi dobili uspešnost nad 90 %, ki je primerljiva z natančnostjo uvrščanja angleških besedil [1], [2]. Že sedanja natančnost uvrščanja je boljša kot natančnost, ki jo je dobil Reuters [2].

Vendar omenjeno ne izključuje možnosti razvoja sorodnega slovenskega orodja.

V naši raziskavi smo ugotovili, da je v primeru zahteve po večji natančnosti, potrebno razviti posebno orodje.

V nadaljnjih raziskavah bomo proučili uspešnost orodja z večjim številom kategorij. Prav tako bomo za boljšo oceno natančnosti uvrščanja uporabili metodo navzkrižne validacije.

## 7 LITERATURA

- [1] C. Apte, F. Damerau, S. M. Weiss, Towards Language Independent Automated Learning of Text Categorization Models, In Proceedings of the ACM SIGIR Conference, 1994.
- [2] P. J. Hayes, P. M. Andersen, I. B. Nirenburg, and L. M. Schmandt. TCS: A Shell for Content-Based Text Categorization. In Proceedings of the Sixth IEEE CAIA, p. 320–326, 1990.
- [3] T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features, In Proceedings of the European Conference on Machine Learning, Springer, 1998.
- [4] Ji He, Ah-Hwee Tan, Chew-lim Tan, A Comparative Study on Chinese Text Categorization Methods, PRICAI Workshop on Text and Web Mining, 2000.
- [5] Ah-Hwee Tan, Fon-Lin Lai, Text Categorization, Supervised Learning, and Domain Knowledge Integration, In proceedings, KDD'2000 International Workshop on Text Mining, Boston, pp. 113–114, August 2000.
- [6] Yiming Yang, Xin Liu, A re-examination of text categorization methods, Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99, p. 42–49), 1999.
- [7] IBM Intelligent Miner for Text Version 2 Release 3 Text Analysis Tools, Documentation on CD, Third edition, December 1998.
- [8] IBM Intelligent Miner for Text Version 2 Release 3, <http://www-3.ibm.com/software/data/iminer/fortext/>, zadnji obisk 15. 3. 2003.
- [9] P. R. Cohen, Empirical Methods for Artificial Intelligence, The MIT Press, Cambridge, USA, 1995.
- [10] Delo, 2002, <http://www.delo.si>, zadnji obisk 15. 3. 2003.
- [11] Večer, 2002, <http://www.vecer.com>, zadnji obisk 15. 3. 2003.
- [12] Dnevnik, 2002, <http://www.dnevnik.si>, zadnji obisk 15. 3. 2003.
- [13] Primorski dnevnik, 2002, <http://www.primorski.it>, zadnji obisk 15. 3. 2003.
- [14] Elektronski časopis Morel, 2002, <http://www.morel.si/>, zadnji obisk 15. 3. 2003.
- [15] CD Večer, ČZP Večer, 2001.
- [16] Naslov strani s testno in učno množico člankov, <http://lpt.uni-mb.si/projekti/rsb2002/index.html>.

Marko Hölbl in Petra Grm sta študenta četrtega letnika dodiplomskega študija računalništva in informatike na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru. Zanimata se za inteligentno obdelavo podatkov z metodami strojnega učenja.

Boštjan Brumen je doktorski študent na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru, kjer je zaposlen kot asistent za področje informatike. Na raziskovalnem področju se ukvarja s podatkovnimi bazami in podatkovnim rudarjenjem.

Tatjana Welzer je izredna profesorica na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru, kjer predava na dodiplomski in podiplomski stopnji in vodi Laboratorij za podatkovne tehnologije. Na raziskovalnem področju se ukvarja predvsem s podatkovnimi bazami, kakovostjo podatkov in podatkovnim modeliranjem.

Izidor Golob je doktorski študent na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru, kjer je tudi zaposlen kot asistent za področje informatike. Na raziskovalnem področju se ukvarja s podatkovnimi bazami, kakovostjo informacij in podatkovnimi skladišči.



# Uporaba desetiških SI predpon in predpon v informatiki

Ladislav Mikola

Fakulteta za elektrotehniko, računalništvo in informatiko  
Univerza v Mariboru, Smetanova 17, 2000 Maribor  
ladislav.mikola@uni-mb.si

## Povzetek

Kljub temu, da je začetek uporabe desetiških predpon k merskim enotam mogoče zaslediti že pred več kot dvesto leti, je v tehniškem komuniciranju, pa tudi tehniških aktih, slovarjih in celo pravopisu mogoče opaziti nedoslednosti in celo napake. V tem članku želimo navesti nekatere najpogostejše, prav tako pa opozoriti na uporabo predpon v informatiki. To področje je že nekaj let urejeno z ustreznim standardom.

## Abstract

### The Use of Decimal SI Prefixes and Prefixes in Information Science

Although the practice of adding decimal prefixes to units was introduced some 200 years ago, we can still find errors and inconsistent usage in engineering communication, technical documents, dictionaries and even in orthography. In this paper we present the most widespread examples of incorrect usage and highlight the proper use of prefixes in information science. This field is regulated by a standard that has come into force a few years ago.

## Uporaba desetiških SI predpon

Pri podajanju numeričnih vrednosti fizikalnih veličin, npr. rezultatov meritev, bi pogosto morali zapisati zelo majhne ali zelo velike številke, če bi želeli dodati enoto te veličine (npr. V, A, W, F idr.). V takšnih primerih je praktično in skladno s priporočilom standarda SIST ISO 1000 (3) (slovenski standard, prevzet od International Organization for Standardization), numerično vrednost zapisati med 0,1 in 1000 in ji dodati

desetiško mersko enoto. Ta je sestavljena iz desetiške predpone in imena enote. Predpona predstavlja določen faktor, npr. mega, oznaka M, predstavlja faktor 10<sup>6</sup>. Tako se število številok, to je dolžina numeričnega zapisa, skrajša in ta postane bistveno preglednejši.

Predpone mednarodnega sistema merskih enot SI (angl. International System of Units), so mednarodno dogovorjene, navedene pa so v različnih aktih. V

Faktor	Predpona		
	Slov. ime	Angl. ime	Simbol
10 <sup>24</sup>	jota	yotta	Y
10 <sup>21</sup>	zeta	zetta	Z
10 <sup>18</sup>	eksa	exa	E
10 <sup>15</sup>	peta	peta	P
10 <sup>12</sup>	tera	tera	T
10 <sup>9</sup>	giga	giga	G
10 <sup>6</sup>	mega	mega	M
10 <sup>3</sup>	kilo	kilo	k
10 <sup>2</sup>	hekto	hecto	h
10 <sup>1</sup>	deka	deca	da

Faktor	Predpona		
	Slov. ime	Angl. ime	Simbol
10 <sup>-1</sup>	deci	Deci	d
10 <sup>-2</sup>	centi	Centi	c
10 <sup>-3</sup>	mili	Milli	m
10 <sup>-6</sup>	mikro	Micro	μ
10 <sup>-9</sup>	nano	Nano	n
10 <sup>-12</sup>	piko	Pico	p
10 <sup>-15</sup>	femto	Femto	f
10 <sup>-18</sup>	ato	Atto	a
10 <sup>-21</sup>	zepto	Zepto	z
10 <sup>-24</sup>	jokto	Yocto	y

Tabela 1: Desetiške SI predpone

Republiki Sloveniji jih predpisuje odredba o merskih enotah (4). Navajamo jih v tabeli 1, kjer smo slovenskemu dodali tudi angleško ime. Simboli so enaki ne glede na državo ali področje, kjer se uporabljajo.

### Najpomembnejša pravila za uporabo desetiških SI predpon

- Zapis predpon in enot mora biti pokončen, medtem ko je simbol za veličino pisan poševno. Tako npr. za induktivnost vrednosti 3,2 henrija zapišemo  $L = 3,2 \text{ H}$ . Pogosto je enota, pa tudi desetiška merska enota, napačno poševno zapisana.
- Med številčnim zapisom rezultata in enoto je presledak.
- Med decimalno predpono in simbolom za enoto ni presledka.
- Uporabiti smemo le eno predpono. Ne npr. mmF, ampak mF.
- Ne smemo zamenjevati pojmov faktorja (npr. 1012) in desetiške predpone (npr. piko, p).
- Paziti je treba, da pravilno uporabljamo velike in male črke in da ne pride do zamenjav, ki bi pomenile bistveno drugačno vrednost (npr. zamenjava m in M pomeni razliko za faktor 109). Manj problematično je, vendar pa nedopustno, zapisati Kg namesto kg. Ta napaka ni redka.
- Včasih, sicer ne več tako pogosto, zasledimo oznako dkg za dekagram. Pomensko je to zapis z dvema predponama. Pravilno je seveda dag.

Iz SIST ISO 1000 povzemamo nekaj primerov neustreznega in ustreznega zapisa.

Neustrezní zapis	Ustrezní zapis
$1,2 \times 10^4 \text{ N}$	12 kN
0,00394 m	3,94 mm
1401 Pa	1,401 kPa
$3,1 \times 10^{-9} \text{ s}$	3,1 ns

Pri praktični uporabi desetiških predpon je treba upoštevati nekaj pravil, ki izhajajo iz tradicije: npr. pri enoti °C predpon nismo uporabljali, čeprav se že povsem legalno pojavlja desetiška enota m°C, pri enoti K je uporaba enote mK povsem primerna, neobičajna pa kK.

Podobno je npr. pri sekundi, kjer uporabljamo ms, ne pa ks. Vendar so v tem primeru na razpolago večje enote npr. minuta (min), ura (h), ki sicer nista enoti SI. Lahko pa se seveda uporabljata brez omejitev.

### Uporaba predpon v informatiki

V informatiki sta se ustalili informacijski enoti bit (simbol bit, iz angleških besed binary digit) in bajt (simbol B, iz angleške besede byte). Z rastjo zmogljivosti računalniških sistemov se je pojavila potreba, da se informacijska vsebina izrazi z večjimi enotami, kot kakor sta bit in bajt. Glede na dolgoletno uporabo desetiških SI predpon so začeli uporabljati te, kar pa pripelje do nedoslednosti ali dvopomenskosti. Najbolj očiten je primer uporabe predpone kilo, oznaka k in v informatiki včasih tudi K, ki pomeni faktor 1000 (npr. kilometer, km,  $1 \text{ km} = 1000 \text{ m}$ ), vendar pa tudi 1024 (npr. kilobajt, KB,  $1 \text{ KB} = 1024 \text{ B}$ ). Tak pristop se je razširil tudi na predponi mega (M) in giga (G). S tem je negirano temeljno izhodišče standardiziranja na področju merskih enot, po katerem naj ima vsaka enota le eno vrednost, eno ime in en znak.

Tehnični komite TC 25 (angl. Technical Committee), pri mednarodni elektrotehniški komisiji IEC (angl. International Electrotechnical Commission), je z močno podporo mednarodnega odbora za uteži in mere CIPM (fr. Comité International des Poids et Mesures) in ameriške inženirske strokovne organizacije IEEE (angl. Institute of Electrical and Electronics Engineers), dopolnil standard IEC 60027 (6) in v njegovem drugem delu, poglavju 13: Data processing and data transmission, tabelarično navedel oz. določil veličine, imena, simbole in enote v informatiki. Nekatere od njih navajamo v tabeli 2.

Ime veličine	Simbol veličine	Ime enote	Simbol enote
(traffic intensity) intenzivnost prometa	A	erlang	E
(storage capacity) pomnilniška zmogljivost	M	bit, bajt	bit, B
(transfer rate) hitrost prenosa (8)	v	bajt na sekundo	B/s
(bit rate) bitna hitrost (9)	$v_b$	bit na sekundo	bit/s

Tabela 2: Nekatere veličine in enote v informatiki

V poglavju 14 istega standarda [6](6) je navedena tabela s predponami, uporabljanimi v informatiki. Pri njihovi tvorbi se je komisija zgledovala po desetiških SI predponah in je izhajala iz tega, da je 210 zelo bližje 1000. Lestvico predpon je začela z besedo kilobinary, kar pomeni (210)1 in jo poimenovala kibi ter

Faktor	Ime	Simbol	Pomen	Po zgledu
$2^{10}$	kibi	Ki	kilobinary: $(2^{10})^1$	kilo: $(10^3)^1$
$2^{20}$	mebi	Mi	megabinary: $(2^{10})^2$	mega: $(10^3)^2$
$2^{30}$	gibi	Gi	gigabinary: $(2^{10})^3$	giga: $(10^3)^3$
$2^{40}$	tebi	Ti	terabinary: $(2^{10})^4$	tera: $(10^3)^4$
$2^{50}$	pebi	Pi	petabinary: $(2^{10})^5$	peta: $(10^3)^5$
$2^{60}$	exbi	Ei	exabinary: $(2^{10})^6$	exa: $(10^3)^6$

Tabela 3: Predpone za binarne mnogokratnike

označila s simbolom Ki. Prvi zlog pomensko kaže na zvezo z SI predponami, drugi pa, da gre za binarno tvorbo. Standardizirane predpone, teh je šest, navajamo v tabeli 3, take kot so v originalu standarda [6](6). Poudarjamo, da te predpone niso v SI.

V tabeli 4 navajamo nekaj primerov predpon za binarne mnogokratnike in primerjavo z doslej uporabljanimi desetiški SI predponami.

kibibit	1 Kibit = $2^{10}$ bit = 1024 bit
kilobit	1 kbit = $10^3$ bit = 1000 bit
mebibajt	1 MiB = $2^{20}$ B = 1 048 576 B
megabajt	1 MB = $10^6$ B = 1 000 000 B
gibibajt	1 GiB = $2^{30}$ B = 1 073 741 824 B
gigabajt	1 GB = $10^9$ B = 1 000 000 000 B

Tabela 4: Primeri predpon za binarne mnogokratnike in primerjava s SI predponami

### Obveznost uporabe predpon

Standard IEC 60027 še ni razglašen za slovenski standard (SIST), vendar pričakujemo, da se bo to zgodilo v bližnji bodočnosti. Glede na načelo sodobne standardizacije, da upoštevanje določil standardov ni obvezno, dokler z določenim postopkom ali aktom to ne postane, v Republiki Sloveniji predpon standarda IEC 60027 še ni obvezno uporabljati. Gotovo pa se bo to uredilo ob spremembi odredbe o merskih enotah ali pa s katerim drugim meroslovnim predpisom. Primereno pa bi bilo, da se standard začne uporabljati čimprej povsod, predvsem v šolstvu.

### Sklep

V članku smo želeli opozoriti predvsem na stanje in uporabo predpon. Nismo se spuščali v jezikovne dileme, ki so na področju poimenovanja merskih enot in predpon še vedno nerešene, kar je vidno tudi iz primerov. Gre namreč za jezikoslovno opredelitev (7), da imena enot in predpon zapisujemo tako, kot jih izgovarjamo. Kljub temu smo opazili, da v praksi, predvsem med tehnično inteligenco, vladajo tudi drugačna mnenja.

### Literatura

- [1] The International System of Units (SI), Bureau International des Poids et Mesures, 7th edition 1998, Sevres.
- [2] Brezinščak, M.: Tvorba decimalnih i binarnih mjernih jedinica, Mjeriteljski vjesnik, 18, št. 1–2, september 2000.
- [3] SIST ISO 1000: SI units and recommendation for the use of their multiples and of certain other units, third edition 1992.
- [4] Odredba o merskih enotah, Uradni list RS, št. 26, april 2001.
- [5] Thor, A. J.: Prefixes for binary multiples, Metrologia 37, št. 1, 2000.
- [6] IEC 60027- 2: Letter symbols to be used in electrical technology, Part2: Telecommunications and electronics, 2000.
- [7.] Slovenski pravopis, SAZU in ZRC SAZU Ljubljana, 2001.
- [8] Računalniški slovarček, 3. razširjena izdaja, Cankarjeva založba, Ljubljana, 1993.
- [9] Pahor, D., Drobnič, M.: Leksikon računalništva in informatike, Pasadena, Ljubljana, 2002.

Mag. Ladislav Mikola je višji predavatelj na Fakulteti za elektrotehniko, računalništvo in informatiko v Mariboru, kjer na visokošolskem strokovnem programu predava na predmetnem področju električnih meritev ter tehničnih in pravnih normativov. Njegovo raziskovalno delo je usmerjeno predvsem v meroslovje.

## Digitaliziranje madžarske kulturne dediščine<sup>1</sup>

Zsuzsana Tószegi  
tzs.@neumann-haz.hu

**Omrežne podatkovne storitve so na razpolago od konca osemdesetih, prve madžarske digitalne publikacije pa so se pojavile v prvi polovici devetdesetih let prejšnjega stoletja. Od leta 1994 so v Madžarski elektronski knjižnici dostopna prek interneta leposlovna in znanstvena dela. Poleg Digitalne knjižnice John von Neumann (John von Neumann Digital Library), ki je bila ustanovljena leta 1997, opravlja omrežne storitve in spodbuja proces digitaliziranja madžarske narodne kulturne dediščine še nekaj sto knjižnic, muzejev in arhivov.**

Omrežna podatkovna storitev Digitalne knjižnice John von Neumann WebKat.hu je odličen vir za študij in analizo elektronskih dokumentov na madžarskih strežnikih, saj je bilo npr. februarja 2003 na razpolago v WebKat.hu 171.754 dokumentov. Z uporabo OPAC je na madžarskih strežnikih mogoč dostop do vseh monografij, znanstvenih del, člankov, slik idr. dokumentov. Pričujoči prispevek opisuje možnosti informatiziranja kulturne dediščine, področja, kjer nas čaka še precej dela.

Na Madžarskem je bil internet v uporabi že v času železne zavese, ko so s strogimi predpisi omejevali uvoz modernih orodij informacijske tehnologije (IT). Strokovnjaki na področju visokega šolstva in v knjižnicah pa so vendarle uporabljali računalnik kot pomembno delovno sredstvo; ne le, da so znali pristopati do tujih podatkovnih baz, začeli so razvijati tudi svoje. Prvo knjižnično podatkovno bazo so leta 1975 vzpostavili knjižničarji univerzitetne knjižnice v Szegedu. Od leta 1976 so računalnike uporabljali v Madžarski nacionalni knjižnici Széchényi za obdelovanje in publiciranje madžarske nacionalne bibliografije. V zadnjih petindvajsetih letih je prevedlo kartične kataloge v računalniške podatkovne baze prek sto knjižnic. V začetku leta 2003 je vseboval madžarski nacionalni kataloški sistem podatke petnajst največjih knjižnic s prek 2,5 milijona<sup>2</sup> zapisi. Madžarski seznam (HUNOPAC) obsega skupaj 112 organizacij.

Nov izziv uporabe IT v knjižnicah je bil digitaliziranje kulturne dediščine. Prve madžarske zgoščenke iz leta 1990 so vsebovale podatkovne baze, prva zgoščenka s popolnim besedilom je bila biblija, prve multimedijska zgoščenka pa je bila izdana leta 1990. Med letoma 1990 in 2000 je bilo na Madžarskem izdanih okoli 2.000 zgoščenk, od katerih se jih 400 uvršča v kategorijo ohranjanja kulturne dediščine. V zgodnjih devetdesetih letih so se pojavila tudi prva digitalna gradiva v madžarsščini. Danes je na razpolago storitev »Naši znanstveniki – virtualna jubilejna razstava«, poleg te pa še madžarski knjižnični portal.

S pobudo za ustanovitev Madžarske elektronske knjižnice (MEK) iz leta 1994 so želeli zbrati vse pomembne intelektualne dosežke v digitalni obliki in javnosti omogočiti brezplačen pristop do njih. Pobudo je po petih letih uresničila madžarska narodna knjižnica. Digitalna knjižnica John von Neumann je bila ustanovljena leta 1997. Glavna področja njenega delovanja so digitaliziranje madžarske kulturne dediščine, koordiniranje projektov digitaliziranja, katalogiziranje madžarskih internetnih virov in razvijanje podatkovne baze madžarskih knjižnic, muzejev in arhivov, pa tudi organiziranje podiplomskega študija in sodelovanje pri razvoju standardov. Danes je v knjižnici okoli 50.000 literarnih del.

Edinstvena ustanova te vrste je *Digitalna literarna akademija*, ki zbira opus najpomembnejših madžarskih pisateljev, dobitnikov *Kossuthove nagrade* in *lvorjevega venca*. Opus obsega vsa v madžarsščini objavljena dela. Število članov Digitalne literarne akademije stalno raste in poleg časti prinaša članstvo v akademiji tudi denarne koristi, s čimer država pomaga kulturnikom: samo v letu 2003 z več kot 150 milijoni forintov. Trenutno obsega zbirka 24.290 del 61 avtorjev. Storitve so od leta 2000 uporabili že trije milijoni virtualnih bralcev, ki so npr. leta 2002 odpri več kot 440.000 del in prepisali več kot 97.000 dokumentov. Upošteva vso zbirko je bilo leta 2002 dnevno povprečje 1.205 obiskov in 226 prepisov dokumentov. Danes je v okviru programa *Bibliotheca Hungarica Internetiana* na internetu dostopen celotni opus vidnih madžarskih avtorjev od stare madžarske literature do največjih klasikov zgodnjega 20. stoletja. Digitalne knjižnice omogočajo hiter odziv

na aktualne dogodke; v počastitev Nobelove nagrade za literaturo leta 2003 Imreju Kertészju je izšla zgoščenka z zbirko dvanajstih Kertészevih del prav na dan slovesnosti ob podelitvi nagrad.

Ena najpomembnejših storitev Neumannove knjižnice je obveščanje o madžarskih elektronskih dokumentih v omrežnem katalogu *WebKat.hu*. Dokumenti so obdelani na analitični ravni, kar pomeni, da je vsak intelektualni produkt posebej, npr. esej, pesem, članek, slika, poseben zapis v podatkovni bazi skupaj s svojim URL. Model za vzpostavitev tega kataloga so vodilni svetovni internetni projekti katalogiziranja. Pravila katalogiziranja digitalnih knjižnic upoštevajo mednarodne trende. WebKat.hu je izdelan skladno s standardoma ISBD(ER) in MARC. Opisi so opremljeni z zaglavjem naslova za iskanje po vsebini in za prepis zapisov v formatu MARC. Februarja 2003 je obsegal katalog WebKat.hu 171.754 zapisov.

Za slovenske bralce je zanimiv podatek, koliko dokumentov je razpoložljivih o slovenski kulturi in zgodovini; 28. februarja 2003 je bilo v madžarsščini 200 takih dokumentov. O Sloveniji (149 zadetkov), Slovencih (39 zadetkov) in s tem povezanimi stranmi (201 zadetek) je precej zanimivih gradiv, med njimi npr. dokument o dvestoti obletnici rojstva Franceta Prešerna.

Ponudniki vsebin objavljajo na internetu vse več javnih zbirk, od enostavnih domačih strani do več tisoč zapisov obsegajočih katalogov. Neumannova knjižnica zbira podatke knjižnic, muzejev, arhivov in galerij. Stalno ažurna javna podatkovna baza je v začetku leta 2003 obsegala podrobne podatke o 1.850 zbirkah, od katerih jih je 1.500 dostopnih prek interneta. Med institucijami je 817 knjižnic, 89 arhivov in 921 muzejev. Del podatkovne baze WebKat.hu je baza metapodatkov o madžarskih javnih zbirkah, ki je februarja 2003 obsegala 1.850 zapisov o različnih ustanovah. To so knjižnice, muzeji, arhivi in komunalni centri. V letu 2002 so podatkovno bazo razširili z vsemi dosegljivimi podatki o projektih digitaliziranja javnih zbirk.

Na koncu naj omenimo še sodelovanje med dvema organizacijama, ki nosita ime Johna von Neumanna: Digitalna knjižnica John von Neumann in Društvo informatikov John von Neumann. Obe si prizadevata na Madžarskem razvijati družbo znanja, ki bo dostopna najširši javnosti. V zadnjih nekaj letih je bil najpomembnejši cilj Društva John von Neumann razširjati računalniško pismenost. Društvo informatikov je aktivno pri uvajanju sistema izobraževanja mednarodno priznanega spričevala ECDL. Rezultat njunega sodelovanja je več kot 50.000 udeležencev, ki so se usposobili v 260 testnih centrih in prek 100.000 kandidatov, ki opravljajo izpite.

Priložnost za afirmacijo madžarske digitalne kulturne dediščine je bila tudi stoletnica rojstva Johna von Neumanna, enega največjih matematikov dvajsetega stoletja. Več podatkov o njegovi intelektualni dediščini je mogoče najti na lokaciji [www.neumann-centenarium](http://www.neumann-centenarium).

Podrobnejši podatki in informacije so na voljo na naslednjih naslovih:

- [1] [www.neumann-centenarium.hu](http://www.neumann-centenarium.hu)
- [2] [www.webkat.hu](http://www.webkat.hu)
- [3] [www.irodalmiakademia.hu](http://www.irodalmiakademia.hu)
- [4] [www.neumann-haz.hu](http://www.neumann-haz.hu)

Dr. Zsuzsana Tószegi je podpredsednica madžarskega Društva John von Neumann in direktorica Digitalne knjižnice John von Neumann.

<sup>1</sup> Referat avtorice Z. Tószegi na posvetovanju Dnevi slovenske informatike 2003 je z njenim privoljenjem prevedel in priredil Niko Schlamberger.

<sup>2</sup> URL angleške lokacije je [http://www.mokka.hu/e\\_index.html](http://www.mokka.hu/e_index.html).

IEEE International Conference on Engineering of Complex Computer Systems	14.-16. apr. 2004	Firence, Italija	INFOS, DSIT-DSI	<a href="http://www.dsi.unifi.it/iceccs04">http://www.dsi.unifi.it/iceccs04</a>
ACM International Conference on Computing Frontiers (CF '04)	14.-16. apr. 2004	Ischia, Italija	ACM - SIGMICRO	<a href="http://www.computingfrontiers.org/beaty@emess.mscd.edu">http://www.computingfrontiers.org beaty@emess.mscd.edu</a>
IPSI - Internet, Processing, Systems and Interdisciplinary	9. maj 2004	Frankfurt, Nemčija	IPSI	<a href="http://www.internetconferences.net">http://www.internetconferences.net</a>
5 <sup>th</sup> Working Conference on Knowledge Management in Electronic Government (KMGov2004)	17.-19. maj 2004	Krems, Avstrija	Krems Danube University Krems, IFIP WG 8.3 & WG 8.5	<a href="http://falcon.ifs.uni-linz.ac.at/kmgov2004/">http://falcon.ifs.uni-linz.ac.at/kmgov2004/</a> <a href="http://falcon.ifs.uni-linz.ac.at/wimmer@ifs.uni-linz.ac.at">http://falcon.ifs.uni-linz.ac.at wimmer@ifs.uni-linz.ac.at</a>
ACM SIGPLAN Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES'04)	11.-13. jun. 2004	Washington, ZDA	LCTES	<a href="http://www.acm.org/sigplan/lctes.htm">http://www.acm.org/sigplan/lctes.htm</a> <a href="http://lctes04.flux.utah.edu">http://lctes04.flux.utah.edu</a>
Deveta konferenca OTS'2004 - Objektna tehnologija v Sloveniji	16.-17. jun. 2004	Maribor, Slovenija	COT - Center za objektno tehnologijo, FERi Maribor	<a href="http://cot.uni-mb.si/ots2004/">http://cot.uni-mb.si/ots2004/</a>
31 <sup>st</sup> Annual International Symposium on Computer Architecture - ISCA-2004	19.-23. jun. 2004	Muenchen, Nemčija		<a href="http://isca.in.tum.de/">http://isca.in.tum.de/</a>
ICS-2004 - 18 <sup>th</sup> International Conference on Supercomputing	26. jun.-1. jul. 2004	St. Malo, Francija	ICS	<a href="http://ics04.irisa.fr">http://ics04.irisa.fr</a>
16 <sup>th</sup> Euromicro Conference on Real-time Systems (ECRTS 04)	30. jun.-2. jul. 2004	Catania, Italija	Euromicro Technical Committee on Real-time Systems	<a href="http://www.dit.unict.it/ecrts2004">http://www.dit.unict.it/ecrts2004</a>
8 <sup>th</sup> World Multi-Conference on SYSTEMICS, CYBERNETICS And INFORMATICS - SCI 2004	18.-21. jul. 2004	Orlando, Florida, ZDA	SCI	<a href="http://www.iis.org/sci2004/">http://www.iis.org/sci2004/</a>
10 <sup>th</sup> International Conference on Information Systems Analysis and Synthesis - ISAS 2004	21.-25. jul. 2004	Orlando, Florida, ZDA		<a href="http://www.infocybnetics.org/citsa2004">http://www.infocybnetics.org/citsa2004</a> WebSite/Default.asp
International Conference on Cybernetics and Information Technologies, Systems and Applications - CITSA 2004	21.-25. jul. 2004	Orlando, Florida, ZDA		<a href="http://www.infocybnetics.org/citsa2004/">http://www.infocybnetics.org/citsa2004/</a> WebSite/Default.asp
2 <sup>nd</sup> International Conference on Politics and Information Systems: Technologies and Applications - PISTA'04	21.-25. jul. 2004	Orlando, Florida, ZDA	PISTA	<a href="http://www.confinf.org/Pista04">http://www.confinf.org/Pista04</a>
IFIP World Computer Congress (WCC 2004)	22.-27. avg. 2004	Toulouse, Francija	IFIP	<a href="http://www.wcc2004.org">http://www.wcc2004.org</a> dervillers@wcc2004.org
2 <sup>nd</sup> IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis	8.-10. sep. 2004	Stockholm, Švedska	IFIP	<a href="http://www.ida.liu.se/codes">http://www.ida.liu.se/codes</a>
IPSI - Internet, Processing, Systems and Interdisciplinary e-Challenges e-2004 Conference and Exhibition	25. sep. 2004	Stockholm, Švedska	IPSI	<a href="http://www.internetconferences.net">http://www.internetconferences.net</a>
IPSI - Internet, Processing, Systems and Interdisciplinary	27.-29. okt. 2004	Dunaj, Avstrija	CEPIS	<a href="http://www.irmg.com/www.eChallenges.org">www.irmg.com/www.eChallenges.org</a>
IPSI - Internet, Processing, Systems and Interdisciplinary	31. okt. 2004	Milano, Italija	IPSI	<a href="http://www.internetconferences.net">http://www.internetconferences.net</a>
WCCE 2005-World Conference on Computers in Education	4.-7. jul. 2005	Južna Afrika	IFIP	<a href="http://www.wcce2005.org.za">http://www.wcce2005.org.za</a>

# Pristopna izjava

Želim postati član Slovenskega društva INFORMATIKA

Prosim, da mi pošljete položnico za plačilo članarine SIT 6.700 (kot študentu SIT 2.900) in me sproti obveščate o aktivnostih v društvu.

(ime in priimek, s tiskanimi črkami)

(poklic)

(domači naslov in telefon)

(službeni naslov in telefon)

(elektronska pošta)

Datum:

Podpis:

Članarina SIT 6.700,- (plačljiva v dveh obrokih) vključuje tudi naročnino za revijo Uporabna informatika. Študenti imajo posebno ugodnost: plačujejo članarino SIT 2.900,- in za to prejemajo tudi revijo. Izpolnjeno naročilnico ali pristopno izjavo pošljite na naslov:

**Slovensko društvo INFORMATIKA, Vožarski pot 12, 1000 Ljubljana.**

Lahko pa izpolnite obrazec na domači strani društva: <http://www.drustvo-informatika.si>

## Naročilnica na revijo UPORABNA INFORMATIKA

Revijo naročam(o)  s plačilom letne naročnine SIT 5.900

izvodov po pogojih za podjetja SIT 17.800 za eno letno naročnino in SIT 11.900 za vsako nadaljnjo naročnino

po pogojih za študente letno SIT 2.800

(ime in priimek, s tiskanimi črkami)

(podjetje)

(davčna številka)

(ulica, hišna številka)

(pošta)

Datum:

Podpis:

Naročnino bomo poravnali najkasneje v roku 8 dni po prejemu računa.

### INTERNET

Vse bralce revije obveščamo, da lahko najdete domačo stran društva na naslovu: <http://www.drustvo-informatika.si>

Obiščite tudi spletne strani mednarodnih organizacij, v katere je včlanjeno naše društvo: IFIP: [www.ifip.or.at](http://www.ifip.or.at), ECDL: [www.ecdl.com](http://www.ecdl.com), CEPIS: [www.cepis.com](http://www.cepis.com)

# Popoln E-Business Suite



Vse aplikacije zasnovane enotno.  
Vse informacije na enem mestu.

**ORACLE®**

[www.oracle.si](http://www.oracle.si)

## ▶ **Razprave**

Tomaž Dogša

**Problematika sistemov za vodenje popraviljanja in vzdrževanja**

Stjepan Pervan, Martina Kern

**Reševanje problema pomanjkanja unikatnih naslovov internetnega protokola**

Krista Rizman Žalik

**Podatkovna skladišča in kakovost podatkov**

Miro Ribič, Andrej Kovačič, Marjan Lončarič

**Sistem za upravljanje delovnih procesov**

Marko Hölbl, Petra Grm, Boštjan Brumen, Tatjana Welzer, Izidor Golob

**Natančnost uvrščanja slovenskih besedil**

## ▶ **Poročila**

Ladislav Mikola

**Uporaba desetiških SI predpon v informatiki**

Zsuzsana Toszegi

**Digitaliziranje madžarske kulturne dediščine**

## ▶ **Koledar prireditev**

ISSN 1318-1882



9 771318 188001