

InfoGen: Informacijski sistem za sledenje, analizo in vodenje vzorcev na področju funkcijske genomike

Staš Hvala¹, Žiga Pušnik¹, Živa Rejc², Uršula Prosenc Zmrzljak³,
Kaja Blagotinšek³, Žiga Urlep³, Nejc Nadižar³, Damjana Rozman³, Miha Moškon¹

¹Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Večna pot 113, SI-1000 Ljubljana

²Fakulteta za kemijo in kemijsko tehnologijo, Univerza v Ljubljani, Večna pot 113, SI-1000 Ljubljana

³Center za funkcijsko genomiko in biočipe, Inštitut za biokemijo, Medicinska fakulteta, Univerza v Ljubljani, Zaloška 4, SI-1000 Ljubljana

E-pošta: ziga.pusnik@gmail.com, stas.hvala@gmail.com, miha.moskon@fri.uni-lj.si

InfoGen: Information system for tracking, analysis and management of functional genomics samples

We present the establishment of the information system as a support system for tracking, analysis and management of animal and/or human samples acquired in the biochemical research laboratories and/or clinical environment. The proposed information system supports various tasks. These can be divided among (1) acquisition, storage and identification of samples, (2) management of acquired data and (3) various computational and statistical analyses. The system incorporates different experimental workflows, i.e. from the initial collection of samples to the biochemical analysis of obtained data. It is divided into three different segments that are designated to different research laboratories, which are focused to the analysis of (1) animal collected samples, (2) samples acquired from suicide victims and (3) samples for the prognosis of Alzheimer's disease. Each segment includes the support for designated analyses that can be performed on the target samples. The system is compatible with the majority of commercial software products for data management and analysis, since it support the data import and export in a CSV (Comma Separated Values) format. The proposed system presents a platform tailored to specific users and is an excellent alternative to the popular general-purpose computational tools that are currently prevalent within research laboratories.

1 Uvod

V raziskovalnih laboratorijih in na klinikah se dnevno zbere enormno število laboratorijskih vzorcev, kot so npr. vzorci krvi ali tkiva [1]. Odvzeti vzorci so v nadaljevanju podvrženi številnim analizam, pri katerih lahko njihovo število še narašča (npr. izolacija DNA ali RNA). Zaradi trenutnega načina vodenja podatkov o vzorcih, ki je v večini primerov nesistematičen in pomanjkljiv ter se lahko bistveno razlikuje že v različnih laboratorijih znotraj iste organizacije, prihaja do izgubljanja podatkov in sledljivosti nad shranjenimi vzorci.

V prispevku predstavimo razvoj informacijskega sistema, ki predstavlja alternativo splošno-namenskemu orodju za vodenje in analizo zajetih podatkov ter je prilagojen trem ciljnim laboratorijem. Informacijski sistem podpira sledenje, analizo in vodenje vzorcev pri (1) delu z

mišjimi vzorci, (2) pri delu s humanimi vzorci žrtev samomora in (3) pri delu s humanimi vzorci za diagnostiko Alzheimerjeve demence. Informacijski sistem predstavlja računalniško podporo laboratorijskemu delu, ki zajema cel proces od (1) zajema in hrambe vzorcev, (2) upravljanja z zajetimi podatki ter do (3) izvajanja računalniških in statističnih analiz nad zajetimi podatki. Informacijski sistem podpira uvoz in izvoz podatkov v formatu CSV (ang. Comma Separated Values), s čimer zagotavljamo kompatibilnost z večino ostalih orodij, ki se uporabljajo pri laboratorijskem delu. Označevanje in identifikacija vzorcev je implementirana z uporabo kod QR (ang. Quick Response Code). Sistem je implementiran kot spletna aplikacija, do katere lahko dostopoma iz zaščitene notranjega omrežja ciljnega laboratorija.

V nadaljevanju prispevka predstavimo metode, ki so bile uporabljene pri razvoju informacijskega sistema, njegovo strukturo in njegovo delovanje ponazorimo na študiji primera pri delu z mišjimi vzorci. V zaključku komentiramo rezultate in podamo smernice pri nadaljnjem delu.

2 Informacijski sistem

Informacijski sistem je razdeljen v tri glavne podsegmente, vsem segmentom pa je skupna avtentikacija uporabnika in administracija uporabniških vlog. Gre za integriran večnivojski sistem, kjer ima uporabnik dostop do vseh svojih segmentov. Čeprav gre za integriran sistem, je vsa poslovna logika ločena v različnih izvornih datotekah. Celoten informacijski sistem je namreč razvit v spletnem razvojnem ogrodju *Laravel* v programskem jeziku *PHP* (ang. *Personal Home Page* oz. *PHP: Hypertext Preprocessor*). Prav tako so ločene tudi entitetno-relacijske podatkovne baze. Takšna arhitekturna shema preprečuje neavtorizirane dostope, saj imajo uporabniki omogočen dostop le do dovoljenih segmentov in njihovih podatkovnih baz. Informacijski sistem je dostopen iz enotne vstopne točke v zaščitene notranje omrežje.

2.1 Ogrodje Laravel

Laravel [2] je prosto dostopno, odprtokodno razvojno ogrodje (angl. framework) za razvoj spletnih aplikacij v programskem jeziku *PHP*, ki temeljijo na *MVC* (ang. *Model View Controller*) arhitekturnem vzorcu. Prvi del arhitekture je model (ang. *model*), kjer definiramo objekte

z ustreznimi relacijami. Model nam omogoča avtomatsko preslikavo med objekti in podatkovnimi modeli. V kontrolerju (ang. *controller*) definiramo poslovno logiko, oziroma logiko sistema. Tukaj operiramo z objekti in ostalimi spremenljivkami. Te nato prenesemo v pogled (ang. *view*), kjer se s pomočjo pogona za kreiranje *HTML* (ang. *Hyper Text Markup Language*) kode *Blade* generira spletna stran. Vse našete operacije se izvedejo na strežniku, na koncu pa strežnik vrne zahtevek za ustrezno *HTML* kodo, ki jo je odjemalčev brskalnik sposoben prikazati. Glavne prednosti *Laravel* ogrodja so: integrirana orodja za avtomatsko postavitev in vzdrževanje sistema v strežniškem okolju, različni načini dostopa do entitetno-relacijskih baz, migracije podatkovnih modelov v podatkovne baze in sintaktično urejena koda. Ena izmed pomembnejših orodij, ki nam omogoča dostop do entitetno-relacijskih podatkovnih baz je *Eloquent API* (ang. *Application programming interface*), ki hkrati podpira tudi objektno relacijsko preslikavo. *Laravel* je poleg *Symphony*, *CodeIgniter* ter *Yii2* ena izmed najbolj popularnih PHP ogrodij [3].

3 Sledljivost vzorcev in izvoz podatkov

V informacijskem sistemu lahko vnašamo nove vzorce, jih urejamo in brišemo. Za sledenje laboratorijskih vzorcev uporabljamo kodiranje *QR*. Koda *QR* je dvodimenzionalna matrična črtna koda, ki jo lahko preberemo z ustreznim črtnim čitalcem. Zaradi visoke učinkovitosti lahko na površino v velikosti poštne znamke lahko zakodiramo več tisoč bajtov, poleg tega pa redundanca kodiranja omogoča detekcijo in korekcijo potencialnih napak [4]. S tem so kode *QR* primerne tudi za označevanje laboratorijskih vzorcev. Pri kreaciji novega vzorca v informacijskem sistemu nam ta omogoči izvoz kode *QR* v obliki ukazne datoteke, ki vsebuje navodila kje in kako naj tiskalnik to kodo natisne. Nalepka na katero natisnemo kodo *QR* je tako dovolj majhna, da jo lahko nalepimo na epruveto laboratorijskega vzorca, ki je namenjena trajni hrampi. Poleg zaporedne številke vzorca lahko koda *QR* vsebuje tudi zaporedno številko laboratorija in projekta. Takšen pristop omogoča sledljivost vzorcev. Iz vsakega označenega laboratorijskega vzorca lahko odčitamo njegovo zaporedno številko in v sistemu poiščemo potrebne informacije, kot je lokacija shrambe ali npr. kdo je nazadnje rokoval z vzorcem.

Za izvajanje statističnih analiz se poslužujemo izvoza podatkov. Implementacija osnovnih statističnih analiz bi bila nesmiselna, saj bi s tem zadostili le osnovnim potrebam, še vedno pa bi za naprednejše analize raziskovalci morali izvažati podatke in jih analizirati v drugih orodjih za obdelavo podatkov. Informacijski sistem omogoča izvoz vseh podatkov v obliki *CSV* (ang. *Comma Separated Values*) datoteke. Gre za standarden format podatkovnih datotek podprt v vseh večjih orodjih za delo s podatki. Takšen pristop nam omogoča izvajanje poljubne statistike nad različnimi podatki z poljubnimi orodji.

Mouse ID	Sample ID	Genotype	Sex	Date of birth	Diet	Edit
2	573	low/Cre	Female	2011-10-16	Altramin	Q E
3	574	0/low/Cre	Female	2011-10-16	Altramin	Q E
4	579	low/Cre	Male	2011-10-16	Altramin	Q E
5	580	0/low/Cre	Female	2011-10-16	Altramin	Q E
6	582	low/Cre	Female	2011-10-16	Altramin	Q E
7	583	0/low/Cre	Female	2011-10-19	Altramin	Q E
8	586	low/Cre	Female	2011-10-19	Altramin	Q E
9	588	low/Cre	Female	2011-10-19	Altramin	Q E
10	593	low/Cre	Male	2011-10-22	Altramin	Q E
11	599	0/low/Cre	Female	2011-11-12	Altramin	Q E

Slika 1: Izgled tabele podatkov za miši.

Sample Type: Plasma

Cholesterol [mg/dL]	LDL [mg/dL]
2.27	0.19
HDL [mg/dL]	Free Fatty Acid [mg/dL]
1.46	1.49
Triglyceride [mg/dL]	TNFa [mg/dL]
0.00	0.23
AST [mg/dL]	ALT [mg/dL]
2.36	0.54
Total bilirubin [mg/dL]	Direct conjugated bilirubin [mg/dL]
0	0
Control [mg/dL]	Side Note
0	

Slika 2: Forma za vnos podatkov za vzorec plazma.

4 Študija primera: podpora delu z mišjimi vzorci

Za demonstracijo aplikacije bomo v pričujočem razdelku podrobneje predstavili segment za delo z mišjimi vzorci. Spletni vmesnik (ang. *interface*) ponuja možnost vnašanja, urejanja, brisanja in pregleda podatkov o laboratorijskih miškah in podatkov o zajetih vzorcih. Kratica *CRUD* (ang. *Create, Read, Update and Delete*) zajema osnovne funkcionalnosti vsake aplikacije, ki temelji na hranjenju podatkov. Kot primer uporabe sistema lahko uporabnik najprej vnese podatke za laboratorijsko miš, nato pa izpolni formo za vsak pripadajoči vzorec. Kasneje lahko vzorce ter miši ureja ali izbriše. Podatki se pravzaprav nikoli ne izbrišejo, pač pa se le označijo za izbrisane. S tem se izognemo neželeni izgubi podatkov. Za vsako spremembo podatkov o vzorcu se zabeleži tudi elektronski naslov uporabnika, ki je spremembo naredil. S tem zagotovimo sledljivost sistema. Primer pregledne tabele za vnesene miši je prikazan na sliki 1, primer forme za kreiranje vzorcev pa lahko vidimo na sliki 2.

Uporabnik ima možnost vnosa devetih tipov vzorcev: plazma, genotip, DNA (ang. *DeoxyriboNucleic Acid*), RNA (ang. *RiboNucleic Acid*), cDNA (angl. *complementary DNA*), metabolit, protein, *western* prenos (ang. *western blot*) ter tkivna rezina (ang. *tissue slice*) s prisotnimi patološkimi stanji. Potek zajema vzorcev v laboratorijskem okolju je prikazan na sliki 3. Diagram poteka

je služil tudi kot osnovno izhodišče pri zasnovi izgleda aplikacije in uporabniške izkušnje (ang. *User Experience*, UX). Vsi podatki o vzorcih in miših se zapisujejo v podatkovno bazo, ker ponuja zanesljivo hrambo, konsistenco pri večjem številu uporabnikov ter boljšo skalabilnost pri naraščajočem številu podatkov. Entitetno-relacijska shema za segment sistema za delo z mišjimi vzorci je prikazana na sliki 4.

5 Zaključek

Predstavljeni informacijski sistem uporabniku zagotavlja ustrezno podporo pri organizaciji zbranih vzorcev in podatkov. Še posebej je primeren za uporabo v laboratorijih, v katerih se dnevno zbere zelo veliko podatkov, prav tako pa podpira izmenljivost podatkov med različnimi laboratoriji. Sistem je enostaven za uporabo. Ogrodje, ki je bilo vzpostavljeno v projektu je prilagodljivo za uporabo v laboratorijih oziroma nad vzorci, ki v trenutnem delu še niso bili zajeti in podprti. Uporaba odprtokodnih orodij in platform pri uporabi ne predstavlja kakršnihkoli birokratskih omejitev. Trdno verjamemo, da je predstavljeni informacijski sistem dobra alternativa splošno-namenskimi računalniškimi orodjem, ki so večinoma v uporabi trenutno.

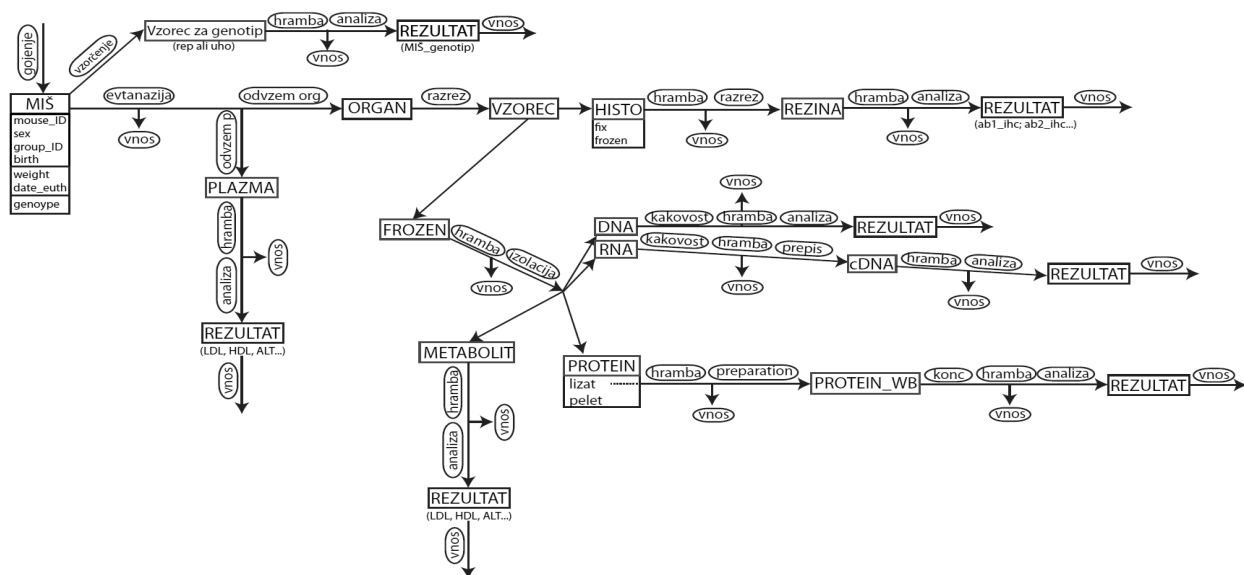
Smatramo, da bomo lahko z delom na vzpostavljenem informacijskem sistemu nadaljevali tudi v prihodnosti. Pri tem bomo posebno pozornost posvetili njegovi posplošitvi, tako da bo uporaben na širšem področju funkcijske genomike kot tudi na ostalih področjih, ki zahtevajo natančno vodenje in podporo dela z laboratorijskimi vzorci.

Zahvala

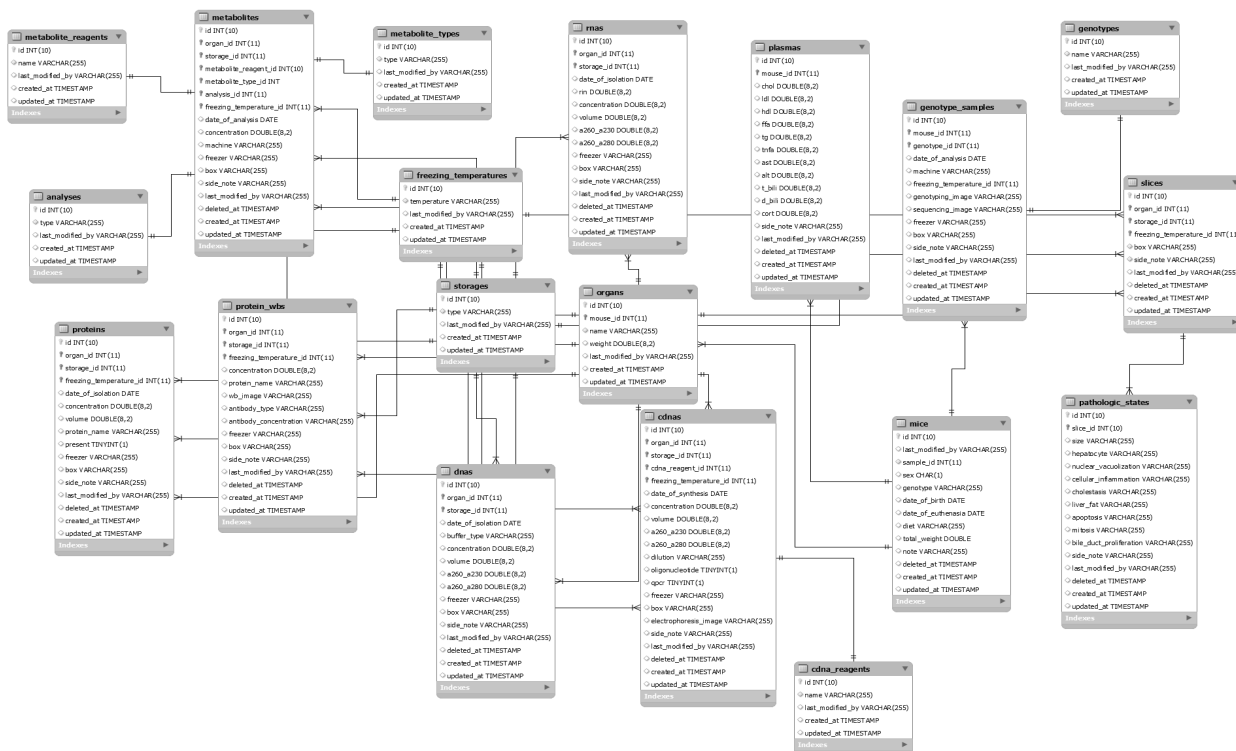
Projekt so sofinancirali Republika Slovenija in Evropska unija iz Evropskega socialnega sklada (25-13-3), Agencija za raziskave in razvoj Republike Slovenije v okviru znanstveno raziskovalnih programov P1-0390 in P2-0359 in infrastrukturni program ELIXIR.

Literatura

- [1] Goodwin, S., McPherson, J. D., McCombie, R. W.: Coming of age: Ten years of next generation sequencing technologies, *Nature Reviews Genetics*, **17**, 333–351 (2016)
- [2] Otwell, T.: *Laravel, The PHP Framework For Web Artists*, (2016)
- [3] Skvorc, B.: *The Best PHP Framework for 2015: SitePoint Survey Results, SitePoint–Learn HTML, CSS, JavaScript, PHP, Ruby & Responsive Design*, (2015)
- [4] Kieseberg, P., Leithner, M., Mulazzani, M., Munroe, L., Schrittwieser, S., Sinha, M., Weippl, E.: QR code security, *Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia*, 430–435, (2010)



Slika 3: Shema poteka dela v ciljnem laboratoriju.



Slika 4: Shema entitete relacijske podatkovne baze za delo z mišjimi vzorci.