

Univerza v Ljubljani  
Fakulteta za računalništvo in informatiko

5

10 **ZDRUŽEVANJE HETEROGENIH PODATKOVNIH VIROV**  
**NAPRAV ZA OBDELAVO PODATKOV**

15 Predloženi izum se nanaša na postopek združevanja heterogenih podatkovnih nabo-  
rov znotraj naprave z možnostjo zajema in analize podatkov, kot na primer osebni  
ali tablični računalnik, podatkovni strežnik in podobno, kjer omenjene podatkovne  
nabore tvorijo bodisi porazdeljeni podatkovni viri, kot na primer podatkovne baze,  
skladišča, prosto-dostopni registri, spletne storitve in podobno, pri čimer za enega  
20 ali več omenjenih podatkovnih virov ni znana formalna preslikava, ki bi omogočala  
enoznačno združevanje omenjenih podatkovnih naborov v omenjeni vsakokratni  
napravi. Pri tem je omenjena naprava poljubno programirljiv stroj z omrežno  
povezavo in možnostjo zahtevnejšega procesiranja ter pomnjenja večje količine po-  
datkov. Izum sodi v razred G06F 17/30 mednarodne patentne klasifikacije.

25

Znanih je več različnih postopkov združevanja podatkovnih naborov, ki jih tvo-  
rijo različni podatkovni viri. Takšno združevanje podatkovnih naborov v splošnem  
temelji na predpostavki o enotni homogeni predstavitvi, pri čimer izbira uporablje-  
nih algoritmov združevanja določa omenjeno predstavitev podatkovnih naborov.  
30 Pomanjkljivost dosedanjih rešitev na področju združevanja podatkovnih naborov,  
ki jih tvorijo različni podatkovni viri, je tako zlasti predpostavka o enotni pred-  
določeni predstavitvi omenjenih podatkovnih naborov, ki bodisi ne upošteva ali  
preddoloči na primer pomen ali relacije med posameznimi deli omenjenih podat-  
kovnih naborov, ali zaupanje uporabnika v omenjene podatkovne vire.

Naloga predloženega izuma je ustvariti postopek združevanja heterogenih podatkovnih naborov, ki jih tvorijo različni podatkovni viri, ki v okviru poljubne naprave za obdelavo podatkov omogoča enostavno in zanesljivo obravnavo splošnih podatkovnih virov, dočim je omenjeni postopek neodvisen od predstavitve in vsebine omenjenih podatkovnih naborov ter uporabljenih algoritmov združevanja. Tehnični problem, ki ga rešuje izum, je pridobivanje, združevanje, analiza in hramba heterogenih podatkovnih naborov v okviru poljubne naprave za obdelavo podatkov, pri čimer omenjena naprava prilagodi svoje delovanje omenjenim vsakokratnim podatkovnim virom ter tako zagotovi večji učinek pri njihovi obdelavi. Slednje uporabniku omenjene naprave omogoča enotno obravnavo omenjenih podatkovnih naborov, kar močno poveča učinkovitost preiskovanja, preučevanja in obdelave splošnih podatkovnih virov. Združevanje heterogenih podatkovnih naborov, ki jih tvorijo podatkovni viri, ki se nahajajo na fizično ločenih lokacijah in so realizirani z različnimi tehnologijami, je namreč ključnega pomena v številnih uporabniških storitvah in napravah, kot na primer sodobni elektronski pripomočki, navigacijske naprave, multimedijski portali, vremenske postaje in podobno.

Zastavljeni problem je po izumu rešen s postopkom, ki v okviru omenjene vsakokratne naprave glede na zadostno mero podobnosti enotne predstavitve omenjenih podatkovnih naborov določi enoznačne entitete, ki jih opisujejo posamezni deli omenjenih podatkovnih naborov. V omenjeni napravi na osnovi omenjenih entitet odpravi redundantne dele ter združi omenjene podatkovne nabore s poljubnim algoritmom združevanja. Po izumu je predvidena splošna tri-nivojska predstavitev omenjenih podatkovnih naborov, ki upošteva lastnosti entitet, ki jih opisujejo omenjeni podatkovni nabori, relacije med omenjenimi entitetami, pomen omenjenih lastnosti in relacij entitet ter zaupanje uporabnika v omenjene podatkovne vire. Rezultat postopka po izumu je združen podatkovni nabor shranjen v omenjeni napravi, ki zajema vse informacije vsebovane v omenjenih podatkovnih naborih.

Izum je v nadaljevanju opisan s sklicevanjem na priložene skice, ki prikazujejo

sl. 1 shematski prikaz tri-nivojske predstavitve podatkovnih naborov,

sl. 2 shematski prikaz realizacije tri-nivojske predstavitve,

sl. 3 shematski potek postopka po izumu.

Izum je v nadaljevanju podrobneje opisan na osnovi izvedbenega primera, kjer je omenjena tri-nivojska predstavitev podatkovnih naborov v okviru omenjene vsakokratne naprave realizirana na osnovi grafov ali omrežij, ontologij in podatkovnih skupkov. Nadalje je za omenjen algoritem združevanja omenjenih podatkovnih naborov izbrano iterativno razvrščanje relacijskih podatkov.

Tri-nivojska predstavitev, ki je shematsko prikazana na sl. 1, po izumu omenjene podatkovne nabore znotraj omenjene naprave enoznačno predstavi na podatkovnem, semantičnem in abstraktnem nivoju. Podatkovni nivo (oznaka A) omenjene predstavitve zajema lastnosti entitet, ki jih opisujejo omenjeni podatkovni nabori, ter relacije med omenjenimi entitetami. Semantični nivo (oznaka B) omenjene predstavitve zajema podatkovni nivo omenjenih podatkovnih naborov ter pomen omenjenih lastnosti in relacij entitet. Abstraktni nivo (oznaka C) omenjene predstavitve zajema preslikavo med podatkovnim in semantičnim nivojem omenjenih podatkovnih naborov, ter posledično oba omenjena nivoja.

Ločitev omenjene predstavitve podatkovnih naborov na podatkovni in semantični nivo po izumu je smiselna, saj omogoča uporabo poljubnih algoritmov združevanja v okviru omenjene vsakokratne naprave, ki na primer upoštevajo bodisi zgolj omenjene lastnosti in relacije entitet ali pa tudi pomen lastnosti in relacij entitet. Abstraktni nivo omenjene predstavitve na osnovi preslikave med omenjenima nivojema omogoča enotno obdelavo omenjenih podatkovnih naborov, ne glede na konkreten nivo obravnave omenjenih vsakokratnih algoritmov združevanja.

Tri-nivojska predstavitev omenjenih podatkovnih naborov je po izumu realizirana kot to prikazuje shema na sl. 2. Podatkovni nivo (oznaka A) in semantični nivo (oznaka B) sta realizirana neodvisno med seboj, kar pomeni kot dve fizično ločeni podatkovni strukturi v omenjeni vsakokratni napravi. Abstraktni nivo (oznaka C) je realiziran kot ustrezna enoznačna preslikava med posameznimi deli omenjenih podatkovnih struktur.

V konkretnem izvedbenem primeru je podatkovni nivo omenjenih podatkovnih naborov realiziran na osnovi grafov ali omrežij, pri čimer so omenjene entitete, ki

jih opisujejo omenjeni podatkovni nabori, predstavljene z vozlišči omrežja, relacije omenjenih entitet predstavljene s povezavami omrežja ter lastnosti omenjenih entitet in relacij predstavljene z oznakami in utežmi omenjenih vozlišč in povezav. Semantični nivo omenjenih podatkovnih naborov je realiziran na osnovi ontologij, pri čimer so lastnosti in relacije omenjenih entitet predstavljene kot na podatkovnem nivoju, entitete pa so nadalje označene z aksiomi, ki opisujejo pomen lastnosti in relacij omenjenih entitet. Abstraktni nivo omenjenih podatkovnih naborov je nazadnje realiziran na osnovi podatkovnih skupkov kar v zadevni patentni prijavi označuje trojice  $(p, i, v)$ , pri čimer je  $p$  enoznačen indikator omenjenega vsakokratnega podatkovnega vira,  $i$  enoznačen indikator omenjene vsakokratne lastnosti, relacije ali pomena lastnosti ali relacije entitete, in  $v$  vrednost, ki ustreza  $i$ .

Pri tem je seveda popolnoma očitno, da predstavitev omenjenih podatkovnih naborov na abstrakten nivoju po izumu omogoča enotno obdelavo z omenjenimi algoritmi združevanja v okviru omenjene vsakokratne naprave, ne glede na konkreten nivo obravnave omenjenega vsakokratnega algoritma.

Postopek združevanja heterogenih podatkovnih virov v okviru naprave za obdelavo podatkov, ki je shematsko prikazan na sl. 3, po izumu obsega dve glavni fazi, in sicer fazo preslikave in fazo združevanja omenjenih podatkovnih naborov.

Prva faza postopka po izumu obsega preslikavo omenjenih podatkovnih naborov v omenjeno tri-nivojsko predstavitev znotraj omenjene vsakokratne naprave, ki poteka kot sledi. Preslikava v predstavitev na podatkovnem nivoju (korak 1) in sočasno preslikava v predstavitev na semantičnem nivoju (korak 2). Nadalje sledi preslikava v enotno predstavitev na abstraktnem nivoju (korak 3). Omenjene preslikave potekajo v korakih, ki so popolnoma očitni iz opisa izvedbenega primera omenjene tri-nivojske predstavitve. Rezultat prve faze postopka po izumu je podatkovni nabor v omenjeni enotni tri-nivojski predstavitvi, ki zajema vse podatke vsebovane v omenjenih podatkovnih naborih in je shranjen v omenjeni napravi.

Druga faza postopka po izumu obsega združevanje omenjenih podatkovnih naborov v enoten podatkovni nabor znotraj omenjene vsakokratne naprave, ki poteka kot sledi. Določanje enoznačnih entitet (korak 4), ki jih opisujejo posamezni

deli omenjenih podatkovnih naborov, ki so si v zadostni meri podobne glede na omenjeno enotno tri-nivojsko predstavitev, pri čimer je uporabljen poljuben vsakokratni algoritem združevanja. Nadalje sledi odpravljanje redundantnih delov (korak 5) omenjenih podatkovnih naborov, ki jih tvori en ali več omenjenih podatkovnih virov, na osnovi obdelave omenjenih enoznačnih entitet. Podvojene dele omenjenih podatkovnih naborov, ki ustrezajo omenjenim enoznačnim entitetam in imajo enak zapis na abstraktnem nivoju omenjene predstavitve, zavržemo, dele, ki ustrezajo omenjenim entitetam in imajo različen zapis na omenjenem nivoju, pa sprejmemo na osnovi največjega zaupanja omenjenemu vsakokratnemu podatkovnemu viru. Nadalje sledi združevanje (korak 6) omenjenih podatkovnih naborov z odpravljenimi redundantnimi deli, ki je tedaj popolnoma očitno določeno na osnovi omenjenih enoznačnih entitet. Rezultat druge faze postopka po izumu je podatkovni nabor v omenjeni enotni tri-nivojski predstavitvi, ki zajema vse informacije vsebovane v omenjenih podatkovnih naborih in je shranjen v omenjeni napravi.

15

V konkretnem izvedbenem primeru je za omenjen algoritem združevanja izbrano razvrščanje relacijskih podatkov realizirano z iteracijo, ki obravnava omenjene podatkovne nabore na podatkovnem nivoju omenjene tri-nivojske predstavitve, pri čimer upošteva lastnosti in relacije omenjenih entitet. Zaradi enotne obdelave na abstraktnem nivoju omenjene predstavitve, omenjen algoritem posredno deluje tudi na semantičnem nivoju, ter upošteva pomen omenjenih lastnosti in relacij.

20

Seveda je po izumu mogoče združevanje podatkovnih naborov, ki jih tvorijo omenjeni podatkovni viri, pri čimer je en ali več omenjenih podatkovnih naborov shranjen v omenjeni vsakokratni napravi (t.i. bogatenje podatkovnih virov), ne da bi se s tem oddaljili od smisla in obsega zaščite.

25

30

**PATENTNI ZAHTEVKI**

1. Postopek združevanja heterogenih podatkovnih naborov v okviru naprave z  
možnostjo zajema in analize podatkov, kot na primer osebni ali tablični računal-  
5       nik, podatkovni strežnik in podobno, kjer omenjene podatkovne nabore tvorijo  
podatkovni viri, kot na primer podatkovne baze ali skladišča, prosto-dostopni  
registri, spletne storitve in podobno, pri čimer za enega ali več omenjenih  
podatkovnih virov ni znana formalna preslikava, ki bi omogočala enoznačno  
združevanje omenjenih podatkovnih naborov v omenjeni vsakokratni napravi,  
10       *značilen po tem*, da obsega:
  - a) preslikavo omenjenih podatkovnih naborov v enotno predstavitev znotraj  
omenjene naprave, pri čimer omenjena predstavitev zajema lastnosti enti-  
tet, ki jih opisujejo omenjeni podatkovni nabori, relacije med omenjenimi  
entitetami ter pomen omenjenih lastnosti in relacij;
  - 15       b) v omenjeni napravi določanje enoznačnih entitet, ki jih opisujejo posamezni  
deli omenjenih podatkovnih naborov, ki so si v zadostni meri podobne glede  
na omenjeno enotno predstavitev omenjenih podatkovnih naborov;
  - c) v omenjeni napravi odpravljanje redundantnih delov omenjenih podatkov-  
nih naborov, ki jih tvori en ali več omenjenih podatkovnih virov, na osnovi  
20       obdelave omenjenih enoznačnih entitet; in
  - d) v omenjeni napravi združevanje omenjenih podatkovnih naborov z odpra-  
vljenimi redundantnimi deli, glede na omenjene enoznačne entitete, ki jih  
opisujejo omenjeni podatkovni nabori.
  
- 25   2. Postopek po zahtevku 1, *značilen po tem*, da omenjena preslikava podatkov-  
nih naborov v enotno predstavitev v omenjeni vsakokratni napravi obsega:
  - a) preslikavo omenjenih podatkovnih naborov v enoznačno predstavitev na po-  
datkovnem nivoju, ki zajema lastnosti entitet, ki jih opisujejo omenjeni po-  
datkovni nabori, ter relacije med omenjenimi entitetami;
  - 30       b) preslikavo omenjenih podatkovnih naborov v enoznačno predstavitev na se-  
mantičnem nivoju, ki zajema podatkovni nivo omenjenih podatkovnih na-  
borov ter pomen omenjenih lastnosti in relacij; in
  - c) preslikavo omenjenih podatkovnih naborov v enotno predstavitev na abstrak-  
tnem nivoju, ki zajema preslikavo med predstavitvama na podatkovnem in

semantičnem nivoju omenjenih podatkovnih naborov.

3. Postopek po zahtevku 2, *značilen po tem*, da je omenjena tri-nivojska predstavitev podatkovnih naborov v okviru omenjene vsakokratne naprave na podatkovnem nivoju realizirana na osnovi grafov ali omrežij, na semantičnem nivoju realizirana na osnovi ontologij, na abstraktnem nivoju realizirana na osnovi podatkovnih skupkov oziroma kombinacija omenjenih možnosti.  
5
4. Postopek po kateremkoli od predhodnih zahtevkov, *značilen po tem*, da je en ali več omenjenih podatkovnih naborov, ki jih tvorijo omenjeni podatkovni viri, shranjen v omenjeni vsakokratni napravi (t.i. bogatenje podatkovnih virov).  
10
5. Postopek po kateremkoli od predhodnih zahtevkov, *značilen po tem*, da so omenjeni podatkovni viri bodisi lokalni, na eni fizični lokaciji, ali porazdeljeni med fizično ločenimi lokacijami.  
15

**POVZETEK**

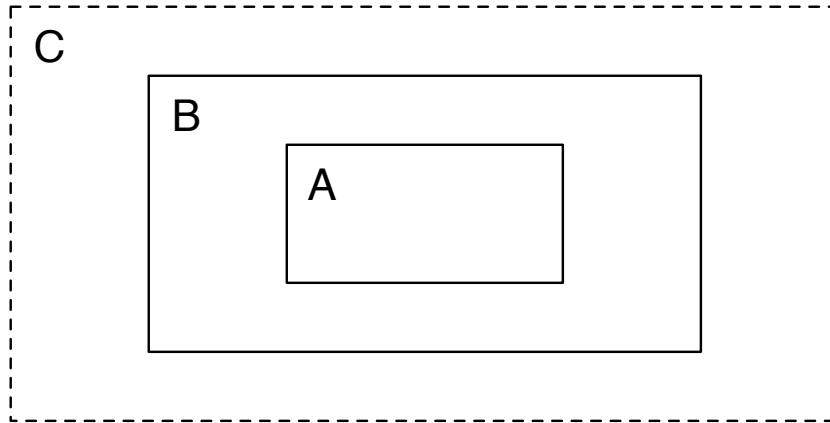
Predloženi izum se nanaša na postopek združevanja heterogenih podatkovnih naborov znotraj naprave z možnostjo zajema in analize podatkov, kot na primer osebni ali tablični računalnik, podatkovni strežnik in podobno, kjer omenjene podatkovne nabore tvorijo bodisi porazdeljeni podatkovni viri, pri čimer za enega ali več omenjenih podatkovnih virov ni znana formalna preslikava, ki bi omogočala enoznačno združevanje omenjenih podatkovnih naborov v omenjeni vsakokratni napravi. Slednje uporabniku omenjene naprave omogoča enotno pridobivanje, obdelavo, analizo in hrambo heterogenih podatkovnih naborov. Po izumu je predvideno, da omenjena naprava glede na podobnost enotne predstavitve omenjenih podatkovnih naborov določi enoznačne entitete, ki jih opisujejo posamezni deli omenjenih podatkovnih naborov. Na osnovi omenjenih entitet se v omenjeni napravi odpravi redundantne dele ter združi omenjene podatkovne nabore. Po izumu je predvidena splošna tri-nivojska predstavitev omenjenih podatkovnih naborov, ki upošteva lastnosti entitet, ki jih opisujejo omenjeni podatkovni nabori, relacije med omenjenimi entitetami ter pomen omenjenih lastnosti in relacij.

20

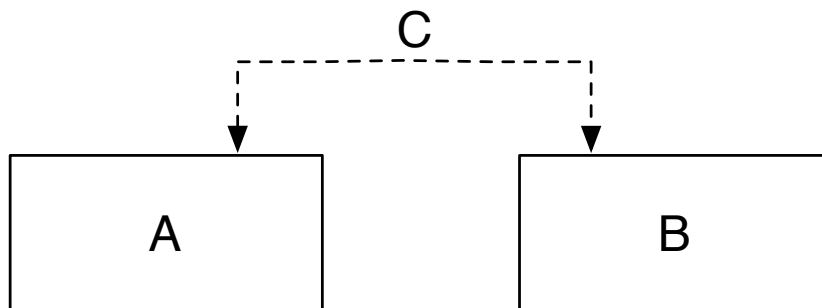
Univerza v Ljubljani

Fakulteta za računalništvo in informatiko





sl. 1



sl. 2

