

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

FIIT-5212-47914

Ivan Srba

SLEDOVANIE INTENZITY VZŤAHOV MEDZI POUŽÍVATEĽMI
V RÁMCI SOCIÁLNYCH SIETÍ

Bakalárska práca

Študijný program: Informatika

Študijný odbor: 9.2.1 Informatika

Miesto vypracovania: Ústav informatiky a softvérového inžinierstva, FIIT STU v Bratislave

Vedúca práce: prof. Ing. Mária Bieliková, PhD.

máj 2010

ZADANIE BAKALÁRSKEHO PROJEKTU

Meno študenta: **Ivan Srba**
Študijný odbor: INFORMATIKA
Študijný program: Informatika
Názov projektu: **Sledovanie intenzity vzťahov medzi používateľmi v rámci sociálnych sietí**

Zadanie:

Súčasný web charakterizuje najmä stále zvyšujúca sa interaktivita medzi používateľmi. Web sa mení z úložiska informácií na sociálny priestor, v rámci ktorého ľudia informácie nielen vyhľadávajú, ale aj navzájom komunikujú. Zrejme najväčšími koncentrátormi používateľov sú sociálne portály, v rámci ktorých používatelia nadväzujú a aj explicitne potvrdzujú rôzne druhy vzťahov. Iným príkladom sociálneho priestoru, kde vznikajú vzťahy medzi používateľmi sú rôzne diskusné fóra, či internetové noviny. Vzťahy medzi používateľmi však existujú s rôznou intenzitou. Analyzujte problematiku vývoja vzťahov medzi používateľmi pri práci s informáciami na webe. Navrhňte a implementujte webovú aplikáciu, ktorá bude na základe činnosti používateľa aproximovať intenzitu jeho vzťahu k ostatným používateľom v závislosti od času. Overte navrhnuté riešenie v konkrétnom vybranom prostredí (sociálny portál, internetové noviny alebo diskusné fórum).

Práca musí obsahovať:

- Anotáciu v slovenskom a anglickom jazyku
- Analýzu problému
- Opis riešenia
- Zhodnotenie
- Technickú dokumentáciu
- Zoznam použitej literatúry
- Elektronické médium obsahujúce vytvorený produkt spolu s dokumentáciou

Miesto vypracovania: Ústav informatiky a softvérového inžinierstva, FIIT STU, Bratislava
Vedúci projektu: prof. Ing. Mária Bieliková PhD.

Termín odovzdania práce v zimnom semestri: dňa 10. decembra 2009

Termín odovzdania práce v letnom semestri: dňa 14. mája 2010

Bratislava, dňa 21. septembra 2009



prof. Ing. Pavol Návrát, PhD.
riaditeľ ÚISI

Anotácia

Slovenská technická univerzita v Bratislave
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ
Študijný program: Informatika

Autor: Ivan Srba

Bakalársky práca: Sledovanie intenzity vzťahov medzi používateľmi v rámci sociálnych sietí

Vedúca bakalárskej práce: prof. Ing. Mária Bieliková, PhD.

Máj 2010

Súčasný web charakterizuje najmä stále zvyšujúca sa interaktivita medzi používateľmi. Web sa mení z úložiska informácií na sociálny priestor, v rámci ktorého ľudia informácie nielen vyhľadávajú, ale aj navzájom komunikujú a spolupracujú. Zrejme najväčšími koncentrátormi používateľov sú sociálne portály, v rámci ktorých používatelia nadväzujú a aj explicitne potvrdzujú rôzne druhy vzťahov. Ak vieme o silnom alebo naopak slabom vzťahu dvoch používateľov, môžeme vykonať akcie, ktoré závisia od konkrétnej oblasti nasadenia. Ako príklad môžeme uviesť adaptívne odporúčanie. Iným príkladom je, že používatelia môžu kontrolovať vývoj svojich vzťahov. Ak niektorý z nich začne mať tendenciu slabnúť, môžeme používateľa upozorniť a navrhnúť spôsob, ako to zmeniť (napr. poslaním správy alebo virtuálneho darčeka).

Naším hlavným cieľom je navrhnúť metódu pre analýzu vývoja používateľových vzťahov a jej overenie prostredníctvom webovej aplikácie, ktorá bude aproximovať intenzitu týchto vzťahov v čase. Táto aproximácia je založená na rozličných používateľových aktivitách. Ich príkladom môže byť posielanie správy alebo publikovanie spoločnej fotografie. Takéto aktivity nazývame hodnotiace faktory. Čiastočná intenzita vzťahu vyjadrená jedným hodnotiacim faktorom závisí od váhy hodnotiaceho faktora, od počtu všetkých výskytov daného hodnotiaceho faktora a pre niektoré typy aktivít aj od času, kedy sa aktivita udiala a od doby vplyvu tejto aktivity na spoločný vzťah. Všetky tieto vplyvy sú zahrnuté v sekvencii výpočtov.

Na overenie navrhutej metódy môžeme použiť rozličné sociálne siete. Pre experiment sme vybrali známy a populárny sociálny portál Facebook. Pre potreby realizácie metódy sme vyvinuli softvérový nástroj s názvom IRe, ktorý používa obalovač, pomocou ktorého sa pripája na sociálnu sieť Facebook a doluje výskyty hodnotiacich faktorov. V experimente sme dosiahli tri výsledky. Prvým je, že navrhnutá metóda vypočítala intenzitu vzťahov pre prvých 10 najlepších priateľov s úspešnosťou 88%. Druhým získaným výsledkom je, že nami vypočítaná intenzita opisuje podobné rozloženie interakcie medzi priateľmi, aké je výsledkom druhých výskumných prác. Tretím je, že vypočítaná intenzita reprezentuje podobný vývoj vzťahov v čase, ako je opísaný v súvisiacich výskumoch.

Annotation

Slovak University of Technology Bratislava
FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES
Degree course: Informatics

Author: Ivan Srba

Bachelor Theses: Tracing strength of relationships between users in social networks

Supervisor: prof. Ing. Mária Bieliková, PhD.

May 2010

Current web is known as a space with constantly growing interactivity among its users. It is changing from a place for data storage to the social place. Place where people not only search interesting information, but also communicate and collaborate each with other. Obviously, the best and biggest places for common interaction are social networks where people can arrange and explicitly express their relationships. If we know about strong or weak relationship intensity between two users we can provide actions which depend on actual context of deployment. For example we can provide adaptive recommendation. Also the user can control evolution of his or her relationships to friends. If some of these relationships have a tendency to weak, we can notify this user and propose the way how to change current state (e.g., send a message or send some virtual gift).

We present a method for analysis of the evolution of user's relationships and its evaluation by means of developed web-based application, which approximates the user's relationships with other users in time. This approximation is based on varied user's activities performed in social networks. Example of this activity is sending a message or uploading common photography. Such activity we denote as a rate factor. Partial intensity expressed by one rate factor depends on rate factor's weight, the count of all appearances of the rate factor and for some types of activities also on time when activity happens and the duration of influence. All this effects are included in sequence of calculation.

We can use many sources of user's activities to evaluate proposed method. We chose for the experiment well known and popular social portal Facebook. We developed software tool named IRe to realize the proposed method. This application uses a wrapper to connect to social network Facebook and to data mine rate factors via Facebook API. In the experiment we achieved three results. First one is that the method calculated relationship intensity for first ten best friends with 88% precision. Second one is that the calculated intensity described similar distribution of user interaction among friends as in related works. Finally, third achieved result is that the calculated intensity represented similar evolution of relationships in time as it was described in other researches.

Obsah

1	ÚVOD	1
2	SOCIÁLNE SIETE NA WEBE	3
2.1	Pojem sociálna sieť	3
2.2	Metódy a techniky práce so sociálnymi sieťami	3
2.2.1	Techniky zbierania dát	4
2.2.2	Ohodnotenie vzťahov	4
2.2.3	Základné vlastnosti siete a jej uzlov	5
2.3	Existujúce sociálne siete	6
2.4	Aplikácie pre analýzu sociálnych sietí	9
2.5	Zhodnotenie analýzy sociálnych sietí	11
3	METÓDA PRE ANALÝZU INTENZITY VZŤAHOV	13
3.1	Zabezpečenie prístupu k údajom sociálnych sietí	13
3.2	Spracovanie údajov sociálnych sietí	13
3.3	Výpočet intenzity vzťahov	15
4	REALIZÁCIA METÓDY PRE ANALÝZU INTENZITY VZŤAHOV	21
4.1	Prístup k údajom sociálnych sietí	21
4.2	Štruktúra softvérového nástroja pre analýzu vzťahov IRe	24
4.2.1	Architektúra softvérového nástroja	24
4.3	Štruktúra údajov v softvérovom nástroji IRe	25
4.4	Implementácia a testovanie	26
5	EXPERIMENTÁLNE OVERENIE METÓDY ANALÝZY INTENZITY VZŤAHOV	29
5.1	Definovanie experimentu	29
5.2	Vyhodnotenie experimentu	29
5.2.1	Úspešnosť metódy	30
5.2.2	Rozloženie vypočítanej intenzity medzi priateľmi	30
5.2.3	Rozloženie vypočítanej intenzity v čase	31
6	ZHODNOTENIE	35
	LITERATÚRA	37
	INTERNETOVÉ ZDROJE	39
	POUŽITÉ SKRATKY	41
	PRÍLOHA A – POŽIADAVKY NA SOFTVÉROVÝ NÁSTROJ	43
	A.1 Požiadavky na komponent IReA	43
	A.2 Požiadavky na komponent IReP	44
	PRÍLOHA B – MODEL PRÍPADOV POUŽITIA	45
	PRÍLOHA C – HODNOTIACE FAKTORY PRE SOCIÁLNU SIETĚ FACEBOOK	49
	PRÍLOHA D – TECHNICKÁ DOKUMENTÁCIA	51
	D.1 Diagram fyzického modelu údajov	51
	D.2 Štruktúry údajových entít fyzického modelu údajov	53
	D.3 Ukážka zdrojového kódu	57
	PRÍLOHA E – OPIS POUŽITIA SOFTVÉROVÉHO NÁSTROJA IRE	61

PRÍLOHA F – PRÍSPEVOK NA KONFERENCIU WEB INTELLIGENCE	67
PRÍLOHA G – OBSAH ELEKTRONICKÉHO MÉDIA.....	73

1 Úvod

V súčasnosti môžeme v priestore webu sledovať prudký rozmach novej generácie služieb známej pod označením Web 2.0. Časť týchto služieb tvoria sociálne siete, ktoré umožňujú používateľom navzájom realizovať rôzne typy interakcie. Ich popularita neustále rastie a stávajú sa súčasťou nášho každodenného života. Je to spôsobené predovšetkým tým, že sa mení prístup používateľov a ich dôvera k Internetu. Používatelia už nemajú tendenciu vystupovať v plnej anonymite ako tomu bolo v minulosti.

Práve naopak, poskytujú rozsiahle informácie o sebe, svojich záujmoch a svojich priateľoch. Tieto informácie nám dávajú možnosť bližšie spoznať používateľa z pohľadu jeho informačných potrieb a jeho preferencie. Musíme ich však vedieť efektívne spracovať a vhodne prezentovať. S využitím týchto údajov vieme vytvoriť širokú paletu rozličných aplikácií. Jednou z nich je práve aplikácia, ktorá realizuje nami navrhnutú metódu určenú na aproximáciu intenzity vzťahov medzi používateľmi sociálnych sietí.

Motiváciou pre vytvorenie takejto aplikácie je spoznanie vzťahov používateľa či už po stránke rôzne silných vzťahov k rozličným priateľom, alebo po stránke vývoja intenzity vzťahov v čase. Informáciu o silnom, resp. slabom spoločnom vzťahu dvoch používateľom potom vieme využiť pri riešení mnohých úloh.

Cieľom tohto projektu je návrh a overenie metódy, ktorá vyhodnotí intenzitu vzťahov daného používateľa k ostatným používateľom v priestore sociálnych sietí. Metóda je navrhnutá natoľko univerzálne, že ako zdroj údajov bude možné použiť ľubovoľnú sociálnu sieť, ktorá spĺňa vopred zadané kritériá.

Práca sa skladá zo šiestich kapitol. V druhej kapitole sa venujeme analýze sociálnych sietí, na ktorých bude celá metóda postavená. Zadefinujeme pojem sociálna sieť a predstavíme si niektoré z techník práce so sociálnymi sieťami, ktoré sme následne použili aj v samotnom návrhu metódy. Analyzujeme vedecké výskumy, ktoré sa zaoberajú rovnakou problémovou oblasťou a výsledky, ktoré boli v nich dosiahnuté. Taktiež analyzujeme existujúce sociálne siete a softvérové riešenia, ktoré sa zaoberajú podobnými problémami.

Tretia kapitola definuje predpoklady potrebné pre použitie metódy pre analýzu vzťahov. Tieto predpoklady sa delia na dve samostatné časti – požiadavky na sociálne siete a požiadavky na spracovanie údajov. Ďalej opisujeme samotný návrh metódy a možnosti využitia jej výsledkov.

V štvrtej kapitole opisujeme realizáciu navrhutej metódy a realizáciu softvérového nástroja. Na základe analýz sme zvolili sociálnu sieť, pre ktorú sme navrhli softvérový nástroj. Ďalej opisujeme životný cyklus vytvoreného softvéru od špecifikácie požiadaviek, cez návrh, výber technológií, implementáciu, až po testovanie.

V piatej kapitole určujeme ciele, ktoré sme sledovali v experimente. Definujeme priebeh experimentu a prezentujeme jeho vyhodnotenie.

Šiesta kapitola obsahuje zhrnutie práce vykonanej pri návrhu, realizácii a overení metódy analýzy intenzity vzťahov a opisuje ďalšie možné smerovanie projektu.

2 Sociálne siete na webe

Súčasný web charakterizuje najmä stále zvyšujúca sa interaktivita medzi používateľmi. Web sa mení z úložiska informácií na sociálny priestor, v ktorom ľudia informácie nielen vyhľadávajú, ale aj navzájom komunikujú. Zrejme najväčšími koncentrátormi používateľov sú sociálne portály, v rámci ktorých používatelia nadväzujú a aj explicitne potvrdzujú rôzne druhy vzťahov.

2.1 Pojem sociálna sieť

Pre správnu interpretáciu priestoru problémovej oblasti treba v prvom rade vysvetliť pojem sociálna sieť. Existuje viacero terminologických vysvetlení, vyberáme niekoľko z nich:

- a) Pojem sociálna sieť použil už v 50-tych rokoch J.A. Barnes, ktorý definoval sociálnu sieť ako:
 - „Sociálnu štruktúru zloženú z uzlov, ktorými sú všeobecne jednotlivci alebo organizácie. Sociálna sieť reprezentuje vzťahy a toky medzi ľuďmi, skupinami, organizáciami, zvieratami, počítačmi alebo inými entitami schopnými spracovať informácie/znalosti.“
- b) Kautz, Selman a Shah vo svojom príspevku *Referral Web: Combining Social Networks and Collaborative Filtering* [1] definujú sociálnu sieť takto:
 - „Pod pojem *sociálna sieť* zahŕňame skupiny ľudí, ktorí sú prepojení spoločnými profesionálnymi činnosťami. Nepoužívame slovo sociálny v užšom význame slova *priateľský*.“
- c) Autori Hanneman a Riddle vo svojej publikácii *Introduction to social network methods* [2] definujú sociálnu sieť prostredníctvom dát ako:
 - „*Sieťové dáta* (vo svojej najčistejšej podobe) sa skladajú z matice štvorcových rozmerov. Riadky matice sú prípady, predmety alebo pozorovania. Stĺpce matice sú rovnaké súbory prípadov, predmetov alebo pozorovaní (uvedomte si zásadný rozdiel od konvenčných dát, kde stĺpce sú atribúty). Každá bunka matice potom opisuje vzťah medzi aktérmi.“

Pre správne porozumenie pojmu sociálna sieť treba rozlišovať medzi sociálnymi sieťami v zmysle uvedených definícií a medzi webovými sociálnymi portálmi, ktoré sa často označujú pojmom sociálna sieť (napr. sociálny portál Facebook). Tieto portály sú webové aplikácie, ktoré poskytujú rozličné služby, pričom využívajú sociálnu sieť ako svoj základný koncept.

Na základe uvedených definícií môžeme chápať sociálnu sieť ako prostriedok podporujúci výmenu informácií a znalostí medzi entitami, ktorými môžu byť jednotlivci alebo organizácie. V prípade nášho projektu môžeme zaviesť konkrétnejšiu definíciu, kde prostriedkom je web, informácie a znalosti môžeme obmedziť len na údaje vyjadrujúce vzťah príslušných aktérov a všeobecnú skupinu jednotlivcov môžeme nahradiť za skúmaného používateľa s ďalšími používateľmi, ku ktorým explicitne vyjadril svoj vzťah.

2.2 Metódy a techniky práce so sociálnymi sieťami

Metódy a techniky na zbieranie dát a ohodnocovanie vzťahov, ktoré analyzujeme sú statické a priamo nezahŕňajú dynamiku vyjadrujúcu dimenziu času.

2.2.1 Techniky zbierania dát

Ak máme zadanú množinu používateľov vieme zozbierať potrebné údaje viacerými technikami. Na jednej strane spektra sú to metódy analýz celej siete, na opačnej strane zase metódy zamerané na analýzu len jedného samotného uzla s abstrahovaním vzťahov ďalších naviazaných uzlov. Autori Hanneman a Riddle vo svojej publikácii *Introduction to social network methods* [2] uvádzajú tieto techniky:

a) Analýza celej siete

- Zbierajú sa vzťahy všetkých uzlov z určitej skupiny so všetkými susednými uzlami.
- Poskytuje kompletný obraz o vzťahoch v rámci zvolenej skupiny.
- Získavanie dát je pre veľké siete pamäťovo a výpočtovo náročné, preto je táto metóda vhodná na analyzovanie malých skupín s dobre viditeľnými hranicami (napr. trieda v škole).

b) Metóda snehovej gule

- Táto metóda je založená na počiatočnom výbere jedného uzla alebo skupiny uzlov, pre ktoré sa vyhľadajú susedné uzly. Pre všetky tie susedné uzly, ktoré sa ešte nenachádzajú v zozname už navštívených uzlov, sa nájdu ďalšie susedné uzly a algoritmus sa zopakuje.
- Algoritmus sa končí, ak sa nenájdu už žiadne nové uzly alebo všetky ďalšie uzly sú už veľmi vzdialené od prvej skupiny uzlov.
- Nevýhodou tejto metódy je, že sa nemusia nájsť izolované uzly.
- Dôležitý je výber prvej skupiny uzlov, pretože pri nevhodnom výbere môže nastať situácia, že celé podskupiny uzlov zostanú nepreskúmané.
- Táto metóda je vhodná, keď skúmame úzko špecializovanú vrstvu uzlov medzi mnohými ostatnými uzlami (napr. ľudí, ktorí sa venujú jednému konkrétnemu záujmu).

c) Egocentrický (s pridanými vzťahmi)

- V egocentrickej metóde s pridanými vzťahmi vyberieme jeden uzol alebo skupinu uzlov a vyhľadáme ich susedov. Následne zistíme vzájomné vzťahy medzi týmito susedmi.
- Tento prístup je vhodný, keď nepotrebujeme skúmať makroštruktúry siete.
- Touto technikou sa dajú skúmať lokálne podskupiny väčších skupín, dá sa odhadnúť hustota siete a reciprocita vzťahov.
- Touto technikou sa nedajú určiť centrálné uzly, vzdialenosti medzi uzlami, ani porovnávanie pozícií subjektov.

d) Egocentrický (len ego)

- V egocentrickej metóde bez prídavných vzťahov vyberieme jeden uzol alebo skupinu uzlov a vyhľadáme len priamych susedov.
- Zamiera sa na jednotlivca, nie na sieť ako celok.
- Touto metódou sa nedá určiť väčšina štatistických hodnôt danej siete.

2.2.2 Ohodnotenie vzťahov

Hanneman a Riddle uvádzajú nasledujúce možnosti ohodnotenia vzťahov [2]:

1. Nominálne

a) Binárne

- Určuje či vzťah existuje (existencia je reprezentovaná číslom 1), alebo neexistuje (reprezentácia číslom 0).
- Väčšina teórií grafov používa práve binárne ohodnotenie vzťahov.
- Príkladom môže byť či dvaja skúmaní používatelia sú študentmi tej iste fakulty, alebo nie.

b) Multi-kategórie

- Vyjadruje nielen existenciu (príp. neexistenciu vzťahu), ale určuje aj vopred definovanú kategóriu typu vzťahu.
- Pre každú kategóriu možno následne vyjadriť vzťah v binárnom ohodnotení.
- Príkladom môže byť vzťah rodinného príslušníka, kde možno vybrať z preddefinovaných kategórií – rodič, súrodenec, syn, dcéra, iný rodinný príslušník.

2. Ordinálne

a) Skupinové

- Snaží sa opísať silu vzťahu.
- Určuje či vzťah je pozitívny (reprezentované číslom 1), neutrálny (reprezentované číslom 0), alebo negatívny (reprezentované číslom -1).
- Ordinálne dáta poskytujú viac informácií ako nominálne, ale existuje málo algoritmov na ich spracovanie, preto sa často prevádzajú na interval alebo na binárnu formu prostredníctvom zlomového bodu.

b) Plnorozsahové

- V niektorých prípadoch je možné zoradiť vzťahy od najsilnejšieho po najslabší.
- Tento prístup nie je prirodzený pre sociálne siete, a preto ani nenašiel široké uplatnenie.
- Príkladom môže byť zoradenie kolegov, kde číslo 1 priradím kolegovi, s ktorým sa mi spolupracuje najlepšie, číslo 2 druhému kolegovi v poradí atď.

c) Intervalom

- Najpresnejšie ohodnotenie vzťahov.
- Vhodným zvolením otázky (hodnotiaceho faktoru) sa dá vzťah kvantifikovať.
- Vzhľadom na to, že väčšina algoritmov bola navrhnutá pre binárne dáta, sa intervalové ohodnotenia často konvertujú do binárnej formy cez zlomový bod, ktorý však v niektorých prípadoch nie je jednoduché určiť a neexistuje žiadny formálny spôsob pre jeho definovanie.
- Príkladom môže byť frekvencia posielania emailov.

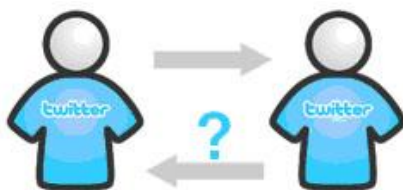
2.2.3 Základné vlastnosti siete a jej uzlov

Na sociálnu sieť sa môžeme pozeráť z dvoch pohľadov – na sieť ako celok a na jednotlivé uzly. Pre nami analyzovanú problematiku nie je zaujímavý prvý pohľad na sieť ako celok, keďže našim cieľom je skúmať len vzťahy konkrétneho používateľa sociálnej siete k ostatným používateľom, ku ktorým explicitne potvrdil vzťah. Pri pohľade na jednotlivé uzly môžeme na základe analýzy vzťahov so susednými uzlami rozdeliť uzly na diády a triády [2].

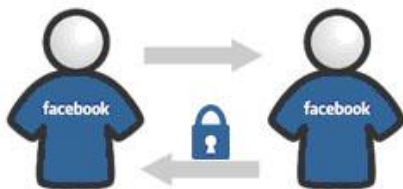
Diády sú dva subjekty a ich ľubovoľné vzájomné spojenie. Ak berieme do úvahy, že spojenie je orientované, tak môžu nastať 4 rôzne prípady:

- a) subjekty nie sú spojené,
- b) prvý subjekt vysiela, ale neprijíma,
- c) druhý subjekt vysiela, ale neprijíma,
- d) obidva subjekty vysiela aj prijímajú.

Konkrétne rozdiely v spojeniach môžeme poukázať na sociálnych sieťach Twitter a Facebook. V prípade sociálnej siete Twitter existuje asymetrický model, zobrazený na obr. 2.1, ktorý používa všetky 4 typy spojení. Na druhej strane sociálna sieť Facebook používa symetrický model vzťahov, ktorý je zobrazený na obr. 2.2. To znamená, že vzťah medzi dvoma používateľmi môže existovať až po jeho explicitnom akceptovaní oboma stranami a je nevyhnutne vzájomný.



Obr. 2.1 Asymetrický model spojení v diádach [15]



Obr. 2.2 Symetrický model spojení v diádach [15]

V triádach ide o vzťah troch subjektov. Tu je možných až 64 kombinácií. Vzhľadom na to, že sa budeme zaujímať len o priamy vzťah uzla s jedným naviazaným uzlom, nie je pre nás toto zoskupenie uzlov z pohľadu práce zaujímavé.

2.3 Existujúce sociálne siete

V súčasnosti existuje viacero vedeckých výskumov, ktoré sú zamerané na analýzu štruktúry a vlastností sociálnych sietí [2], [3], [4]. Tieto štúdie opisujú sociálne siete na vysokej úrovni a poskytujú len všeobecné informácie o uzloch a hranách medzi nimi. Nezachytávajú žiadne bližšie informácie, ktoré sa vzťahujú na spoločný vzťah dvoch konkrétnych používateľov.

Niektoré výskumy, napríklad výskum, ktorý vypracoval Maia a kol. [5] poskytuje konkrétnejšie informácie o používateľoch sociálneho portálu YouTube a rozdeľuje týchto používateľov do niekoľkých skupín vzhľadom na ich správanie v prostredí sociálneho portálu. Aj napriek tomu, že takáto informácia môže byť užitočná, neposkytuje nám dostatočne detailné ohodnotenie vzťahu dvoch používateľov.

Existuje len niekoľko výskumov, ktoré sú zamerané na skúmanie aktivít konkrétnych používateľov a spracovávajú tzv. *siete aktivít*. Pod týmto pojmom rozumieme sieť, ktorá je vytvorená používateľmi pri vyjadrovaní svojich vzťahov s využitím funkcionality sociálnej siete. Chun a kol. vypracoval prácu o sieti aktivít v sociálnom portáli Cyworld [6]. Výsledkom výskumu bolo, že

sieť aktivít bola veľmi podobná rozloženiu sociálnej siete. Napriek tomuto záveru, druhý výskum, ktorý vypracoval Wilson a kol. poskytuje zhodnotenie, že sieť aktivít na sociálnom portáli Facebook je značne odlišná od rozloženia sociálnej siete [7]. Tento fakt je spôsobený predovšetkým prirodzenými obmedzeniami, ako je napríklad limitovaný čas, ktorý používateľ strávi v sociálnom priestore. Na základe tohto faktu môžeme očakávať, že budú existovať používatelia, pre ktorých nebudú existovať žiadne spoločné aktivity. Z toho vyplýva, že aj celková intenzita vzťahu bude veľmi nízka alebo až nulová.

Výskum, ktorý má najbližšie zameranie k tejto práci, vypracoval Viswanath a kol. [8]. Autori opisujú metódu analýzy vzťahov používateľov v priestore sociálneho portálu Facebook na základe odkazov umiestnených na tzv. *stene*. Navrhnutá metóda má niekoľko výrazných obmedzení:

- a) na analýzu vzťahov používa len odkazy napísané na stene, ignoruje akékoľvek iné aktivity vykonané v priestore sociálnej siete (napr. zdieľanie fotografií, posielanie súkromných správ),
- b) zdrojové údaje boli obmedzené na skupinu používateľov len jediného územného regiónu (New Orleans),
- c) metóda pracovala len s verejne dostupnými profilmi (pričom štandardné nastavenie profilu je privátne),
- d) najväčším obmedzením je, že metóda nie je natoľko univerzálna, aby ju bolo možné aplikovať na akúkoľvek sociálnu sieť, prípadne spájať údaje viacerých sietí do jedného spoločného výsledku.

V tejto práci riešime uvedené obmedzenia.

Sociálne sieťovanie je jedno z najrýchlejšie rastúcich trendov v priestore Internetu. Existujú stovky sociálnych sietí, ktoré sú dostupné prostredníctvom webu. Niektoré majú len niekoľko stoviek používateľov, iné rádovo milióny. Počet ich používateľov a ich popularita v poslednom období rapídne rastie.

Niektoré sociálne siete sú zamerané predovšetkým na sociálne sieťovanie (napr. Facebook, Friendster, MySpace), kým ostatné majú sociálne sieťovanie len ako sekundárny cieľ (napr. YouTube, Flickr) [9]. Takéto sociálne siete môžu poskytovať rozličné primárne zamerania, napr.:

- a) mikroblogy,
- b) fotografie,
- c) videá a iné médiá.

V tab. 2.1 uvádzame viaceré sociálne portály a siete spolu s analýzou a porovnaním ich vlastností.

	Odkaz	Zameranie	Používatelia	Prístup k údajom sociálnej siete	Údaje používateľov
Facebook	http://www.facebook.com	všeobecné	viac ako 400 miliónov zaregistrovaných používateľov (31.03. 2010) [16], až 37% používateľov má viac ako 100 priateľov [7]	rozsiahle API, nízky počet oficiálnych klientských knižníc [17]	široké spektrum údajov, len dva typy explicitného vyjadrenia vzťahu (je priateľ/nie je priateľ)
MySpace	http://www.myspace.com	všeobecné	viac ako 125 miliónov zaregistrovaných používateľov	nevyhovujúce API (dostupné len prostredníctvom projektu OpenSocial)	
Twitter	http://www.twitter.com	mikroblogy		explicitné vyjadrenie štyroch typov vzťahov (vzťahy „nasledovania“ a ich kombinácie medzi dvoma používateľmi)	nejednoznačne interpretovateľné údaje
Orkut	http://www.orkut.com	všeobecné	používateľská komunita predovšetkým v Južnej Amerike, len 20% používateľov má viac ako 100 priateľov [7]		široké spektrum údajov
Flickr	http://www.flickr.com	fotografie	sociálnych vzťahov nadväzuje len úzke jadro používateľov [10]	rozsiahle API	jednoznačne interpretovateľné údaje, slabá sociálna interakcia
Diskusia na portáli SME	http://www.sme.sk	spravodajstvo	obmedzenie predovšetkým na slovensky hovoriacich používateľov	dostupnosť údajov a logov v prostredí fakulty, pričom tieto zdroje nie sú všeobecne dostupné [11]	slabá sociálna interakcia
Emailová komunikácia		súkromná komunikácia	emailovú komunikáciu má k dispozícii väčšina používateľov	efektívny prístup k údajom prostredníctvom protokolov	nejednoznačne interpretovateľné údaje

Tab. 2.1 Porovnanie existujúcich sociálnych sietí

V analýze existujúcich sociálnych sietí sme sa zamerali bližšie na sociálnu sieť Facebook.

Facebook

Sociálnu sieť Facebook založil v roku 2004 študent Harvardskej univerzity Mark Zuckerberg. Pôvodne bola určená len pre študentov Harvardskej univerzity, neskôr sa do nej začali zapájať ďalšie univerzity a od roku 2006 je otvorená pre všetkých používateľov starších ako 13 rokov. V januári 2009 sa podľa štúdie Compete.com stal Facebook najpoužívanejšou sociálnou sieťou na svete [18].

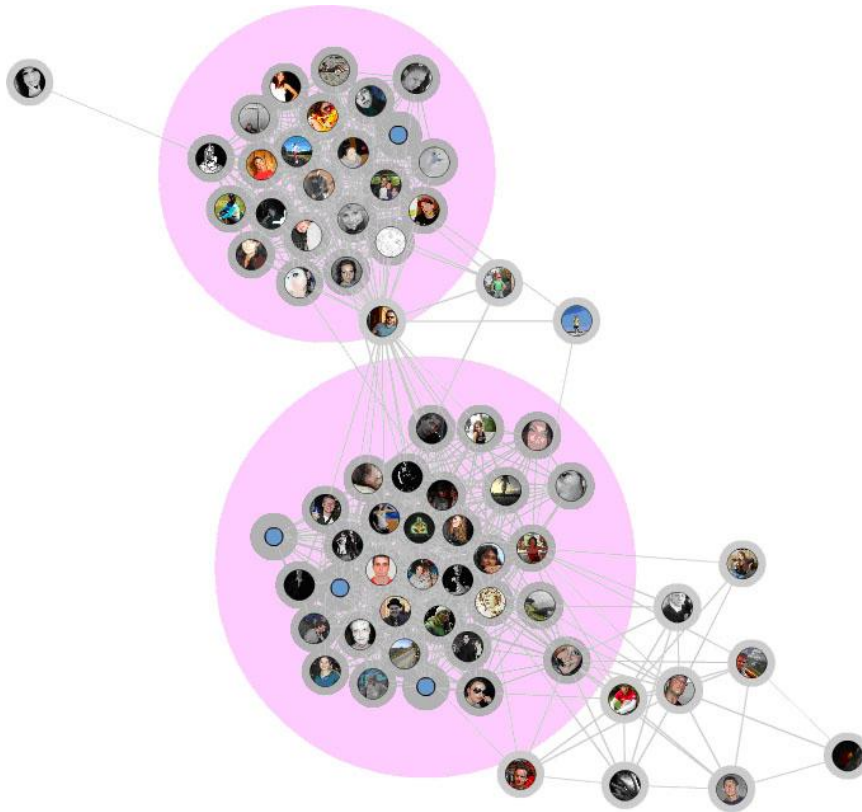
Facebook umožňuje používateľom vyplniť osobný profil, ktorý zahŕňa základné informácie ako meno, priezvisko, dátum narodenia. Taktiež má používateľ možnosť doplniť širokú paletu nepovinných informácií – štúdium, osobné záujmy, kontaktné informácie atď. Každý profil zahŕňa priestor pre písanie odkazov, tzv. *stenu* (*Wall*), ktorá je jedným z primárnych mechanizmov asynchrónnej výmeny správ medzi používateľmi. Okrem toho Facebook poskytuje služby pre zdieľanie fotografií, označovanie osôb na fotografiách (tzv. tagovanie), komentovanie používateľských aktivít, rozposielanie pozvánok na akcie, atď. Všetky tieto údaje môžu byť použité ako hodnotiace faktory intenzity vzťahu.

2.4 Aplikácie pre analýzu sociálnych sietí

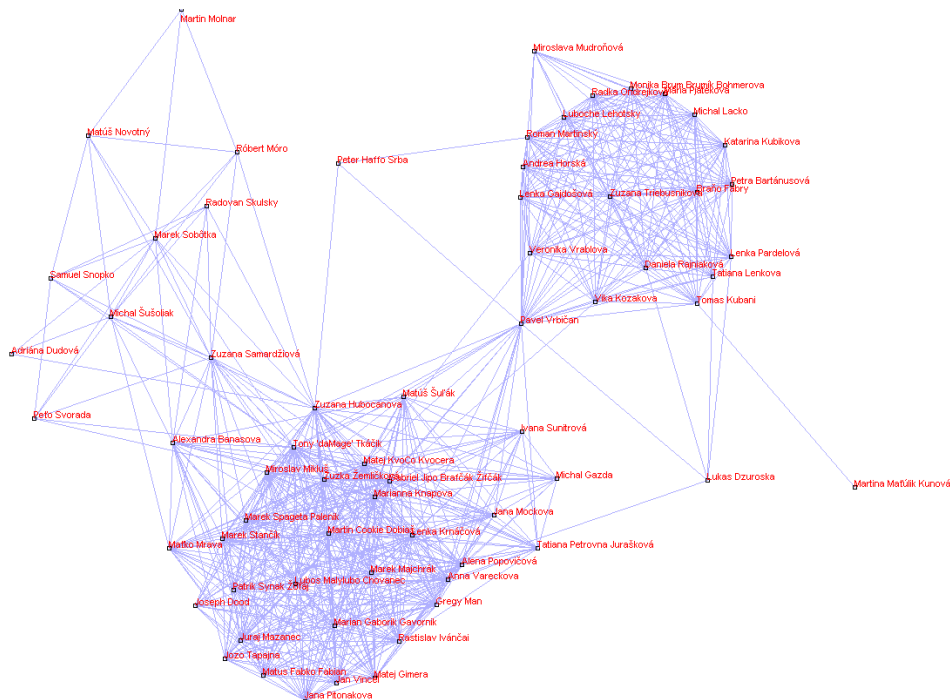
Pri analýze existujúcich riešení sa nám nepodarilo nájsť aplikáciu, ktorá by poskytovala analýzu intenzity vzťahov v čase. Väčšina aplikácií analýzy sociálnych sietí sa zameriava na analýzu sociálnych sietí ako celkov. Využíva techniky analýz celej siete alebo metódu snehovej gule, pričom vyhodnocuje všeobecné štatistické údaje ako vzájomná vzdialenosť uzlov, veľkosť, hustota a stupeň siete, reciprocita, tranzitivita atď. Tieto údaje sú však nezaujímavé z pohľadu ohodnotenia intenzity vzťahu medzi dvoma konkrétnymi uzlami. Príkladom takýchto aplikácií sú univerzitné výskumné projekty ako UCINET (<http://www.analytictech.com/ucinet/>) alebo Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>).

Analyzované aplikácie používajú binárne hodnotenie vzťahov (aj v dôsledku toho, že väčšina grafových algoritmov pracuje len s binárnym ohodnotením vzťahov). Binárne ohodnotenie však neposkytuje informácie o intenzite daného vzťahu. Príkladom takýchto aplikácií sú SocialGraph a FriendGraph, ktoré pracujú s dátami dostupnými zo sociálnej siete Facebook. Obe aplikácie používajú rovnakú techniku získavania dát – egocentrický prístup s pridanými vzťahmi.

Majú podobný výstup formou grafu (obr. 2.3 a obr. 2.4), v ktorom prezentujú vzájomné vzťahy medzi používateľmi. Aplikácia SocialGraph poskytuje aj dodatočnú funkcionálnu identifikáciu blízkych, silne naviazaných skupín používateľov, ktoré sú na výstupe zvýraznené kruhovým vyfarbeným pozadím. Hlavnou nevýhodou je práve binárne ohodnotenie vzťahov.



Obr. 2.3 Výstup aplikácie SocialGraph (<http://apps.facebook.com/socgraph/>)



Obr. 2.4 Výstup aplikácie FriendGraph (<http://apps.facebook.com/friendgraph/>)

2.5 Zhodnotenie analýzy sociálnych sietí

Stúpajúci nárast používateľov sociálnych sietí je sprevádzaný záujmom o spracovávanie údajov a vlastností sociálnych sietí. V tejto oblasti existuje veľa metód a techník, ktorých použitie závisí od konkrétneho cieľa výskumu. V analýze uvedenej v tejto kapitole uvádzame len tie, ktoré sú relevantné k problematike analýzy intenzity vzťahov. Veľké množstvo výskumov a aplikácií pre sociálne siete len potvrdzuje rastúci záujem o spoznávanie sociálnych sietí. Motiváciou a dôvodom tohto záujmu je nielen rastúci počet používateľov, ale predovšetkým aj čoraz väčšia intenzita s akou využívajú funkcionality sociálnych sietí a rozsah osobných informácií, ktorý používatelia publikujú.

Publikovaním svojich osobných informácií nám používatelia dávajú možnosť analyzovať ich spoločné vzťahy a záujmy, pomocou ktorých dokážeme vypočítať intenzitu vzťahov tak, ako je našim cieľom v tejto práci. Len málo používateľov si uvedomuje citlivosť údajov, ktoré každý deň dopĺňajú do priestoru sociálnych sietí. A to nielen svojich osobných údajov, ale aj údajov o svojich priateľoch a vzťahoch. Na jednej strane nám tento fakt pomáha pri analýze intenzity vzťahov alebo všeobecne pri získavaní dát pre potreby výskumu, na druhej strane musíme pri výskume garantovať dodržanie súkromia používateľov, ktorí sa zapoja do experimentov.

3 Metóda pre analýzu intenzity vzťahov

Metóda, ktorú sme navrhli sa primárne zaoberá výpočtom sily intenzity vzťahu. Tento výpočet umožňuje aj sledovanie intenzity vzťahu v čase. Na vykonanie výpočtu potrebujeme zabezpečiť prístup k údajom sociálnych sietí, predspracovať tieto údaje a realizovať samotný výpočet intenzity vzťahov.

3.1 Zabezpečenie prístupu k údajom sociálnych sietí

Zbieranie dát

Pre potreby navrhnutej metódy netreba získavať dáta, ktoré sú nevyhnutné pre štatistické analýzy celej siete vzhľadom na to, že táto oblasť nie je predmetom tejto práce. Pre aproximáciu vzťahov je postačujúce sledovanie jednotlivca voči svojim priamym susedom a následne opísanie vzájomných vzťahov medzi identifikovanými susedmi. Vhodná bude teda metóda egocentrického získavania dát s pridanými vzťahmi, pokiaľ bude možné a vhodné v danej sociálnej sieti sledovať aj pridané vzťahy, inak použijeme metódu egocentrického zbierania dát (len ego).

Ohodnotenie vzťahov

Pri ohodnotení vzťahov budeme vychádzať zo siete aktivít, pričom toto ohodnotenie má dva rozmery. Na jednej strane potrebujeme vyjadriť čiastkovú intenzitu vzťahu na základe jednej elementárnej aktivity. Príkladom takejto aktivity je poslanie správy alebo zverejnenie spoločnej fotografie. Pre takýto druh aktivity zavedieme pojem *hodnotiaci faktor*. Vlastnosti hodnotiacich faktorov sú bližšie analyzované v kapitole 3.2. Na druhej strane potrebujeme vyjadriť ohodnotenie celkovej intenzity vzťahu, ktoré bude výsledkom výpočtu zo všetkých hodnotiacich faktorov. Je evidentné, že bude nevyhnutné vyjadrovať hodnotiace faktory intervalom (napr. ohodnotiť intenzitu vzťahu na základe dĺžky posielaných správ). Nominálne ani ordinálne ohodnotenie nie je v tomto prípade postačujúce. Rovnako je jednoznačné, že ohodnotenie celkovej intenzity vzťahu musí byť vyjadrené prostredníctvom intervalu.

Aby bol zabezpečený prístup metódy k údajom sociálnej siete, musí táto sociálna sieť poskytovať metóde prístup k svojim údajom, resp. k tej časti údajov, ktorá je zviazaná s daným používateľom.

3.2 Spracovanie údajov sociálnych sietí

Po zabezpečení prístupu k údajom sociálnych sietí môžeme spracovať údaje. Spracovanie pozostáva z dolovania hodnotiacich faktorov vyjadrenia vzťahu.

Hodnotiaci faktor reprezentuje identifikovateľné a spočítateľné vyjadrenie vzťahu. Intenzitu vzťahu môže ovplyvňovať pozitívne alebo negatívne na základe toho, či reprezentuje pozitívne, alebo negatívne vyjadrenie vzťahu. Hodnotiaci faktor môže byť zdieľaný viacerými sociálnymi sieťami (*zdrojmi*), pričom pre každú sociálnu sieť môže mať rôznu *váhu*. Váha vyjadruje relatívnu veľkosť vplyvu na celkovú intenzitu vzťahu a bola získaná experimentálne.

Príklady hodnotiacich faktorov s experimentálne určenými váhami pre sociálnu sieť Facebook:

- výskyt na spoločnej fotografii (pozitívny, váha 0,29),
- štúdium na spoločnej strednej škole (pozitívny, váha 0,11),
- spoločný vzťah (pozitívny, váha 0,98).

Okrem váhy veľkosť vplyvu hodnotiaceho faktora na celkovú intenzitu vzťahu ovplyvňuje aj celkový počet výskytu daného hodnotiaceho faktora v celom priestore sociálnej siete (nielen vo vzťahu dvojice používateľov, ktorých intenzitu vzťahu metóda počíta). Zabezpečíme tým, že sa zohľadní aj frekvencia využívania danej sociálnej siete používateľom.

Rovnako treba pri výpočte čiastkovej intenzity vzťahu vyjadrenej z jedného hodnotiaceho faktora zohľadniť aj vplyv času. Na základe takejto závislosti klasifikujeme tri typy hodnotiacich faktorov:

Hodnotiaci faktor jednorazovej udalosti

Reprezentuje jednorazové vyjadrenie vzťahu, ktorého čas vieme určiť na základe údajov sociálnej siete. Rovnako vieme experimentálne určiť dobu, počas ktorej má daný hodnotiaci faktor vplyv na intenzitu vzťahu.

Príklady:

- Odoslanie textovej správy, pričom poznáme dátum a čas jej odoslania a doba vplyvu je experimentálne určená na 2 dni.
- Spoločná fotografia, pričom poznáme dátum a čas jej vytvorenia (prípadne zverejnenia na sociálnej sieti) a doba vplyvu je odhadovaná na 30 dní.

Zoznam všetkých identifikovaných hodnotiacich faktorov spolu s experimentálne stanovenými váhami a dobami vplyvu pre sociálnu sieť Facebook je uvedený v prílohe C.

Hodnotiaci faktor intervalovej udalosti

Reprezentuje vyjadrenie vzťahu platné počas časového intervalu, poznáme dátum a čas začiatku a konca tohto intervalu a rovnako, ako v predchádzajúcom prípade, aj doby vplyvu na intenzitu vzťahu.

Príklad:

- Štúdium na spoločnej škole od septembra 2006 do júna 2010, pričom doba vplyvu je 365 dní.

Hodnotiaci faktor neohraničenej udalosti

Reprezentuje časovo neobmedzené vyjadrenie vzťahu, nevieme určiť ani dátum a čas začiatku, ani dátum a čas konca.

Príklady:

- spoločné záujmy,
- spoločné bydlisko.

Spracovanie hodnotiacich faktorov sa vykoná:

- a) pre každú zvolenú sociálnu sieť,
- b) ku každému zvolenému používateľovi, ku ktorému sledovaný používateľ explicitne potvrdil vzťah (technika egocentrického zbierania dát bez pridaných vzťahov opísaná v kapitole 2.2.1), a prípadne aj medzi všetkými kombináciami týchto používateľov (technika egocentrického zbierania dát s pridanými vzťahmi opísaná v kapitole 2.2.1),
- c) získaním hodnotiacich faktorov vyjadrenia vzťahu, ich celkových výskytov a prípadne aj príslušných dátumov a časov,
- d) a uložením týchto dát do databázy.

3.3 Výpočet intenzity vzťahov

Ako sme uviedli vyššie, čiastková intenzita hodnotiaceho faktora závisí od priradenej váhy, celkového počtu výskytu daného hodnotiaceho faktora a dátumu a času, v ktorom počítame intenzitu vzťahu:

$$I_f(k, j) = \frac{w_{kj} \sum_{i=1}^l f_t}{1 + \ln(1 + l_c)}$$

kde:

- I_f – čiastková intenzita hodnotiaceho faktora,
- w_{kj} – váha j -teho faktora pre k -ty zdroj,
- l – počet výskytov hodnotiaceho faktora v sledovanom vzťahu,
- l_c – celkový počet výskytov hodnotiaceho faktora v priestore sociálnej siete,
- f_t – funkcia vyjadrujúca vplyv času.

Uvedený vzťah bol odvodený na základe experimentovania a sledovania vypočítaných hodnôt pre konkrétnych používateľov v sociálnej sieti Facebook. Logaritmický vplyv celkového počtu výskytov daného hodnotiaceho faktora zabezpečí, že v prípade častého výskytu jedného konkrétneho hodnotiaceho faktora nestúpne nežiaduco jeho váha vzhľadom na ostatné hodnotiace faktory a ani nebude potlačená tak, ako keby bol tento vplyv lineárny.

Funkcia vyjadrujúca vplyv času f_t závisí od typu hodnotiaceho faktora.

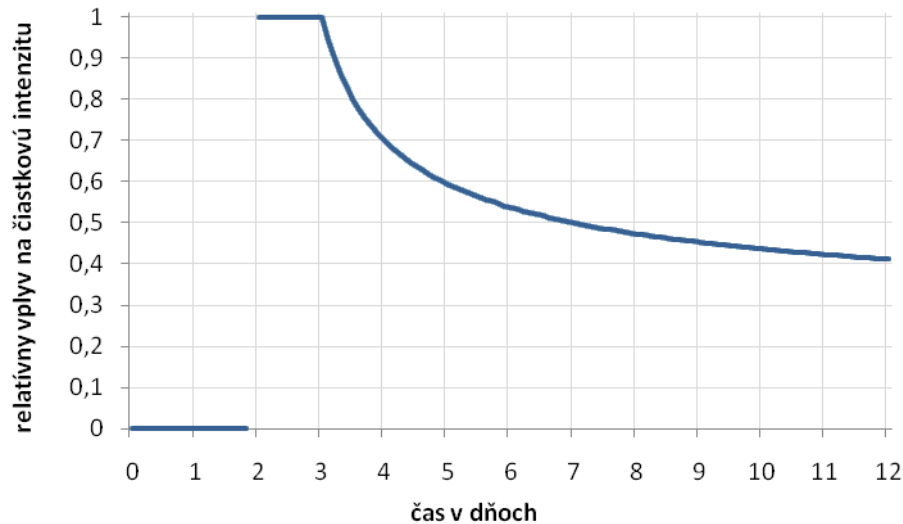
a) Funkcia f_t pre hodnotiaci faktor jednorazovej udalosti má tvar:

$$f_t(t_{kj}, t_m, t_v) = \begin{cases} 1 / (1 + \log_{\sqrt{t_{kj}}}(\max(1, (t_m - t_v)))) & t_m \geq t_v \\ 0 & t_m < t_v \end{cases}$$

kde:

- t_{kj} – doba vplyvu v dňoch, odmocnina t_{kj} vyjadruje dobu, za ktorú logaritmicky klesne intenzita vzťahu na polovicu,
- t_m – čas, pre ktorý meriame intenzitu,
- t_v – čas jednorazovej udalosti,
- $t_m - t_v$ – časový rozdiel v dňoch medzi časom, pre ktorý meriame intenzitu a časom jednorazovej udalosti.

Obr. 3.1 zobrazuje vplyv časovej funkcie jednorazovej udalosti na čiastkovú intenzitu. V zobrazenom prípade aktivita začala v druhý deň, doba vplyvu je nastavená na 25 dní. V siedmom dni môžeme vidieť, že časová funkcia má hodnotu 0,5.



Obr. 3.1 Príklad funkcie vplyvu času pri jednorazovej udalosti

b) Funkcia f_i pre hodnotiaci faktor neohraničenej udalosti nadobúda tvar:

$$f_i = \frac{1}{2}$$

Potom vieme zjednodušiť výpočet čiastkovej intenzity hodnotiaceho faktora na tvar:

$$I_f(k, j) = \frac{w_{kj} l}{2(1 + \ln(1 + l_c))}$$

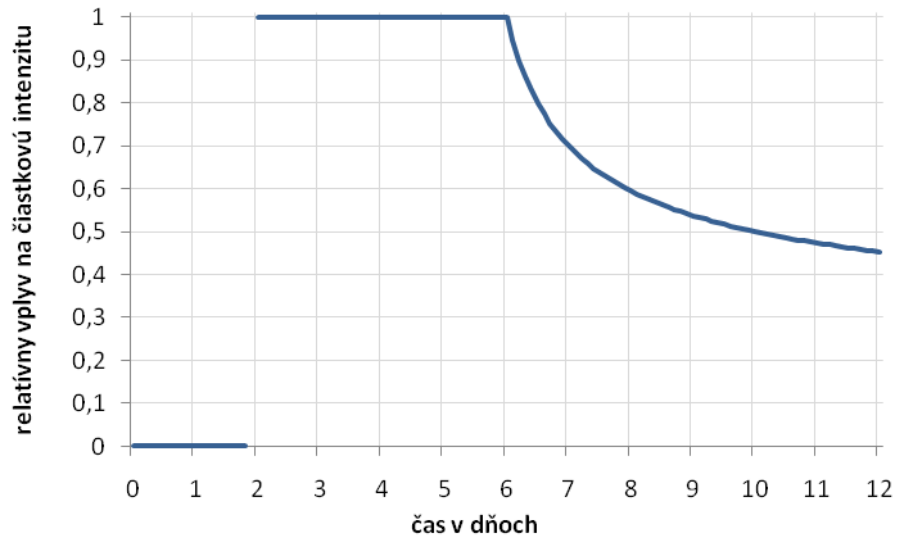
c) Funkcia f_i pre hodnotiaci faktor intervalovej udalosti nadobúda tvar, ktorý je čiastočne upravenou kombináciou predchádzajúcich dvoch tvarov:

$$f_i(t_{kj}, t_m, t_z, t_k) = \begin{cases} 0 & t_m < t_z \\ 1 & t_m \in \langle t_z, t_k \rangle \\ 1/(1 + \log_{\sqrt{t_{kj}}}(\max(1, (t_m - t_k)))) & t_m > t_k \end{cases}$$

kde:

- t_z – čas začiatku intervalovej udalosti,
- t_k – čas konca intervalovej udalosti.

Obr. 3.2 zobrazuje vplyv časovej funkcie intervalovej udalosti na čiastkovú intenzitu. V tomto zobrazenom prípade aktivita začala v druhý deň, skončila v piaty deň, doba vplyvu je nastavená na 25 dní.



Obr. 3.2 Príklad funkcie vplyvu času pri intervalovej udalosti

Použitie funkcie vyjadrujúcej vplyv času nám umožňuje na základe aktuálneho stavu používateľových vzťahov vypočítať intenzitu týchto vzťahov v ľubovoľnom časovom okamihu v minulosti. Nie je tak potrebné pravidelné dolovanie údajov v stanovených časových intervaloch. Taktiež je možné aproximovať intenzitu vzťahov v budúcnosti za predpokladu, že medzi časom dolovania hodnotiacich faktorov a časom, pre ktorý počítame intenzitu nenastanú žiadne nové výskyty hodnotiacich faktorov.

Vzhľadom na to, že jednotlivé hodnotiace faktory sú navzájom nezávislé, tak čiastková intenzita jedného zdroja je sumáciou čiastkových intenzít všetkých hodnotiacich faktorov:

$$I_z(k) = \sum_{j=1}^m I_f(k, j)$$

kde:

- I_z – čiastková intenzita zdroja,
- m – počet hodnotiacich faktorov.

V metóde uvažujeme aj nad výpočtom intenzity vzťahu dvoch používateľov z viacerých zdrojov. Jednotlivé zdroje poskytujú rôzne aspekty používateľových vzťahov, preto celkovú intenzitu vzťahu vypočítame ako aritmetický priemer čiastkových intenzít zo všetkých zdrojov:

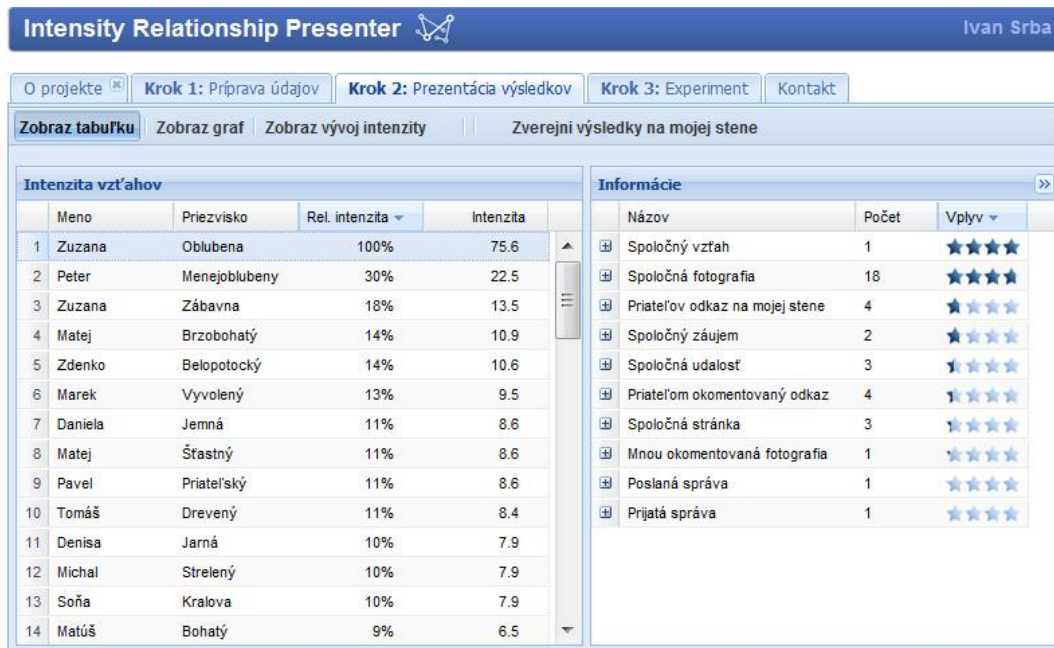
$$I = \frac{\sum_{k=1}^n I_z(k)}{n}$$

kde:

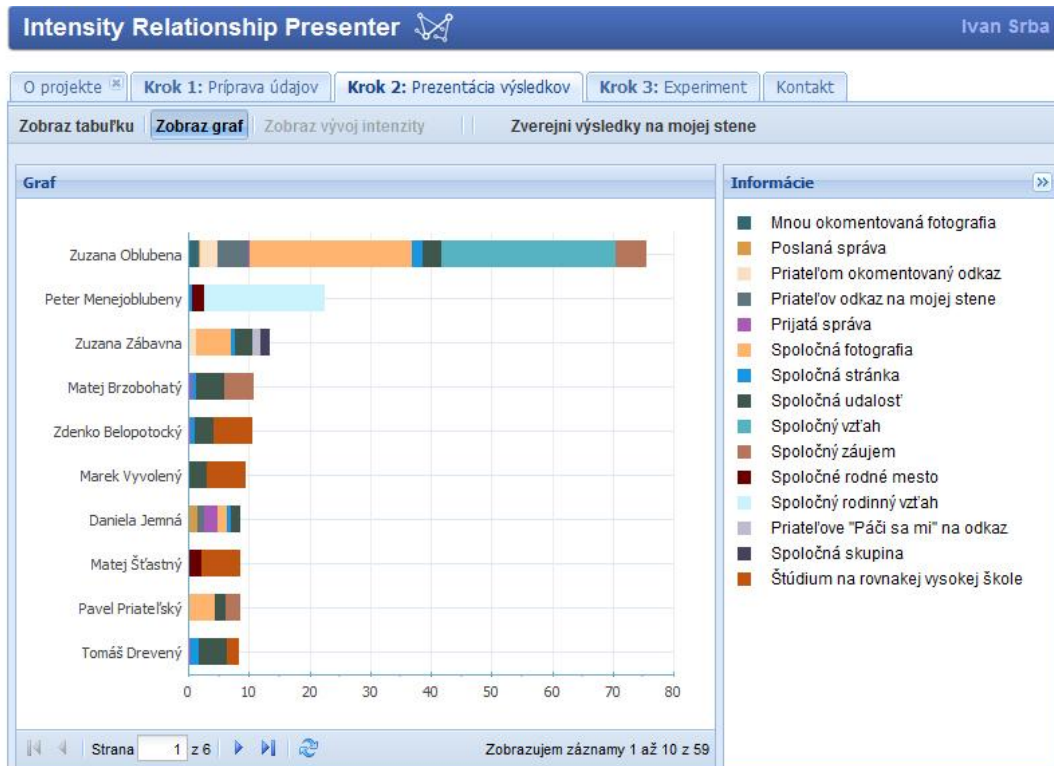
- I – celková intenzita vzťahu,
- n – počet zdrojov.

Pomocou metódy pre určenie intenzity vzťahov je možné vypočítať široké spektrum údajov. Predovšetkým je to intenzita vzťahov k rozličným používateľom a jej porovnanie. Takto

vypočítaná intenzita môže byť používateľovi prezentovaná vo forme tabuľky (obr. 3.3) alebo vo forme grafu (obr. 3.4).



Obr. 3.3 Prezentácia výsledkov vo forme prehľadovej tabuľky



Obr. 3.4 Prezentácia výsledkov vo forme grafu

Vzhľadom na to, že navrhnutá metóda dokáže vypočítať intenzitu v ľubovoľnom časovom okamihu, je možné používateľovi prezentovať aj vývoj intenzity v čase (obr. 3.5).



Obr. 3.5 Prezentácia výsledkov – vývoj intenzity vzťahu v čase

Na základe vedomosti o silnej alebo naopak slabej intenzite vzťahu dvoch používateľov môžeme vykonať akcie, ktoré už závisia od konkrétnej oblasti nasadenia. Ako príklad môžeme uviesť personalizované odporúčanie, kedy môžeme predpokladať, že ak istý objekt zaujal používateľa, ku ktorému máme silný vzťah, tak tento objekt môže byť zaujímavý aj pre nás. Inou aplikáciou je prispôsobovanie vzhladu alebo obsahu webových stránok na základe preferencií, alebo správania iného pre nás blízkeho používateľa, kolaboratívne filtrovanie [11], kolaboratívne vzdelávanie [13], alebo kolaboratívne vyhľadávanie [14].

V rôznych oblastiach nasadenia však sú zaujímavé len niektoré hodnotiace faktory, podľa ktorých sa určuje intenzitu vzťahu. Túto požiadavku dokážeme splniť jednoduchým výberom požadovaných zdrojov údajov a hodnotiacich faktorov.

4 Realizácia metódy pre analýzu intenzity vzťahov

Metódu, ktorú sme navrhli v kapitole 3, sme realizovali softvérovým nástrojom s názvom IRe. Na základe navrhnutej metódy pre výpočet intenzity vzťahov sme ho rozdelili do dvoch hlavných komponentov:

- a) Komponent s funkčnou logikou (jadrom metódy), ktorý sme pomenovali *Intensity Relationships Analyzer (IReA)*.
- b) Klientský komponent, ktorý sme pomenovali *Intensity Relationships Presenter (IReP)*.

4.1 Prístup k údajom sociálnych sietí

Navrhnutú metódu sme realizovali tak, aby dokázala aproximovať intenzitu vzťahov používateľa nad sociálnou sieťou, pre ktorú sú splnené nasledujúce kritéria:

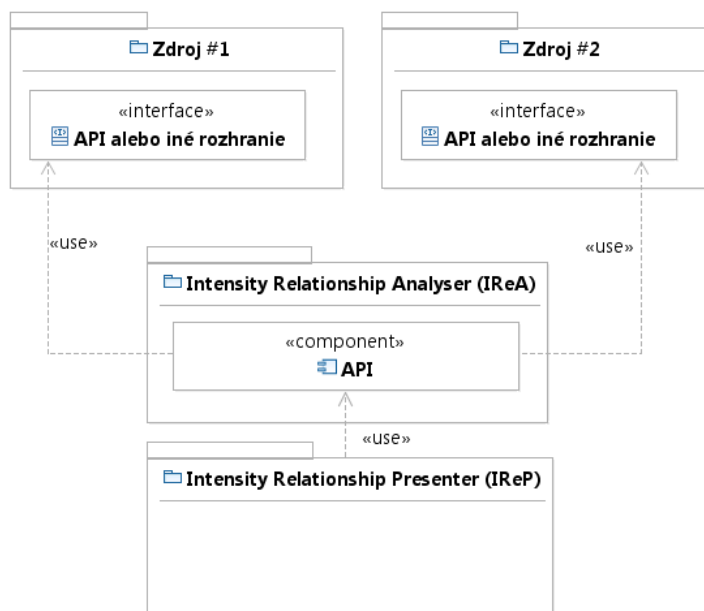
- Údaje sociálnej siete musia byť dostupné cez web.
- Musí existovať jednoznačná identifikácia používateľa, pre ktorého má metóda vyhodnocovať intenzitu vzťahov.
- Musí existovať informácia o tom, s akými používateľmi explicitne vyjadril alebo potvrdil svoje vzťahy sledovaný používateľ. Pre skupinu týchto používateľov zavedieme názov *priatelía*. Táto informácia nám umožní použiť egocentrické zbieranie dát.
- Sociálna sieť poskytuje svojim používateľom možnosti, ako môžu rozličnými spôsobmi vyjadrovať svoje sociálne vzťahy (napr. posielanie správ, písanie komentárov).
- Vzťahy musia byť viditeľné a skúmateľné. Dáta nemusia byť nevyhnutne verejne dostupné (viditeľné pre každého), ale musia byť dostupné po prihlásení daného používateľa.

Samotný prístup k údajom sociálnych sietí sa realizuje prostredníctvom komponentu IReA. Ten obsahuje rozhranie API, ktoré bude slúžiť nielen na vstup údajov zo sociálnych sietí, ale aj na výstup vypočítaných výsledkov pre klientský komponent IReP.

Priamy prístup k údajom

Ideálnym riešením pre realizáciu metódy by bola možnosť prepojiť prostredníctvom API komponent IReA priamo s ľubovoľným zdrojom dát tak, ako je to zobrazené na obr. 4.1. Na strane sociálnej siete by sa k zdroju dát pristupovalo prostredníctvom protokolu API alebo iného systému určeného na prístup k dátam daného zdroja.

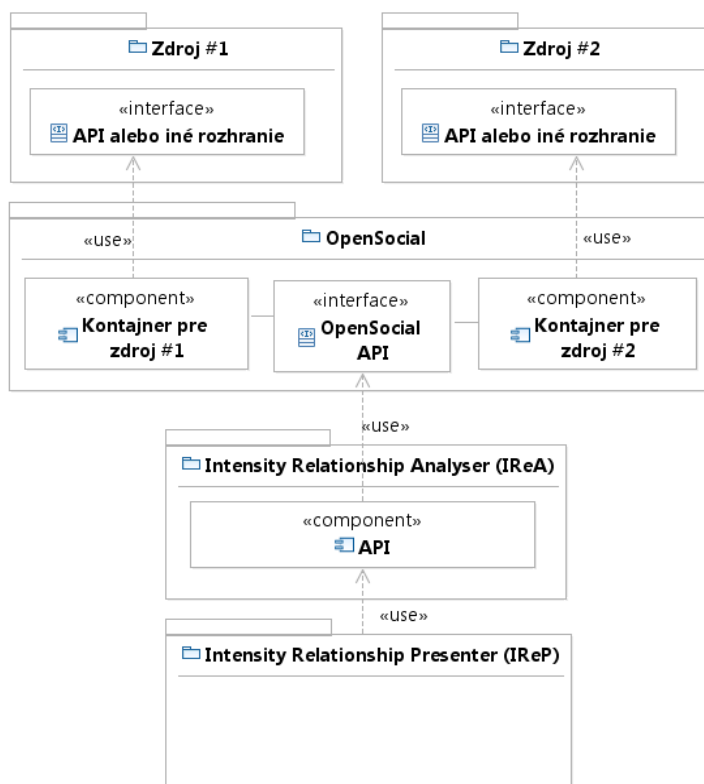
V súčasnosti neexistuje univerzálna architektúra a implementácia všetkých sociálnych sietí, ktorá by umožnila takýto priamy prístup k údajom. Z toho vyplýva, že ku každej sociálnej sieti musíme pristupovať prostredníctvom sprostredkovateľa alebo jednotlivo.



Obr. 4.1 Priamy prístup k údajom

Prístup prostredníctvom sprostredkovateľa

Druhým možným prístupom k údajom sociálnych sietí je integrovanie univerzálneho sprostredkovateľa, ktorý na jednej strane pozná rozhranie zdrojov a na druhej strane poskytuje uniformné rozhranie pre prístup k údajom všetkých zdrojov. V súčasnosti už existuje implementácia takéhoto sprostredkovateľa a je známa ako projekt OpenSocial [20]. Spojenie prostredníctvom sprostredkovateľa je zobrazené na obr. 4.2.

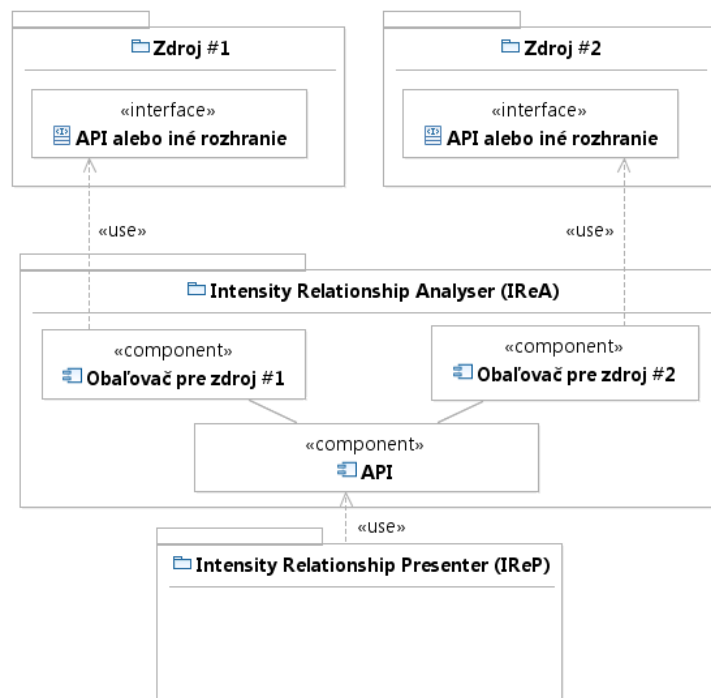


Obr. 4.2 Prístup prostredníctvom sprostredkovateľa

Takéto riešenie má však aj svoje nevýhody. Pre každý zdroj, ku ktorému potrebujeme pristupovať prostredníctvom OpenSocial, je potrebný tzv. *kontajner*. V čase riešenia projektu neboli dostupné kontajnery pre viacero významných sociálnych sietí (napr. Facebook) [21]. Okrem toho existujúce kontajnery neboli dostatočne spoľahlivé, napr. kontajner pre sociálnu sieť MySpace v kontrolných testoch uspel len v 57 bodoch z celkového počtu 161 [22]. Taktiež nie je možné pristupovať k špecifickým dátam danej sociálnej siete, ktoré môžu byť práve najhodnotnejšie pre aproximáciu intenzity vzťahov.

Prístup prostredníctvom vlastných obalovačov

Iným prístupom je integrovanie vlastného obalovača (wrapper) nad každým zdrojom, ku ktorému chceme pristupovať. Takýto prístup je zobrazený na obr. 4.3.



Obr. 4.3 Prístup prostredníctvom vlastných obalovačov

Riešenie prostredníctvom vlastných obalovačov má výhodu v tom, že dokážeme individuálne pristupovať ku každému zdroju a efektívne vyhľadávať aj dáta špecifické pre danú sociálnu sieť. Taktiež je možné pristupovať k dátam, ktoré sú uložené v rôznych formátoch, napr. logy, databázy atď. Nevýhodou je potreba implementovať obalovač pre každý nový zdroj a taktiež ho aktualizovať pri zmenách rozhrania sociálnej siete. Toto obmedzenie sa však môže v budúcnosti odstrániť implementovaním vlastného obalovača nad samotným rozhraním projektu OpenSocial.

Pre realizáciu metódy sme sa rozhodli použiť tretí prístup prostredníctvom vlastných obalovačov vzhľadom na to, že prvý priamy prístup je v súčasnosti nemožný a druhý prostredníctvom sprostredkovateľa OpenSocial nie je dostatočne spoľahlivý.

Ako už bolo uvedené, komponent IReA poskytuje API, ktoré slúži na vstup údajov do funkčnej logiky a taktiež na získanie výsledkov metódy. Ďalej poskytuje rad metód, pomocou ktorých je možné spravovať zdroje, hodnotiace faktory, používateľov a vydolované hodnotiace faktory.

Medzi najdôležitejšie metódy patria:

- prístup k zoznamu používateľov, ku ktorým má sledovaný používateľ vzťah,
- zadefinovanie a spustenie dolovania hodnotiacich faktorov,
- výpočet intenzity vzťahu medzi dvoma používateľmi v stanovený čas,
- vypočítanie vývoja intenzity vzťahu v stanovenom časovom období.

Metóda pre výpočet intenzity vzťahov je navrhnutá tak, že je možné použiť viaceré zdroje. V tom prípade je treba vedieť identifikovať rovnakých používateľov a uložiť vzájomné mapovanie jednotlivých identifikátorov sociálnych sietí. Túto požiadavku dokážeme zabezpečiť v závislosti od dostupných údajov. Pre identifikovanie rovnakej osoby tak môžeme použiť napr. emailovú adresu alebo kombináciu mena, priezviska a dátumu narodenia.

Z analýzy uvedenej v kapitole 2.3 sme sa rozhodli vybrať pre implementáciu softvérového nástroja sociálnu sieť Facebook. Dôvodom zvoleného výberu je dostupnosť veľkého množstva hodnotiacich faktorov a rozsiahla používateľská komunita. Na základe detailnej analýzy sociálnej siete Facebook sme vypracovali zoznam hodnotiacich faktorov, ktorý je dostupný v prílohe C.

Ako alternatívu k sociálnej sieti Facebook sme zvažovali emailovú komunikáciu. Emailová komunikácia nám však neposkytuje tak široké možnosti hodnotiacich faktorov, ako sociálna sieť Facebook a pri súkromnej komunikácii nemusí poskytovať úplne relevantné výsledky (napr. z dôvodu použitia iných komunikačných prostriedkov, ktoré nahrádzujú emailovú komunikáciu). No na druhej strane v prípade pracovnej komunikácie poskytuje jednoznačné ohodnotenie vzťahov, pričom pracovné vzťahy nie je možné zvyčajne ohodnocovať prostredníctvom sociálnej siete Facebook. Vzhľadom na to, že dodatočné implementovanie obalovača aj pre emailovú komunikáciu presahuje rozsah tohto projektu, bolo by jej vhodným rozšírením.

4.2 Štruktúra softvérového nástroja pre analýzu vzťahov IRe

Pri špecifikácii požiadaviek softvérového nástroja IRe zohľadňujeme funkcionálne a nefunkcionálne požiadavky uvedené v prílohe A. Softvérový nástroj bude všetkým používateľom sociálnej siete Facebook poskytovať možnosť zadefinovať a spustiť spracovanie údajov, prezentovať výsledky a vykonať experiment. Administrátorovi umožní zobrazit' prehľad údajov v systéme (priebeh spracovania údajov, používateľských účtov, správ a výsledkov experimentu). Detailnú analýzu funkcionálnych požiadaviek vo forme prípadov použitia uvádzame v prílohe B.

4.2.1 Architektúra softvérového nástroja

V softvérovom nástroji implementujeme komponenty IReA a IReP architektúrou klient – server, a to konkrétne takto:

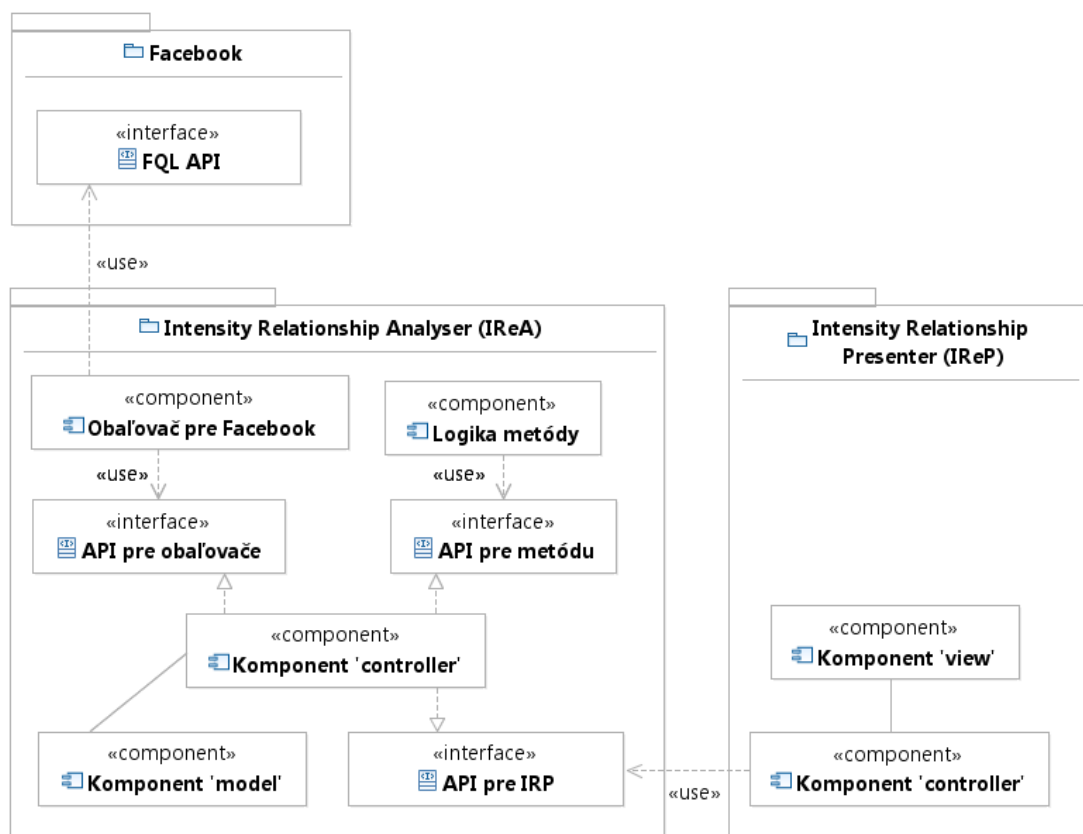
- a) Serverová aplikácia, ktorá bude implementovať IReA.
- b) Klientska aplikácia, ktorá bude implementovať IReP.

Schéma celkovej architektúry softvérového nástroja aj s výberom konkrétnej sociálnej siete je zobrazená na obr. 4.4. V implementácii IReA a IReP je použitý návrhový vzor MVC (Model – View – Controller).

Základom serverovej aplikácie IReA je komponent ovládačov, ktoré riadia celkové vykonávanie funkcionality. Tento komponent realizuje tri API. Prvé slúži pre klienta IReP a poskytuje metódy potrebné pre správu všetkých údajov systému. Druhé API poskytuje obalovačom metódy na

prístup k požiadavkám používateľa a k uloženiu výsledkov spracovania údajov sociálnych sietí. Tretie slúži funkčnej logike metódy na prístup ku spracovaným výsledkom pre potreby vypočítania intenzity vzťahov. Aplikácia IReA ďalej obsahuje komponent model, pomocou ktorého prístupuje k údajom databázy. Tento prístup nám zabezpečuje oddelenie závislosti aplikácie od použitej databázy do jediného komponentu.

Klient IReP sa skladá z komponentu pohľadov (View), ktorý zabezpečuje zobrazovanie GUI a komponentu ovládačov (Controller), ktorý na základe požiadaviek používateľa komunikuje so serverom IReA a výsledky požiadaviek zobrazuje prostredníctvom komponentu pohľadu používateľovi. Komunikácia medzi IReA a IReP je zabezpečená tým, že server IReA je implementovaný ako webová služba. Celkovo poskytuje IReA cez API pre IReP 57 metód. Opis metód, ako aj ich parametrov a návratových hodnôt, je zahrnutý v technickej dokumentácii IReA, ktorá je dostupná na priloženom médiu.



Obr. 4.4 Architektúra softvérového nástroja

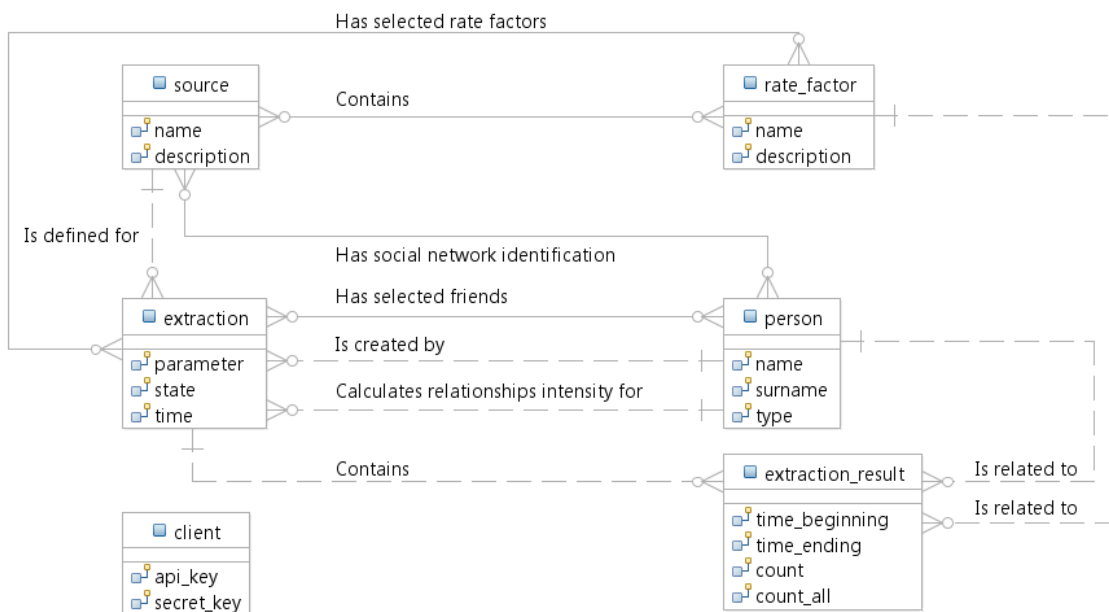
4.3 Štruktúra údajov v softvérovom nástroji IRe

V niektorých sociálnych sieťach nie je legálne ukladať získané údaje do externých databáz. V prípade sociálnej siete Facebook je nelegálne viac ako 24 hodín spracovávať dáta, pokiaľ tak nie je umožnené výnimkou alebo tieto údaje nie sú označené ako skladovateľné [19]. Z tohto dôvodu bude softvérový nástroj ukladať len identifikátory používateľov, skladovateľné identifikátory sociálnej siete, identifikátory hodnotiacich faktorov a počet ich výskytov, dátum a čas ich výskytov, dátum a čas dolovania.

Komponent model serverovej aplikácie IReA je zodpovedný za celkovú komunikáciu softvérového systému s databázou. Skladá sa z niekoľkých častí podľa hlavného objektu, s ktorým pracuje:

- Identifikácia klienta IReP.
- Zdroje (sociálne siete).
- Hodnotiace faktory.
- Osoby vystupujúce v systéme.
- Spracované údaje sociálnych sietí.

Logický model údajov je zobrazený na obr. 4.5. Fyzický model údajov s detailným opisom všetkých tabuliek spolu s informáciami o type a vlastnostiach každého atribútu je uvedený v prílohe D.



Obr. 4.5 Logický model údajov

Databáza obsahuje tabuľku *client* registrovaných klientov. Každému klientovi priradíme API kľúč a tajný kľúč. API kľúč slúži na identifikovanie klienta, ktorý prístupuje k webovej službe IReA. Tajný kľúč slúži na zašifrovanie prihlasovacích údajov.

4.4 Implementácia a testovanie

Výber programovacieho jazyka

Vzhľadom na to, že zadanie určuje webový charakter softvérového nástroja, pri výbere programovacieho jazyka sme sa zamerali na webové technológie. Ako vhodný programovací jazyk pre implementáciu sa javil Ruby on Rails. Pri výbere programovacieho jazyka sme však museli zahrnúť aj limitácie zo strany zvolenej sociálnej siete. Konkrétne v dobe návrhu softvérového nástroja a výberu technológií (november 2009) pre API sociálnej siete Facebook existovali len 4 oficiálne podporované knižnice a to PHP 5, JavaScript, Facebook Connect pre zariadenia iPhone a ActionScript 3.0 pre vytváranie flash aplikácií [23]. Okrem oficiálne podporovaných knižníc existovali aj neoficiálne verzie. Tieto knižnice však neboli dostatočne stabilné pre potreby našej

aplikácie z dôvodu, že sa nestíhali prispôbovať relatívne dynamicky vyvíjajúcemu sa API rozhraniu sociálnej siete Facebook.

Z uvedenej analýzy vyplýva, že pre naše potreby najviac vyhovuje programovací jazyk PHP 5. Základnú funkcionality jazyka PHP sme sa rozhodli rozšíriť o Zend framework, pomocou ktorého vieme efektívne vytvoriť aplikáciu so vzorom MVC. Vzhľadom na to, že PHP je interpretovaný programovací jazyk, rozhodli sme jeho jadro prepojiť s webovým serverom cez FastCGI protokol. Tým zabezpečíme, že webový server môže obsluhovať viacero požiadaviek naraz, čo je nevyhnutná podmienka pre fungovanie komponentu IReA, pretože obalovač pre sociálnu sieť Facebook bude spracovávať dáta pre každého používateľa v piatich vláknach.

Pre implementáciu používateľského grafického prostredia (GUI) sme vybrali programovací jazyk JavaScript s doplnením o ExtJS Framework. Vďaka použitiu tejto technológie bolo možné jednoducho vytvárať prehľadné tabuľky a grafy.

Vzhľadom na kombináciu programovacích jazykov PHP a JavaScript, sme mohli použiť pre zobrazovanie údajov v GUI technológiu AJAX. Na prenos údajov medzi GUI a PHP skriptom sme použili dátový formát JSON.

Technológie zabezpečujúce prístup k sociálnej sieti Facebook

K údajom sociálnej siete Facebook prístupujeme prostredníctvom poskytovaného API [17]. Pre použitie v našej aplikácii sú dostupné dve konkrétne technológie, a to prostredníctvom RESTFull API metód alebo prostredníctvom špeciálneho typu API s názvom FQL (Facebook Query Language), ktorý je obdobou databázového dopytovacieho jazyka SQL. FQL poskytuje jednoduchý a efektívny prístup k všetkým dostupným údajom sociálnej siete. Metódy API sú v skutočnosti len obalené FQL dotazy. Priame použitie FQL má výhodu aj v použití vnorených dotazov, tak ako ich poznáme z jazyka SQL.

Databáza

Vzhľadom na množstvo údajov, s ktorými serverová aplikácia IReA pracuje, je potrebná databáza. Pre implementáciu sme vybrali databázu MySQL. K databáze prístupujeme prostredníctvom knižnice typu PDO, takže v prípade potreby možno jednoducho a s minimálnymi zásahmi zmeniť typ databázy.

Zabezpečenie komunikácie

Aplikáciu IReA sme naprogramovali ako webovú službu. Komunikáciu medzi serverovou aplikáciou IReA a klientom IReP sme zabezpečili prostredníctvom protokolu SOAP. Poskytované rozhranie webovej služby IReA sme opísali jazykom WSDL.

Podporné systémy

Pri implementovaní aplikácie boli potrebné aj ďalšie podporné softvérové systémy. Pre archiváciu a manažment verzií zdrojových kódov sme použili systém SubVersion. Technickú dokumentáciu k aplikáciám IReA a IReP sme zabezpečili prostredníctvom nástroja phpDocumentator.

Používateľské rozhranie

Opis používateľského rozhrania je zahrnutý v opise používania aplikácie IRe, ktorý je dostupný v prílohe E.

Testovanie

Implementácia je dostupná ľubovoľnému používateľovi sociálnej siete Facebook po prihlásení sa do tejto sociálnej siete na adrese <http://apps.facebook.com/ira-irp/>.

Testovanie aplikácie sme vykonali na 7 rozličných používateľských účtoch sociálnej siete Facebook. Pre každý účet sme spracovali niekoľko variant údajov (rôzne kombinácie zvolených priateľov a hodnotiacich faktorov). Na základe týchto spracovaných údajov sme overili správnosť obalovača. Zároveň sme overili správnosť výpočtov metódy.

Na základe vypočítaných výsledkov sme stanovili váhy a doby vplyvu každého hodnotiaceho faktora tak, aby vypočítané intenzity vzťahov zodpovedali rovnakému usporiadaniu vzťahov podľa subjektívneho názoru daných používateľov.

5 Experimentálne overenie metódy analýzy intenzity vzťahov

5.1 Definovanie experimentu

Hlavným cieľom tejto práce je návrh metódy, ktorá ohodnocuje intenzitu vzťahov medzi používateľmi sociálnych sietí. Pre experimentálne overenie navrhutej metódy vyslovujeme tieto ciele:

- a) *Úspešnosť metódy*
 - Predpokladáme, že navrhnutá metóda dokáže vypočítať intenzitu k desiatim najlepším priateľom s úspešnosťou aspoň 80%. Predpokladáme, že 20-percentná chybovosť je ešte akceptovateľná.
- b) *Rozloženie vypočítanej intenzity medzi priateľmi*
 - Predpokladáme, že vypočítaná intenzita bude v rozložení interakcie medzi priateľmi poukazovať na rovnaké trendy, aké sú výsledkom práce [7].
- c) *Rozloženie vypočítanej intenzity v čase*
 - Predpokladáme, že vypočítaná intenzita bude mať rovnaký trend vývoja v čase, ako je opísaný vo výskumnej práci [8].

Pre overenie správnosti vypočítaných výsledkov navrhovaného riešenia použijeme softvérový nástroj IRe opísaný v kapitole 4, ktorý prehľadnou formou tabuliek a grafov zobrazí získané intenzity vzťahov. Táto aplikácia je dostupná priamo v prostredí rozhrania sociálnej siete Facebook a môže ju používať ľubovoľný používateľ tejto sociálnej siete.

Používateľ si môže vybrať zoznam priateľov a hodnotiacich faktorov, ktoré sa použijú pri získavaní informácií o vzťahoch a vytvorení údajovej sady. Na základe týchto získaných údajov vieme vyhodnotiť a overiť druhú a tretiu hypotézu.

Vzhľadom na to, že nedokážeme žiadnym spôsobom zistiť skutočnú intenzitu vzťahu v priestore sociálnej siete Facebook, táto aplikácia slúži aj na subjektívne overenie vypočítaných výsledkov. Používateľ môže poskytnúť explicitnú spätnú väzbu k získaným výsledkom, a tak ohodnotiť korektnosť a presnosť výsledkov. Používateľovi sa zobrazí usporiadaný zoznam priateľov podľa vypočítanej intenzity vzťahu v sociálnej sieti a zvýrazní sa prvých desať pozícií. Následne je vyzvaný, aby usporiadal prvých desať pozícií priateľov tak, aby ich poradie vyjadrovalo používateľove subjektívne vnímanie vzťahu vyjadrovaného v priestore sociálnej siete Facebook. Predpokladáme, že aspoň 80% priateľov z používateľovho upraveného zoznamu bude aj v pôvodnom zozname usporiadanom podľa vypočítanej intenzity. Predpokladáme, že časť priateľov bude presunutá, pretože predvolené všeobecné nastavenie váh a dôb vplyvu nemusí úplne zodpovedať aj preferenciám konkrétneho používateľa.

5.2 Vyhodnotenie experimentu

Experiment prebiehal od 26.03.2010 do 26.04.2010. Aplikáciu si spustilo 344 používateľov sociálnej siete Facebook. Z toho 238 používateľov si nechalo vypočítať aspoň raz intenzitu svojich vzťahov. Celkovo používatelia spustili 358 extrakcii údajov, v ktorých metóda vypočítala intenzitu pre 25 435 vzťahov.

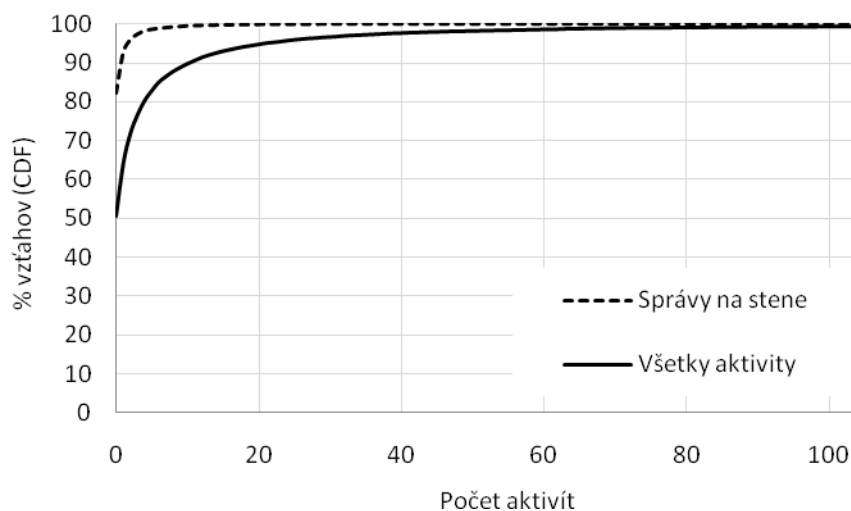
5.2.1 Úspešnosť metódy

V priebehu experimentu odoslali používatelia 95 usporiadaných zoznamov svojich priateľov. Výsledkom experimentu je, že viac ako 88% priateľov sa nachádzalo v oboch zoznamoch súčasne, čo predstavuje relatívne dobrý výsledok.

Dodatočne sme vyhodnotili nielen výskyt v oboch zoznamoch, ale aj ich preusporiadanie. Používatelia celkovo zmenili pozíciu 31% priateľov. To znamená, že zvyšných 69% priateľov vypočítaná intenzita zaradila aj na správnu pozíciu, čo považujeme vzhľadom na disproporcie medzi vzťahmi priateľstva a aktivít v sociálnej sieti Facebook za veľmi dobrý výsledok.

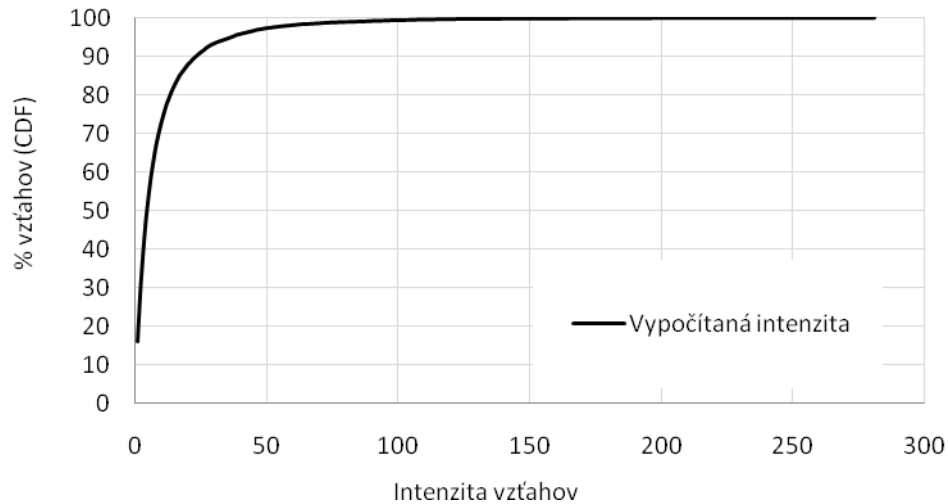
5.2.2 Rozloženie vypočítanej intenzity medzi priateľmi

Vo výskumnej práci [7] sa uvádza, že takmer všetci používatelia komunikujú len so 60% svojich priateľov. V práci [8] autori uvádzajú, že len 12% vzťahov obsahuje aj výmenu správ prostredníctvom steny. Pre porovnanie dát, ktoré sme získali v priebehu experimentu s týmito výsledkami, sme vypočítali kumulatívnu distribučnú funkciu (CDF - Cumulative distribution function) pre hodnotiace faktory vyjadrujúce napísanie odkazu na stenu a pre všetky hodnotiace faktory jednorazových aktivít (obr. 5.1). Z našich dát vyplýva, že až 82% vzťahov nemá žiadne spoločné odkazy na stene a až 50% vzťahov nemá žiadne aktivity. Z týchto výsledkov môžeme odvodiť zaujímavý záver, že väčšina aktivít vykonaných v sociálnej sieti Facebook sa vzťahuje na malú podmnožinu jej používateľov. Podobný záver sme očakávali a potvrdzujeme ním, že sieť aktivít je značne odlišná od sociálnej siete.



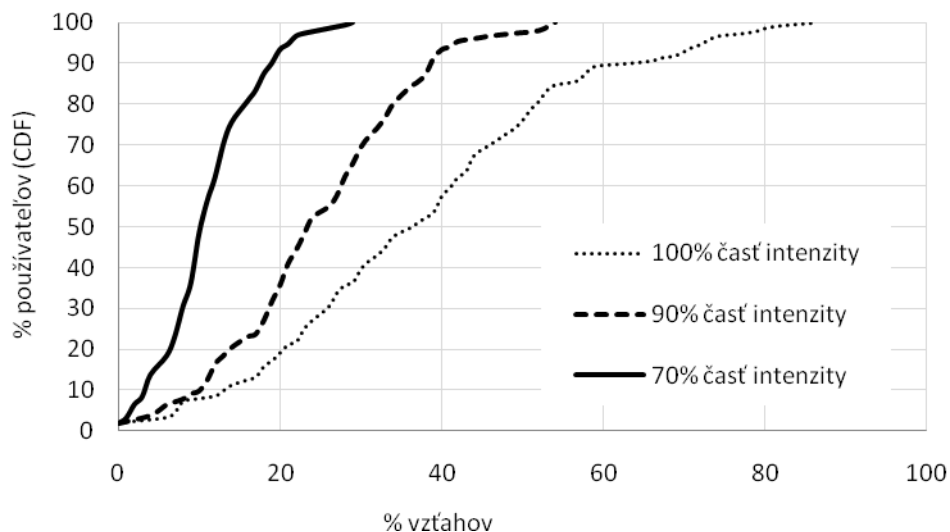
Obr. 5.1 Rozloženie aktivít medzi vzťahmi

Ak zobrazíme distribučnú kumulatívnu funkciu pre intenzitu vypočítanú z hodnotiacich faktorov jednorazových aktivít (obr. 5.2), tak môžeme vidieť, že je veľmi podobná distribučnej kumulatívnej funkcii pre rozloženie aktivít medzi používateľmi.



Obr. 5.2 Rozloženie intenzity medzi vzťahmi

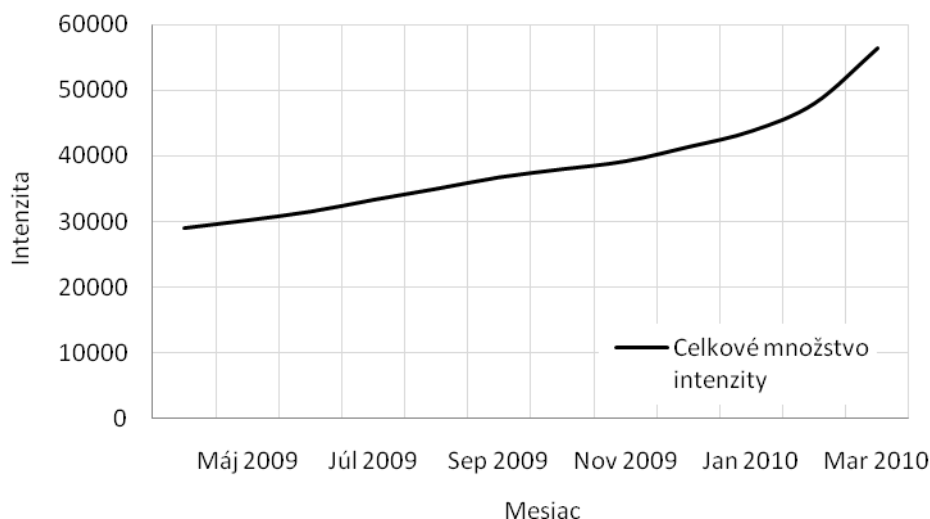
Pre porovnanie s prácou [7] sme vypracovali rozloženie čiastkovej intenzity. Pre každého používateľa sme zráali intenzitu jeho vzťahov a následne sme vypočítali tri hodnoty reprezentujúce časť celkovej intenzity. Konkrétne 100%, 90% a 70% z celkovej intenzity. Následne sme vypočítali koľko percent priateľov sa podieľa na týchto jednotlivých častiach intenzity. Celkový výsledok pre všetkých používateľov sme následne zhrnuli v troch distribučných kumulatívnych funkciách (obr. 5.3). Z grafu vyplýva viaceré zaujímavé výsledky, napr. celú časť z deväťdesiatich percent vypočítanej intenzity vytvorí v priemere len 54% vzťahov každého používateľa.



Obr. 5.3 Rozloženie rôznych častí intenzity medzi priateľmi

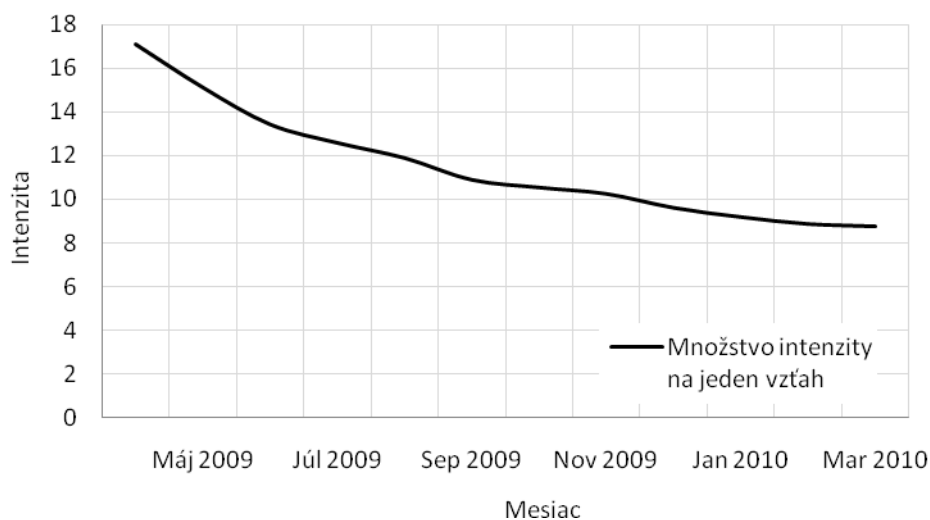
5.2.3 Rozloženie vypočítanej intenzity v čase

Pri analýze rozloženia vypočítanej intenzity v čase sme sa najprv zamerali na globálne výsledky. Na obr. 5.4 sme znázornili pribúdajúce množstvo intenzity všetkých vzťahov za posledných 12 mesiacov. Je evidentné, že s nárastom počtu používateľov musí pribúdať aj celkové množstvo intenzity vzťahov medzi používateľmi.



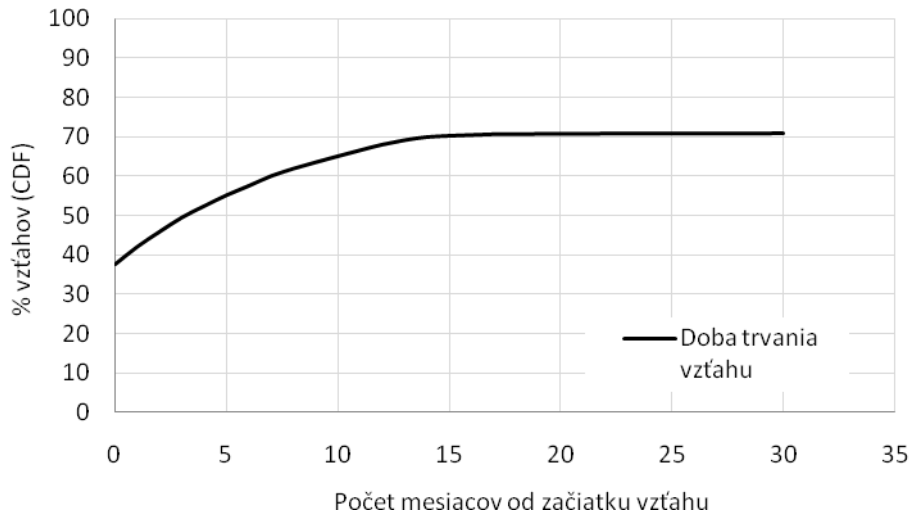
Obr. 5.4 Celkové množstvo intenzity vzťahov

Hodnotnejšiu informáciu nám však poskytuje vývoj priemerného množstva intenzity na jeden vzťah (obr. 5.5). Z grafu je jednoznačne vidno, že celková intenzita vzťahov nestúpa tak rýchlo, ako pribúdajú vzťahy a priemerná intenzita pripadajúca na jeden vzťah tak klesá. Tento trend v súčasnosti registruje viacero významných sociálnych sietí.



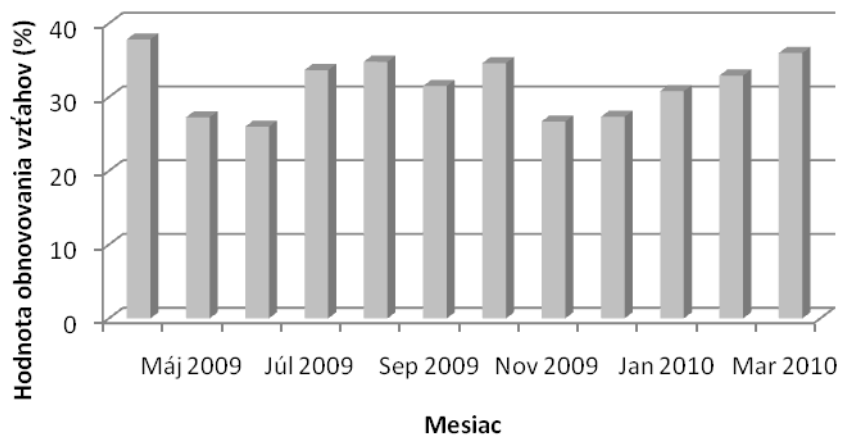
Obr. 5.5 Priemerné množstvo intenzity na jeden vzťah

Ďalej sme sa zamerali na analýzu doby trvania vzťahov. Pre vykonanie analýzy sme z našich údajov vypočítali kumulatívnu distribučnú funkciu pre dobu trvania vzťahu v mesiacoch (obr. 5.6). Za dobu trvania vzťahu považujeme dobu medzi prvým a posledným výskytom hodnotiaceho faktora jednorazovej udalosti. Výsledkom je, že po 12 mesiacoch od vytvorenia vzťahu až 70% všetkých vzťahov zanikne (nie sú už reprezentované žiadnou jednorazovou aktivitou).



Obr. 5.6 Kumulatívna distribučná funkcia pre dobu trvania vzťahu

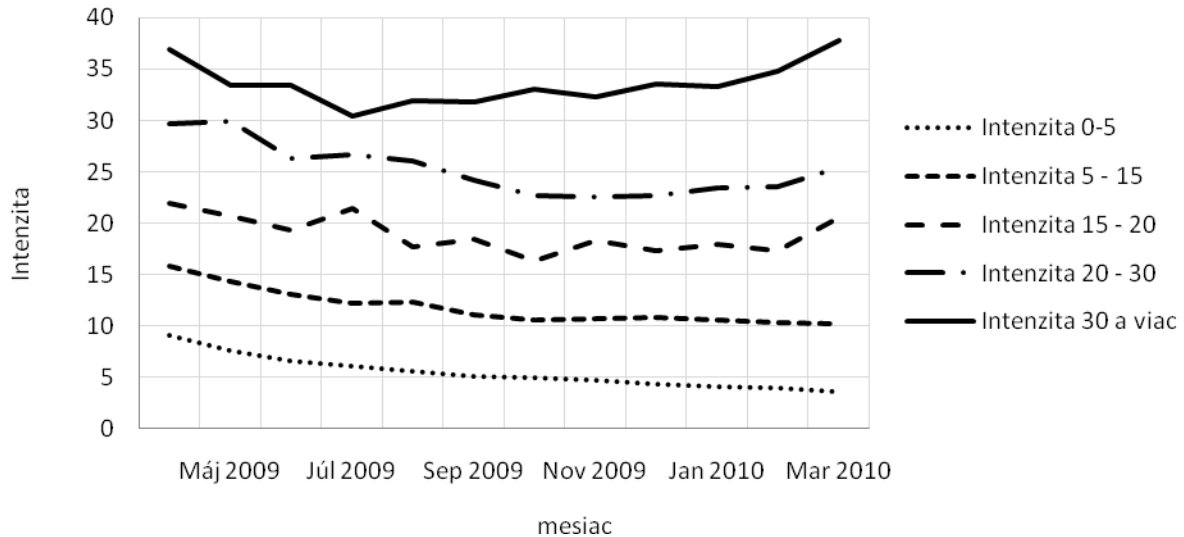
Na základe vysokého percenta zaniknutých vzťahov môžeme predpokladať, že vzťahy sa veľmi dynamicky menia v čase. Pre potvrdenie tejto hypotézy sme vypočítali pre posledných dvanásť mesiacov hodnotu obnovovania vzťahov. Pod hodnotou obnovovania vzťahov rozumieme časť takých vzťahov, ktoré boli v mesiaci n reprezentované aspoň jednou jednorazovou aktivitou a pokračovali vo svojej existencii aj v nasledujúcom mesiaci $n+1$. Hodnoty obnovovania vzťahov sú zobrazené v grafe na obrázku 5.7.



Obr. 5.7 Hodnota obnovovania vzťahov pre posledných 12 mesiacov

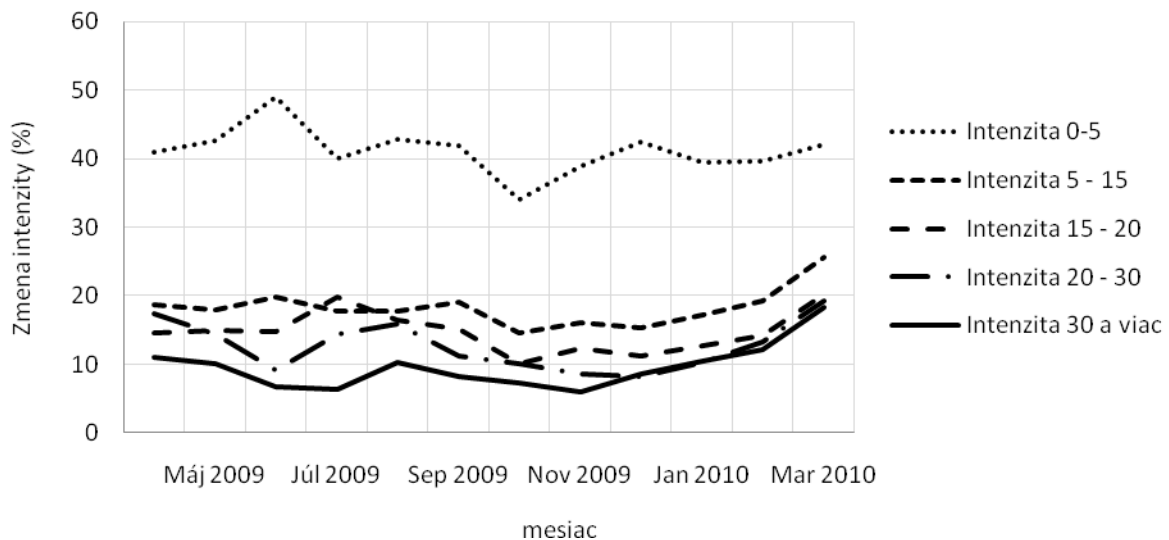
Priemerná hodnota obnovovania vzťahov je približne 31%, čo je veľmi podobná hodnota k 29%, ktorá bola zistená v práci [8].

Práca [8] uvádza, že intenzita komunikácie menej aktívnych používateľov klesá, pokiaľ intenzita komunikácie viac aktívnych používateľov zostáva na rovnakej úrovni. Pre overenie či aj vypočítaná intenzita spĺňa tento výsledok sme zostrojili graf na obrázku 5.8. Rozdelili sme vzťahy do piatich skupín podľa intenzity dosiahnutej v poslednom mesiaci. Ako je vidno, vzťahy s intenzitou menšou ako 15 skutočne slabnú. Na druhej strane vzťahy s intenzitou väčšou ako 15 zostávajú na rovnakej úrovni a vzťahy s intenzitou nad 30 dokonca mierne stúpajú.



Obr. 5.8 Priemerné množstvo intenzity na jeden vzťah pre rôzne silné vzťahy

Už sme ukázali, že priemerná hodnota obnovovania vzťahov je relatívne nízka (31%) a na jej základe môžeme vysloviť záver, že vzťahy sa v čase menia veľmi dynamicky. Rovnaký záver by mala poskytovať aj vypočítaná intenzita. Na obr. 5.9 sme zobrazili priemernú relatívnu zmenu intenzity pre rôzne silné vzťahy. Je evidentné, že veľmi slabé vzťahy s intenzitou menšou ako 5 dokážu zmeniť svoju intenzitu veľmi často a výrazne. Čím silnejšie vzťahy, tým sa ich hodnota v čase stabilizuje. Celkovo však môžeme potvrdiť, že vzťahy sa menia veľmi dynamicky.



Obr. 5.9 Priemerná zmena intenzity pre rôzne silné vzťahy

Celkové výsledky dosiahnuté v experimente považujeme za veľmi dobré, vzhľadom na to, že boli naplnené všetky stanovené ciele. Rozloženie intenzity medzi priateľmi, rovnako ako aj rozloženie intenzity v čase poukazuje na rovnaké trendy, aké sú v súčasnosti známe v sociálnych sieťach.

6 Zhodnotenie

V posledných rokoch sociálne siete získali nezastupiteľné miesto v priestore webu. Svoju pozíciu sa im darí aj naďalej upevňovať vzhľadom na rastúci počet používateľov a množstvo aktivít, ktoré sa v nich vykonávajú.

Sú prostriedkom pre výmenu informácií, komunikáciu, zdieľanie fotografií a video záznamov. V súčasnosti však v sociálnych sieťach začíname vnímať aj negatívne dôsledky ich prudkého rozmachu, napr. globálne oslabovanie vzťahov tak, ako sme na to poukázali aj v našej práci. V dôsledku všetkých pozitívnych a negatívnych dopadov sa vzťahy a ich intenzity menia veľmi dynamicky. Vtedy je výhodné mať možnosť sledovať vývoj a rozloženie intenzity vzťahov, čo bolo aj predmetom výskumu tejto práce.

Navrhnutá metóda pre aproximáciu intenzity vzťahov zahŕňa výpočet intenzity so širokým spektrom vplyvov, ktoré vieme v prostredí sociálnej siete sledovať. Túto metódu sme realizovali ako aplikáciu sociálnej siete Facebook a následne sme ju prostredníctvom tejto aplikácie aj overili. Výsledky experimentu nám poskytujú hodnotné informácie o úspešnosti metódy a rozložení intenzity vzťahov v čase a medzi priateľmi. Teda nielen metóda a jej overenie sú výsledkom tejto práce, ale aj štúdia vzťahov na základe experimentu v sociálnej sieti Facebook. Dôležitým záverom z týchto informácií je, že vypočítaná intenzita poukazuje na rovnaký trend vývoja vzťahov v sociálnej sieti Facebook, ako je opísaný v iných výskumných prácach.

Pri realizácii metódy sme narazili na niekoľko obmedzení zo strany sociálnej siete Facebook. Nebolo možné implementovať všetky navrhnuté hodnotiace faktory z dôvodu chýb v poskytovanom Facebook API. Z dôvodu chýbajúcich hodnotiacich faktorov tak nie sú vo vypočítanej intenzite zahrnuté všetky zrealizované aktivity, a tým sa vnášala do výsledkov istá nepresnosť. Ďalšou negatívnou stránkou realizácie je, že sociálna sieť Facebook neposkytuje dostatočné informácie o pracovných vzťahoch. Tento problém sa však dá vyriešiť zahrnutím údajov z ďalších sociálnych sietí, napr. emailovej komunikácie.

Práca poskytuje široké možnosti svojho rozšírenia či už na spomínanej úrovni doplnenia nových sociálnych sietí a ich hodnotiacich faktorov, alebo na strane poskytovanej funkcionality. Na základe získanej explicitnej spätnej väzby možno prepočítať predvolené nastavenia váh a dôb vplyvu a takouto personalizáciou získať ešte presnejšie výsledky. Tiež možno na základe rozloženia intenzity medzi priateľmi hlbšie analyzovať jednotlivé vzťahy alebo na základe vývoja intenzity v čase predpovedať ďalší vývoj vzťahu. Takýmto rozšírením by sme získali nové zaujímavé údaje, na ktoré by sa bolo možné pozeráť aj z profesionálneho hľadiska psychológie osobnosti. Iným možným smerom rozšírenia je integrovať vypočítanú intenzitu do už existujúcich aplikácií v sociálnej sieti Facebook, napr. v prehľade aktuálnych noviniek doplniť možnosť usporadúvať tieto údaje na základe intenzity vzťahu k príslušnému autorovi.

Literatúra

- [1] Kautz, H., Selman, B., Shah, M.: Referral Web : Combining Social Networks and Collaborative Filtering. In: *Communications of the ACM*, Vol. 40. No. 3, 1997.
- [2] Hanneman, R., Riddle, M.: *Introduction to social network methods* [online], Riverside (California): University of California, 2005. <http://faculty.ucr.edu/~hanneman/nettext> (09. 02. 2010).
- [3] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B.: Measurement and Analysis of Online Social Networks. In: *Proc. of the 7th ACM SIGCOMM conf. on Internet measurement*, San Diego (California): ACM New York, 2007, pp. 29 – 42.
- [4] Ahn, Y., Han, S., Kwak, H., Moon, S., Jeong, H.: Analysis of topological characteristics of huge online social networking services. In: *Proc. of the 16th int. conf. on WWW*, Alberta (Canada): ACM New York, 2007, pp. 835 - 844.
- [5] Maia, M., Alemida, J., Almeida, V.: Identifying user behavior in online social networks. In: *Proc. of the 1st workshop on Social network systems*, Glasgow (Scotland): ACM New York, 2008, pp. 1-6.
- [6] Chun, H., Ahn, Y., Eom, Y., Kwak, H., Moon, S., Jeong, H.: Comparison of online social relations in volume vs interaction: a case study of cyworld. In: *Proc. of the 8th ACM SIGCOMM conf. on Internet measurement*, Vouliagmeni (Greece): ACM New York, 2008, pp. 57-70.
- [7] Wilson, C., Boe, B., Sala, A.: User interactions in social networks and their implications. In: *Proc. of the 4th ACM European conf. on Computer systems*, Nuremberg (Germany): ACM New York, 2009, pp. 205-218.
- [8] Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in Facebook, In: *Proc. of the 2nd ACM workshop on Online social networks*, Barcelona (Spain): ACM New York, 2009, pp. 37-42.
- [9] Golbeck, J.: *The Dynamics of Web-Based Social Networks: Membership, Relationships and Change*, First Monday, 2007-11-05, Vol. 12, No. 11. <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2023/1889>.
- [10] Valafar, M., Rejaie, R., Willinger, W. : Beyond friendship graphs: a study of user interactions in Flickr. In: *Proc. of the 2nd ACM workshop on Online social networks*, Barcelona (Spain): ACM New York, 2009, pp. 25-30.
- [11] Zeleník, D., Bielíková, M.: Dynamika v hierarchickej klasifikácii článkov, In: *Student Research Conference 2010. Vol. 1. 6th Student Research Conference in Informatics and Information Technologies*, Bratislava (Slovakia): STU v Bratislave FIIT, 2010.
- [12] Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: *Proc. of the 2000 ACM conf. on Computer supported cooperative work*, Philadelphia (USA): ACM New York, 2000, pp. 241-250.
- [13] Tvarožek, J., Bielíková M.: The Friend: Socially-Intelligent Tutoring and Collaboration. In: *Frontiers in AI and Applications. Vol. 200 Artificial Intelligence in Education, AIED2009*, Amsterdam, IOS Press, pp. 763-764.
- [14] Tvarožek, M., Bielíková, M.: Collaborative Multi-Paradigm Exploratory Search. In: *Proc. of the Hypertext 2008 WebScience workshop*, Pittsburgh, PA, USA, ACM, New York, NY, pp. 29-33.

Internetové zdroje

- [15] Joshua P.: Relationship Symmetry in Social Networks: Why Facebook will go Fully Asymmetric [online]. <http://bokardo.com/p/1006> (31.03.2010).
- [16] Facebook, Inc. *Facebook Statistics* [online]. <http://www.facebook.com/press/info.php?statistics> (31.03.2010).
- [17] Facebook, Inc. *Facebook API* [online]. <http://wiki.developers.facebook.com/index.php/API> (31.03.2010).
- [18] Kazeniac, A.: *Social Networks: Facebook Takes Over Top Spot, Twitter Climbs* [online]. Compete.com. 2009. <http://blog.compete.com/2009/02/09/facebook-myspace-twitter-social-network> (31.03.2010).
- [19] Facebook, Inc. *Developer Principles and Policies* [online]. <http://developers.facebook.com/policy> (31.03.2010).
- [20] OpenSocial [online]. <http://www.opensocial.org> (31.03.2010).
- [21] OpenSocial [online]. <http://wiki.opensocial.org/index.php?title=Containers> (31.03.2010).
- [22] OpenSocial [online]. <http://opensocial-compliance.appspot.com> (31.03.2010).
- [23] Facebook, Inc. *Facebook API Client Libraries* [online]. http://wiki.developers.facebook.com/index.php/Client_Libraries (31.03.2010).

Použité skratky

API	A pplication P rogramming I nterface <ul style="list-style-type: none">- API je rozhranie, ktoré implementuje daný softvér, aby umožnil inej softvérovej aplikácii spoločnú interakciu.
FastCGI	F ast C ommon G ateway I nterface <ul style="list-style-type: none">- FastCGI je protokol spájajúci webové servery s interaktívnymi programami.
FQL	F acebook Q uery L anguage <ul style="list-style-type: none">- FQL je jedným z typov API, ktorý slúži na získavanie dát zo sociálnej siete Facebook prostredníctvom rozhrania podobnému s SQL.
GUI	G raphic U ser I nterface <ul style="list-style-type: none">- GUI je skratka pre grafické používateľské rozhranie. GUI je jedným z typov používateľskej interakcie, ktorá umožňuje používateľom pracovať s počítačovými programami pohodlnejšie ako prostredníctvom príkazového riadku.
JSON	J ava S cript O bject N otation <ul style="list-style-type: none">- JSON je textový formát zápisu dát, predovšetkým údajových štruktúr a asociatívnych polí (objektov).
PDO	P HP D ata O bject <ul style="list-style-type: none">- PDO je rozšírenie programovacieho jazyka PHP, ktoré definuje konzistentné rozhranie pre prístup k rozličným typom databáz.
PHP	P HP: H ypertext P reprocessor <ul style="list-style-type: none">- PHP je open source skriptovací programovací jazyk, ktorý sa používa najmä na programovanie klient-server aplikácií (na strane servera) a pre vývoj dynamických webových stránok.
SOAP	S imple O bject A ccess P rotocol <ul style="list-style-type: none">- Protokol slúžiaci na výmenu dátových objektov vo forme XML správ prostredníctvom HTTP protokolu.
SQL	S tructured Q uery L anguage <ul style="list-style-type: none">- SQL je počítačový jazyk na manipuláciu (DML) a definíciu dát (DDL). V súčasnosti je to najpoužívanejší jazyk tohto druhu v relačných systémoch riadenia bázy dát.
SVN	S ubversion <ul style="list-style-type: none">- Subversion je systém pre kontrolu verzií. Môže byť použitý pri správe historických a súčasných verzií súborov, ako napríklad zdrojových kódov, web stránok alebo dokumentácie.
WSDL	W eb S ervice D escription L anguage <ul style="list-style-type: none">- WSDL je jazyk, ktorý slúži na opis verejných rozhraní webových služieb. Pre svoj zápis využíva formát XML.

Príloha A – Požiadavky na softvérový nástroj

A.1 Požiadavky na komponent IReA

Funkcionálne požiadavky

Komponent IReA obsahuje funkčnú logiku metódy a prostredníctvom API obal'ovačom sprístupňuje metódy pre vstup údajov a pre IReP metódy na zadefinovanie požiadaviek na spracovanie údajov sociálnej siete, prezentáciu výsledkov a vykonanie experimentu.

Znovu použiteľnosť

Vzhľadom na to, že IReA obsahuje vlastné API, je možné metódy funkčnej logiky použiť v akejkoľvek externej aplikácii. Komponent IReP, ktorý je súčasťou softvérového nástroja je len ukážkou takejto aplikácie.

Modulárnosť

Komponent IReA je modulárny. Oddelené sú všetky funkcie, ktoré od seba navzájom funkčne nezávisia.

Interoperabilita

Komponent IReA je priamo naviazaný na sociálne siete ako externé zdroje údajov.

Modifikovateľnosť

Správa zdrojov údajov, hodnotiacich faktorov a používateľov musí byť jednoduchá a efektívna. Pridanie novej sociálnej siete, resp. zmena existujúcej nesmie výrazne ovplyvniť zvyšok aplikácie IReA.

Práca s údajmi

Navrhovaná aplikácia pracuje so širokým spektrom údajov, ktorých množstvo môže prudko narastať s pridávaním nových sociálnych sietí a nových sledovaných používateľov. Tento fakt je potrebné akceptovať pri návrhu databázovej štruktúry.

Hardvérové požiadavky

Vzhľadom na to, že obal'ovače IReA pristupujú k zdrojom údajov prostredníctvom Internetu, vyžaduje sa rýchle a stabilné internetové pripojenie. Spracovanie údajov vyžaduje aj dostatočný výpočtový výkon procesora.

A.2 Požiadavky na komponent IReP

Funkcionálne požiadavky

Komponent IReP bude implementovaný ako webová aplikácia, ktorá prostredníctvom API využíva metódy komponentu IReA a následne poskytuje grafické používateľské rozhranie (GUI). Slúži na zobrazenie a overenie výsledkov.

Použitelnosť

Prezentácia údajov je v komponente IReP zabezpečená na viacerých úrovniach. Dáta sú prezentované používateľovi jednak v prehľadnej tabuľkovej forme a taktiež prostredníctvom grafov. Pri návrhu GUI je potrebné zohľadniť aj fakt, že k aplikácii sa bude dať pristupovať najmä prostredníctvom rozhrania sociálnej siete Facebook, kde sú limitujúce rozmery zobrazovanej plochy.

Softvérové požiadavky

Vzhľadom na to, že komponent IReP je implementovaný ako webová aplikácia, jej používanie vyžaduje internetový prehliadač.

Príloha B – Model prípadov použitia

Jednotlivé diagramy modelu prípadov použitia sú rozdelené podľa hlavnej funkcionality, ktorú poskytujú a to konkrétne:

- registrácia používateľov,
- spracovanie, prezentácia a vyhodnotenie údajov,
- zobrazenie prehľadu údajov v systéme.

Registrácia používateľov

Na obr. B.1 sa nachádza diagram prípadov použitia, ktoré zabezpečujú registráciu používateľov.



Obr. B.1 Diagram prípadov použitia – registrácia používateľov

Prípad použitia:	UC01 Registruj nového používateľa
Opis:	Registrácia nového používateľa, ktorý má záujem o používanie aplikácie.
Účastníci:	Používateľ, sociálna sieť Facebook, IReP
Kroky základného toku:	<ol style="list-style-type: none"> Používateľ otvorí URL s aplikáciou. IReP na základe identifikátora sociálnej siete Facebook vyhodnotí, že daný používateľ ešte nebol registrovaný. IReP zobrazí všeobecné informácie o potrebe autorizovať aplikáciu a poskytnúť jej povolenia pre prístup k údajom používateľa. Používateľ potvrdí svoj záujem o registráciu v aplikácii. Sociálna sieť Facebook zobrazí autentifikačný formulár. Používateľ autorizuje aplikáciu. Sociálna sieť Facebook oznámi IReP autorizáciu aplikácie. IReP vygeneruje automatické prihlasovacie meno a heslo a zavolá metódu pre vytvorenie nového používateľského konta Aktivuje sa prípad použitia UC02 Nastav povolenia. IReP presmeruje používateľa na zobrazenie všeobecných informácií o projekte Prípad použitia končí.
Alternatívny tok:	<p><u><i>AI IReP na základe identifikátora sociálnej siete Facebook vyhodnotí, že daný používateľ sa už v minulosti registroval</i></u></p> <p>Tok sa aktivuje namiesto kroku 2.</p> <ol style="list-style-type: none"> IReP vykoná automatické prihlásenie. Prípad použitia pokračuje krokom 10 základného toku.

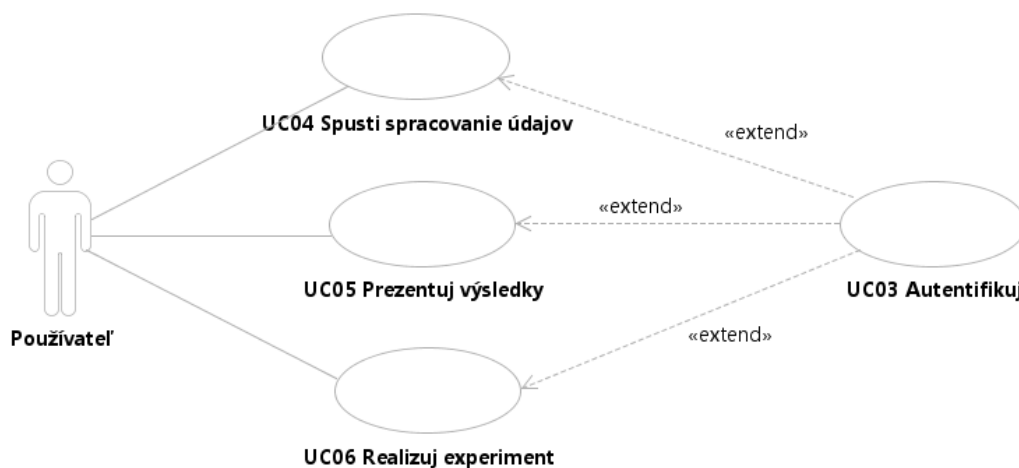
	<p><u>A2 Používateľ neautorizuje aplikáciu</u></p> <ol style="list-style-type: none"> 1. Tok sa aktivuje namiesto kroku 6. 2. Prípad použitia pokračuje krokom 3 základného toku.
--	---

Prípad použitia:	UC02 Nastav povolenia
Opis:	Používateľ nastaví povolenia umožňujúce pristupovať k používateľovým údajom.
Účastníci:	Používateľ, sociálna sieť Facebook, IReP
Kroky základného toku:	<ol style="list-style-type: none"> 1. Sociálna sieť Facebook zobrazí formuláre pre udelenie povolení. 2. Používateľ potvrdí udelenie všetkých povolení. 3. Prípad použitia končí.
Alternatívny tok:	<p><u>A1 Používateľ nepovolí niektoré z povolení</u></p> <p>Tok sa aktivuje namiesto kroku 2.</p> <ol style="list-style-type: none"> 1. Prípad použitia pokračuje krokom 1 základného toku.

Prípad použitia:	UC03 Autentifikuj
Opis:	IReP overí či má prihlásený používateľ právo vykonať danú akciu.
Účastníci:	Používateľ, IReP
Kroky základného toku:	<ol style="list-style-type: none"> 1. Používateľ požiada o vykonanie akcie. 2. Modul správy používateľov overí, že používateľ má oprávnenie na vykonanie danej akcie. 3. Prípad použitia končí.
Alternatívny tok:	<p><u>A1 Používateľ nemá oprávnenie na vykonanie danej akcie</u></p> <p>Tok sa aktivuje namiesto kroku 2.</p> <ol style="list-style-type: none"> 1. Modul správy používateľov zobrazí upozornenie. 2. Prípad použitia končí.

Spracovanie, prezentácia a vyhodnotenie údajov

Na obr. B.2 sa nachádza diagram prípadov použitia zabezpečujúci spracovanie, prezentáciu a vyhodnotenie údajov.



Obr. B.2 Diagram prípadov použitia - Spracovanie, prezentácia a vyhodnotenie údajov

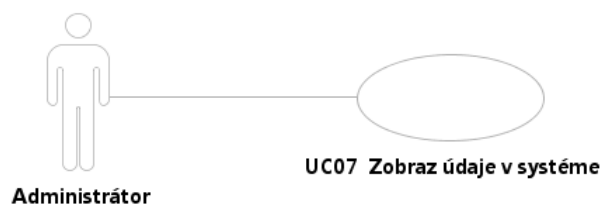
Prípad použitia:	UC04 Spusti spracovanie údajov
Opis:	Zadeinovanie požiadaviek a spracovanie údajov.
Účastníci:	Používateľ, IReP
Kroky základného toku:	<ol style="list-style-type: none"> 1. Používateľ sa prihlási prvýkrát alebo požiadava o vytvorenie nových údajov. 2. IReP zobrazí zoznam používateľových priateľov. Predvolene zvolí náhodne 90 priateľov. 3. Používateľ si zvolí priateľov, pre ktorých chce vypočítať intenzitu vzťahov. 4. Používateľ potvrdí svoj výber. 5. IReP zobrazí zoznam aktivít a údajov, ktoré sa majú použiť pri výpočte intenzity vzťahov. Predvolene zvolí všetky. 6. Používateľ si zvolí požadované aktivity a údaje. 7. Používateľ potvrdí svoj výber. 8. IReP zobrazí informáciu o prijatí požiadavky a spustí jej spracovanie. 9. IReP presmeruje používateľa na zoznam požiadaviek. 10. Prípad použitia končí.

Prípad použitia:	UC05 Prezentuj výsledky
Opis:	Prezentovanie výsledkov vo forme tabuľky, grafu alebo vývoja intenzity vzťahu v čase.
Účastníci:	Používateľ, IReP
Kroky základného toku:	<ol style="list-style-type: none"> 1. Používateľ požiadava o prezentáciu už spracovaných údajov. 2. IReP predvolene zobrazí prehľadovú tabuľku výsledkov. 3. Používateľ si zvolí zmenu spôsobu prezentovania výsledkov na jednu z troch dostupných foriem. 4. IReP zobrazí prezentáciu výsledkov vo forme požadovanej používateľom. 5. Prípad použitia končí.

Prípád použitia:	UC06 Realizuj experiment
Opis:	Vykonanie a odoslanie experimentu.
Účastníci:	Používateľ, IReP
Kroky základného toku:	<ol style="list-style-type: none"> 1. Používateľ požiada o zobrazenie experimentu. 2. IReP zobrazí výsledky v zjednodušenej tabuľke bez vypočítanej intenzity. 3. IReP zobrazí informácie o požadovanom experimente. 4. Používateľ vykoná experiment. 5. Používateľ odošle výsledok svojho experimentu. 6. IReP zobrazí informáciu o prijatí výsledku s poďakovaním. 7. Prípád použitia končí.

Zobrazenie prehľadu údajov v systéme

Na obr. B.3 sa nachádza diagram prípadov použitia zabezpečujúci zobrazenie prehľadu informácií v systéme.



Obr. B.3 Diagram prípadov použitia – spracovanie a vizualizácia údajov

Prípád použitia:	UC07 Zobraz údaje v systéme
Opis:	Spracovanie a zobrazenie údajov v systéme – štatistika využitia systému, zoznam zaregistrovaných používateľov, zoznam odoslaných správ, zoznam a vyhodnotenie prijatých výsledkov z experimentov.
Účastníci:	Administrátor, IReP
Kroky základného toku:	<ol style="list-style-type: none"> 1. Administrátor požiada o zobrazenie údajov v systéme. 2. IReP predvolene zobrazí tabuľku štatistík. 3. Administrátor si zvolí zmenu systémových údajov na jednu zo štyroch dostupných možností. 4. IReP zobrazí systémové údaje požadované administrátorom. 5. Prípád použitia končí.

Príloha C – Hodnotiace faktory pre sociálnu sieť Facebook

Zoznam hodnotiacich faktorov pre sociálnu sieť Facebook uvádzame v tabuľke tab. C.1. Pod pojmom *hyperodkaz* rozumieme internetový odkaz na zaujímavý objekt (napr. odkaz na video). Pod pojmom *odkaz na stene* rozumieme textovú správu napísanú na *stene* daného používateľa.

Názov hodnotiaceho faktora	Váha	Doba vplyvu	Typ	Poznámka
Spoločná fotografia	0.290	20	Jednorazový	
Mnou okomentovaná fotografia	0.120	7	Jednorazový	
Priateľom okomentovaná moja fotografia	0.100	7	Jednorazový	
Moje "Páči sa mi" na fotografiu	0.060		Neohraničený	
Priateľove "Páči sa mi" na fotografiu	0.050		Neohraničený	
Spoločné video	0.310	20	Jednorazový	
Spoločná udalosť	0.200	15	Jednorazový	
Spoločná stránka	0.030		Neohraničený	
Spoločná skupina	0.050		Neohraničený	
Spoločný hyperodkaz	0.200	10	Jednorazový	
Mnou okomentovaný hyperodkaz	0.120	7	Jednorazový	
Priateľom okomentovaný hyperodkaz	0.100	7	Jednorazový	
Moje "Páči sa mi" na hyperodkaz	0.080		Neohraničený	Nie je implementovaný
Priateľove "Páči sa mi" na hyperodkaz	0.070		Neohraničený	Nie je implementovaný
Môj odkaz na priateľovej stene	0.190	10	Jednorazový	
Priateľov odkaz na mojej stene	0.170	10	Jednorazový	
Moje "Páči sa mi" na odkaz	0.080		Neohraničený	
Priateľove "Páči sa mi" na odkaz	0.070		Neohraničený	
Mnou okomentovaný odkaz	0.130	7	Jednorazový	
Priateľom okomentovaný odkaz	0.110	7	Jednorazový	
Poslaná správa	0.060	10	Jednorazový	
Prijatá správa	0.050	10	Jednorazový	
Spoločné rodné mesto	0.070		Neohraničený	
Spoločné bydlisko	0.080		Neohraničený	
Spoločný rodinný vzťah	0.670		Neohraničený	
Spoločný vzťah	0.980		Neohraničený	
Spoločná aktivita	0.130		Neohraničený	
Spoločný záujem	0.140		Neohraničený	
Štúdium na rovnakej vysokej škole	0.110	365	Intervalový	
Štúdium na rovnakej strednej škole	0.140	250	Intervalový	
Spoločná práca	0.200	200	Intervalový	
Spoločná obľúbená hudba	0.090		Neohraničený	
Spoločný obľúbený televízny program	0.080		Neohraničený	
Spoločný obľúbený film	0.110		Neohraničený	
Spoločná obľúbená kniha	0.120		Neohraničený	
Spoločný priateľ	0.080		Neohraničený	Nie je implementovaný

Tab. C.1 Zoznam hodnotiacich faktorov pre sociálnu sieť Facebook

Pre sociálnu sieť Facebook sme identifikovali 36 významných hodnotiacich faktorov. Tri z nich nebolo možné implementovať z dôvodu chyby vo Facebook API alebo z dôvodu, že API neposkytovalo možnosť k daným údajom pristupovať.

Pri návrhu a implementácií sme neuvažovali nad hodnotiacimi faktormi, ktorých výskyt v priestore sociálnej siete je veľmi nízky (napr. pridávanie komentárov k videám), pretože ich spracovanie by predlžovalo extrakciu dát a neposkytlo by významnú výpovednú hodnotu o intenzite vzťahu.

Príloha D – Technická dokumentácia

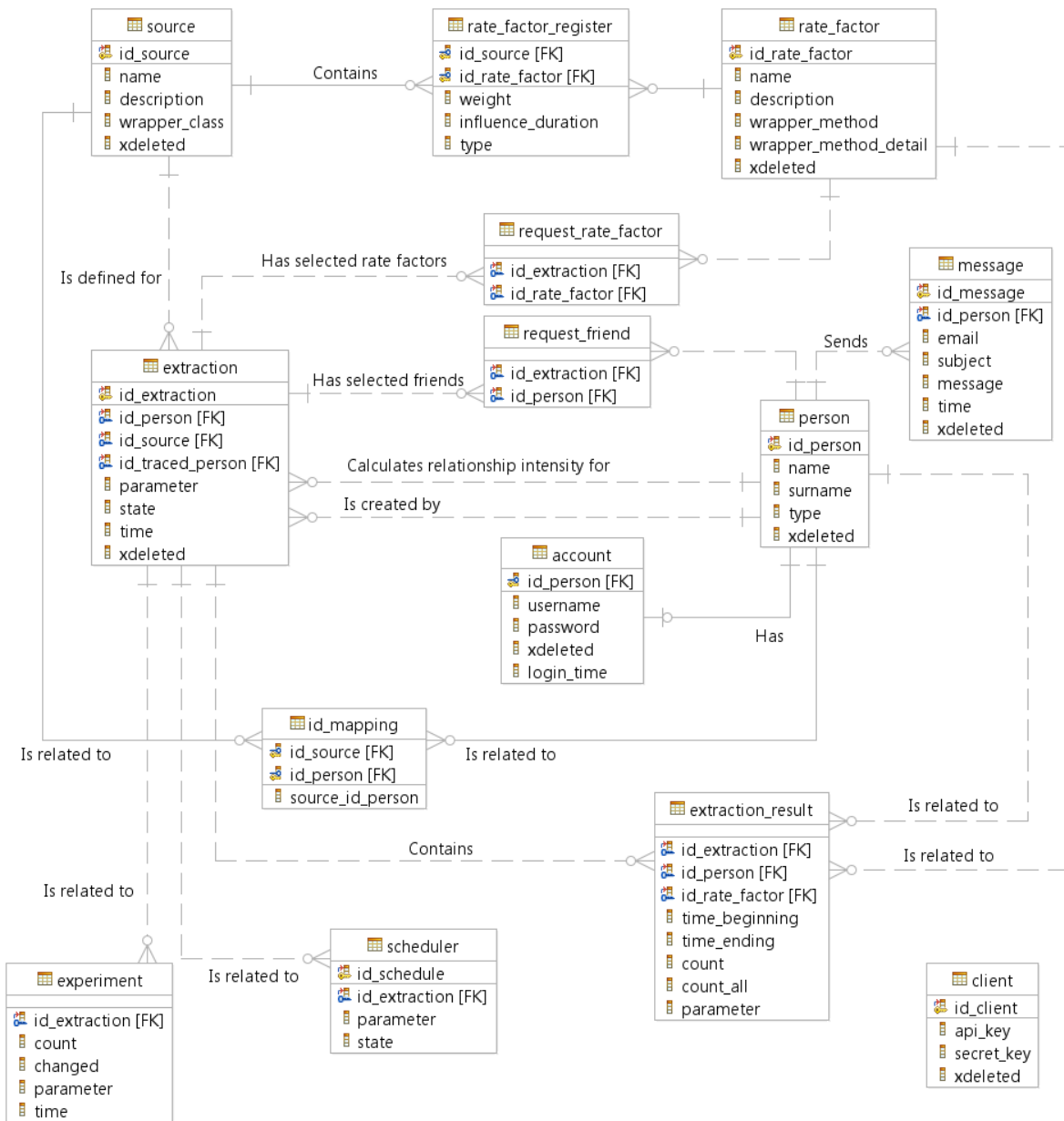
D.1 Diagram fyzického modelu údajov

Navrhnutá databáza obsahuje tabuľky, ktoré definujú:

- dostupné zdroje (*source*),
- dostupné hodnotiace faktory (*rate_factor*),
- mapovanie hodnotiacich faktorov na zdroje (*rate_factor_register*),
- všetky osoby, ktoré vystupujú v systéme (*person*),
- mapovanie identifikátorov sociálnych sietí na osoby vystupujúce v systéme (*id_mapping*),
- účty, ktoré majú registrovaní používatelia (*account*),
- požiadavky používateľov na spracovanie údajov sociálnych sietí (*extraction*),
- priateľov, pre ktorých sa majú údaje spracovať (*request_friend*),
- hodnotiace faktory, ktoré sa majú pri spracovaní použiť (*request_rate_factor*),
- vlákna, ktoré slúžia na spracovanie údajov sociálnych sietí (*scheduler*),
- výsledky extrakcií (*extraction_result*),
- výsledky experimentu (*experiment*),
- správy zaslané používateľmi cez kontaktný formulár (*message*).

Viacere tabuľky obsahujú atribút *xdeleted*. Tento atribút je typu logickej hodnoty a reprezentuje, či daný záznam bol vymazaný.

Fyzický model údajov je zobrazený na obr. D.1.



Obr. D.1 Fyzický model údajov

D.2 Štruktúry údajových entít fyzického modelu údajov

V nasledujúcich tabuľkách opisujúcich štruktúru údajových entít fyzického modelu údajov pre databázu MySQL používame tieto označenia:

- NN (Not Null) - obsahom poľa nemôže byť hodnota *null*
- PK (Primary Key) - primárny kľúč
- AI (Auto Increment) - automatické zvyšovanie hodnoty pri vložení nového záznamu
- FK (Foreign Key) - cudzí kľúč

Názov:	source
Opis:	obsahuje atribúty charakterizujúce dostupné zdroje (sociálne siete)

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_source	INT	Áno	Áno	Áno	Nie	Primárny kľúč
name	VARCHAR(64)	Áno	Nie	Nie	Nie	Názov zdroja
description	TEXT(256)	Nie	Nie	Nie	Nie	Opis zdroja
wrapper_class	VARCHAR(64)	Áno	Nie	Nie	Nie	Názov riadiacej metódy obalovača
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Zdroj je odstránený

Názov:	rate_factor
Opis:	obsahuje atribúty charakterizujúce hodnotiace faktory

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_rate_factor	INT	Áno	Áno	Áno	Nie	Primárny kľúč
name	VARCHAR(64)	Áno	Nie	Nie	Nie	Názov hodnotiaceho faktora
description	TEXT(256)	Nie	Nie	Nie	Nie	Opis hodnotiaceho faktora
wrapper_method	VARCHAR(64)	Áno	Nie	Nie	Nie	Názov metódy obalovača pre daný hodnotiaci faktor
wrapper_method_detail	VARCHAR(64)	Áno	Nie	Nie	Nie	Názov metódy obalovača pre zobrazenie detailných informácií pre hodnotiaci faktor
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Hodnotiaci faktor je odstránený

Názov:	rate_factor_register
Opis:	obsahuje atribúty charakterizujúce prepojenie zdroja a hodnotiaceho faktora

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_source	INT	Áno	Áno	Nie	Áno	Primárny kľúč, cudzí kľúč pre <i>source</i>
id_rate_factor	INT	Áno	Áno	Nie	Áno	Primárny kľúč, cudzí kľúč pre <i>rate_factor</i>
weight	DECIMAL(5, 3)	Áno	Nie	Nie	Nie	Váha hodnotiaceho faktora pre daný zdroj
influence_duration	INT	Áno	Nie	Nie	Nie	Doba vplyvu hodnotiaceho faktora pre daný zdroj
type	SMALLINT	Nie	Nie	Nie	Nie	Zdroj je odstránený

Názov:	person
Opis:	obsahuje atribúty charakterizujúce všetky osoby vystupujúce v systéme

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_person	INT	Áno	Áno	Áno	Nie	Primárny kľúč
name	VARCHAR(64)	Áno	Nie	Nie	Nie	Meno osoby
surname	VARCHAR(64)	Áno	Nie	Nie	Nie	Priezvisko osoby
type	INT	Áno	Nie	Nie	Nie	Typ osoby
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Osoba je odstránená

Názov:	account
Opis:	obsahuje atribúty charakterizujúce všetky osoby vystupujúce v systéme

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>
username	VARCHAR(32)	Áno	Nie	Nie	Nie	Prihlasovacie meno
password	VARCHAR(64)	Áno	Nie	Nie	Nie	Prihlasovacie heslo
login_time	DATETIME	Áno	Nie	Nie	Nie	Dátum posledného prihlásenia (poslednej aktivity)
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Účet je odstránený

Názov:	id_mapping
Opis:	obsahuje atribúty charakterizujúce prepojenie osôb a ich identifikátorov v sociálnych sieťach

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_source	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>source</i>
id_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>
source_id_person	VARCHAR(64)	Áno	Nie	Nie	Nie	Identifikátor osoby v sociálnej sieti

Názov:	extraction
Opis:	obsahuje atribúty charakterizujúce požiadavky na vytvorenie údajov

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_extraction	INT	Áno	Áno	Áno	Nie	Primárny kľúč
id_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>
id_source	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>source</i>
id_traced_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>
parameter	MEDIUMTEXT	Nie	Nie	Nie	Nie	Parameter potrebný pre spustenie extrakcie dát
state	VARCHAR(32)	Áno	Nie	Nie	Nie	Stav extrakcie
time	DATETIME	Áno	Nie	Nie	Nie	Čas zadanie požiadavky na extrakciu dát
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Extrakcia je odstránená

Názov:	request_rate_factor
Opis:	spájacia tabuľka charakterizujúca zoznam zadaných hodnotiacich faktorov v danej extrakcii

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_extraction	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>source</i>
id_rate_factor	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>rate_factor</i>

Názov:	request_friend
Opis:	spájacia tabuľka charakterizujúca zoznam zadaných priateľov v danej extrakcii

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_extraction	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>source</i>
id_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>

Názov:	extraction_result
Opis:	obsahuje atribúty charakterizujúce požiadavky na vytvorenie údajov

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_extraction	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>extraction</i>
id_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>
id_rate_factor	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>rate_factor</i>
time_beginning	DATETIME	Nie	Nie	Nie	Nie	Čas začiatku aktivity
time_ending	DATETIME	Nie	Nie	Nie	Nie	Čas ukončenia aktivity
count	INT	Nie	Nie	Nie	Nie	Počet výskytov hodnotiaceho faktora v skúmanom vzťahu
count_all	INT	Nie	Nie	Nie	Nie	Počet všetkých výskytov hodnotiaceho faktora
parameter	MEDIUMTEXT	Nie	Nie	Nie	Nie	Parameter obsahujúci identifikátory výskytov hodnotiaceho faktora

Názov:	scheduler
Opis:	obsahuje atribúty charakterizujúce správu vlákien zabezpečujúcich extrakciu dát

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_schedule	INT	Áno	Áno	Áno	Nie	Primárny kľúč
id_extraction	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>extraction</i>
parameter	MEDIUMTEXT	Nie	Nie	Nie	Nie	Parameter potrebný pre spustenie vlákna
state	VARCHAR(20)	Áno	Nie	Nie	Nie	Stav vlákna

Názov:	message
Opis:	obsahuje atribúty charakterizujúce správu poslanú používateľom cez kontaktný formulár

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_message	INT	Áno	Áno	Áno	Nie	Primárny kľúč
id_person	INT	Áno	Nie	Nie	Áno	Cudzí kľúč pre <i>person</i>
email	VARCHAR(64)	Nie	Nie	Nie	Nie	Emailová adresa používateľa
subject	VARCHAR(128)	Áno	Nie	Nie	Nie	Predmet správy
message	MEDIUMTEXT	Áno	Nie	Nie	Nie	Text správy
time	DATETIME	Áno	Nie	Nie	Nie	Čas odoslania správy
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Správa je odstránená

Názov:	client
Opis:	obsahuje atribúty charakterizujúce komponenty IReP, ktoré majú oprávnenie komunikovať s IReA

Názov	Dátový typ	NN	PK	AI	FK	Opis
id_client	INT	Áno	Áno	Áno	Nie	Primárny kľúč
api_key	VARCHAR(64)	Áno	Nie	Nie	Nie	API kľúč zabezpečujúci identifikovanie aplikácie (verejná informácia)
secret_key	VARCHAR(64)	Áno	Nie	Nie	Nie	Tajný kľúč pre šifrovanie prihlasovacích údajov (súkromná informácia)
xdeleted	BOOLEAN	Nie	Nie	Nie	Nie	Klient je odstránený

Číselníky atribútov fyzického modelu údajov

Názov:	Číselník hodnôt atribútu type údajovej entity person
Opis:	Atribút identifikuje typ osoby, na základe ktorého sa určujú prístupové práva k funkciám a dátam

Hodnota	Význam
1	Administrátor (plný prístup)
2	Používateľ, ktorý si zaregistroval aplikáciu alebo je priateľom takéhoto používateľa (obmedzený prístup)

Názov:	Číselník hodnôt atribútu type údajovej entity rate_factor_register
Opis:	Atribút identifikuje typ hodnotiaceho faktora na základe vplyvu času

Hodnota	Význam
1	Hodnotiaci faktor jednorazovej udalosti
2	Hodnotiaci faktor neohraničenej udalosti
3	Hodnotiaci faktor intervalovej udalosti

D.3 Ukážka zdrojového kódu

Zdrojový kód skriptu *index.php* zodpovedného za inicializáciu MVC aplikácie a nastavenie potrebných parametrov a ciest.

```
<?php
/**
 * index.php
 *
 * The file handling basic settings and starting application
 * @author Ivan Srba <srba07@student.fiit.stuba.sk>
 * @version 1.0
 */

error_reporting(E_ALL | E_STRICT);
ini_set("soap.wsdl_cache_enabled", "0");
date_default_timezone_set('Europe/Bratislava');

// Define path to application director
define('PUBLIC_PATH', dirname($_SERVER['SCRIPT_NAME']));
define('APPLICATION_PATH', realpath(dirname(dirname(__FILE__)) .
    './application'));
define('LIBRARY_PATH', realpath(dirname(dirname(__FILE__)) . './library'));

// Define application environment
define('APPLICATION_ENVIRONMENT', 'development');
//define('APPLICATION_ENVIRONMENT', 'production');

// PHP ini settings for development
if (APPLICATION_ENVIRONMENT == 'development') {
    ini_set('display_errors', 1);
}

// Ensure library/ is on include_path
set_include_path(implode(PATH_SEPARATOR, array(LIBRARY_PATH,
    APPLICATION_PATH . '/../', get_include_path(), )));

// Zend_Loader_Autoloader
require_once 'Zend/Loader/Autoloader.php';
$autoloader = Zend_Loader_Autoloader::getInstance();
$autoloader->registerNamespace(array('Application', 'Wrapper'));

// Zend_Config_Ini
$config = new Zend_Config_Ini(APPLICATION_PATH .
    '/config/application.ini', APPLICATION_ENVIRONMENT);
$registry = Zend_Registry::getInstance();
$registry->set('configuration', $config);

// Create and store database adapter
$db = Zend_Db::factory($config->database);
$registry->set('db', $db);
$db->query("SET NAMES 'UTF8'");

// Zend_Controller_Front
$frontController = Zend_Controller_Front::getInstance();
$frontController->setControllerDirectory(APPLICATION_PATH . '/controllers');
```

```

$frontController->setBaseUrl(PUBLIC_PATH . '/index.php');
$frontController->throwExceptions(true);
$frontController->setParam('noViewRenderer', true);
$frontController->dispatch();
?>

```

Ukážka zdrojového kódu z API poskytovaného prostredníctvom webovej služby pre IReP.

```

/**
 * Method addPerson creates new person with posted parameters
 * and returns id_person of inserted record
 *
 * @param string $name
 * @param string $surname
 * @param int $type
 * @param string $idSource
 * @param string $sourceIdPerson
 * @param string $username
 * @param string $password
 * @return int
 */

public function addPerson($name, $surname, $type = 2, $idSource,
    $sourceIdPerson, $username = null, $password = null)
{
    $registry = Zend_Registry::getInstance();
    if ($type == 1 && !$registry->isRegistered('loginTypePerson')) {
        throw new SOAPFault('401', 'Unauthorized: Login required');
    }
    if ($type == 1 && $registry->get('loginTypePerson') != 1) {
        throw new SOAPFault('400', 'Bad Request: You do not have permission to
            access this method with posted arguments');
    }

    $tmp = Application_Models_Person::getInstance();

    $idPerson = null;
    $idPerson = $tmp->getIdPerson($idSource, $sourceIdPerson);
    if ($idPerson == null) {
        $idPerson = $tmp->addPerson($name, $surname, $type);
        $tmp->addIdMapping($idSource, $idPerson, $sourceIdPerson);
    }

    if ($username != null && $password != null) {
        $tmp = Application_Models_Extraction::getInstance();
        $tmp->createResultsTable($idPerson);

        $tmp = Application_Models_Person::getInstance();
        $tmp->createAccount($idPerson, $username, $password);
    }

    return $idPerson;
}

```

```

/**
 * Method updatePerson updates selected person
 * with posted parameters
 *
 * @param int $idPerson
 * @param string $name
 * @param string $surname
 * @param int $type
 * @return int
 */
public function updatePerson($idPerson, $name, $surname, $type = 2)
{
    $registry = Zend_Registry::getInstance();
    if (!$registry->isRegistered('loginTypePerson')) {
        throw new SOAPFault('401', 'Unauthorized: Login required');
    }
    if ($type == 1 && $registry->get('loginTypePerson') != 1) {
        throw new SOAPFault('400', 'Bad Request: You do not have permission to
            access this method with posted arguments');
    }

    Application_Models_Person::checkAccessability($idPerson);

    $tmp = Application_Models_Person::getInstance();
    return $tmp->updatePerson($idPerson, $name, $surname, $type);
}

```


Príloha E – Opis použitia softvérového nástroja IRe

Pre používanie aplikácie IRe je potrebné prihlásenie do sociálnej siete Facebook. Aplikácia sa nachádza na adrese <http://apps.facebook.com/ira-irp>. Pri prvom spustení sa od používateľa vyžaduje autorizovať aplikáciu a udeliť jej potrebné povolenia. Po autorizovaní aplikácie a úplnom udelení povolení je používateľ presmerovaný na prvú záložku hlavného menu, ktorá prezentuje informácie o projekte (obr. E.1).

Intensity Relationship Presenter Ivan Srba [Odhliasiť]

O projekte **Krok 1: Príprava údajov** Krok 2: Prezentácia výsledkov Krok 3: Experiment Kontakt Administrácia

O projekte
Intensity Relationship Analyser & Presenter je univerzitný výskumný projekt, ktorého cieľom je vypočítať intenzitu vzťahov medzi používateľmi sociálnych sietí. Táto intenzita je výsledkom netriviálneho výpočtu na rozdiel od mnohých podobných aplikácií, ktoré vyberajú Vašich priateľov len náhodne.

Neváhajte sa dozvedieť, kto je Vaším najlepším priateľom alebo kto si na Vás už dávno nespomenul. Nechajte si vypočítať intenzitu vzťahov a zúčastnite sa, prosím, malého experimentu, vďaka ktorému môžu byť vypočítané výsledky ešte presnejšie.

O motivácii
V súčasnosti Top novinky (Top News) fungujú na základe filtrovania podľa počtu komentárov a typu novinky (fotografia, video, link, ...). Predstavte si, že by toto filtrovanie fungovalo na princípe intenzity Vašich vzťahov. Dokázalo by vybrať novinky od priateľov, s ktorými komunikujete najintenzívnejšie. A nielen to, dokázalo by Vám pomôcť udržiavať Vaše vzťahy aktívne. Priemerný používateľ v sociálnej sieti Facebook má približne 130 priateľov a časom o mnohých z týchto priateľov prichádza práve z dôvodu nerovnomerného rozloženia svojich aktivít. Na základe vypočítanej intenzity Vašich vzťahov by Vás vedel Facebook upozorniť, s ktorým priateľom by ste si mohli "vylepšiť" svoj vzťah.

O autorovi
Ivan Srba je študentom tretieho ročníka Fakulty informatiky a informačných technológií na Slovenskej technickej univerzite v Bratislave. Intensity Relationship Analyser & Presenter je súčasťou bakalárskeho projektu vypracovaného pod vedením prof. Ing. Márie Bielikovej, PhD.

Obr. E.1 Úvodná obrazovka s informáciami o projekte

Práca s aplikáciou sa skladá z troch základných krokov:

- Krok 1.: Príprava údajov.
- Krok 2.: Prezentácia výsledkov.
- Krok 3.: Experiment.

Prvý krok slúži na zadefinovanie a spracovanie údajov zo sociálnej siete Facebook. Prehľad už zadefinovaných údajov sa zobrazuje v základnej prehľadovej tabuľke (obr. E.2). Nové údaje je možné vytvoriť stlačením tlačidla *Pripraviť nové údaje*. Následne má používateľ možnosť zadefinovať, aké údaje majú byť zo sociálnej siete spracované, a to konkrétne v dvoch častiach:

- a) výber priateľov, pre ktorých sa má vypočítať intenzita vzťahov (obr. E.3),
- b) výber aktivít a údajov, ktoré sa majú použiť pri výpočte intenzity vzťahov (obr. E.4).

Po vytvorení nových údajov je používateľ presmerovaný do základnej prehľadovej tabuľky, kde má možnosť kontrolovať aktuálny stav spracovania. Po dokončení spracovania môže pokračovať na druhý krok, ktorým je prezentácia výsledkov.

Intensity Relationship Presenter  Ivan Srba

O projekte **Krok 1: Príprava údajov** Krok 2: Prezентация výsledkov Krok 3: Experiment Kontakt

Prípraviť nové údaje

Zoznam dostupných údajov			
Sledovaná osoba	Zdroj údajov	Stav údajov	Dátum a čas
Ivan Srba	Facebook	Prípravené	28.03.2010 23:46:15

Informácie

Príprava údajov

Pre výpočet intenzity je potrebné spracovať informácie o Vašich aktivitách v sociálnej sieti Facebook. Pri spracovaní nebudú trvalo ukladané žiadne Vaše osobné údaje ani údaje Vašich priateľov.


Nové údaje vytvoríte pomocou tlačidla "Prípraviť nové údaje".

Ako ďalej?

Po zadení nových údajov sa vytvorí v zozname nová položka. Spracovanie údajov môže trvať v závislosti od počtu zvolených priateľov až 5 minút. Akonáhle budú údaje pripravené, zmení sa text stavu údajov na hodnotu "Prípravené". Potom môžete zvoliť tieto údaje dvojklikom alebo kliknutím a vybratím záložky s druhým krokom - Prezентация výsledkov.

Strana 1 z 1 Zobrazujem záznamy 1 až 1 z 1

Obr. E.2 Prehľadová tabuľka zadefinovaných spracovaní údajov

Intensity Relationship Presenter  Ivan Srba

O projekte **Krok 1: Príprava údajov** Krok 2: Prezентация výsledkov Krok 3: Experiment Kontakt

Návrat na zoznam údajov Pokračovať...

Zoznam dostupných priateľov		Zoznam zvolených priateľov	
Meno	Priezvisko	Meno	Priezvisko
Adriána	Malá	Andrea	Prvá
Alena	Múdra	Braňo	Zaujímavý
Alexandra	Plachá	Dašena	Pekná
Anna	Veselá	Denisa	Jarná
Daniela	Rýchla	Dušan	Letný
Daša	Pomalá	Gabriel	Jesený
Ivan	Nezbeda	Gabriela	Zimný
Ivana	Šikovná	Gregy	Usporiadany
Jana	Maličká	Jan	Tichý
Jana	Veliká	Jana	Smutná
Juraj	Bezdeda	Janka	Biela
Juraj	Novák	Janka	Čierna
Ján	Kováč	Jozef	Nízky
Lenka	Pomýlená	Jozo	Vysoký
Lenka	Zmenená	Karolina	Pražská
Lubos	Výnimočný	Katarina	Bratislavská
Lukas	Posledný	Lenka	Konečná

Informácie

Príprava nových údajov

Vyberte zoznam priateľov (maximálne 100), ku ktorým chcete vypočítať intenzitu svojho vzťahu. Prvých 50 priateľov sme Vám už náhodne vybrali. Ak sa Vám takýto výber nepáči, môžete ho zrušiť ľavou dvojklikom a zvoliť si vlastný výber priateľov. Potom pokračujte prostredníctvom tlačidla "Pokračovať..."

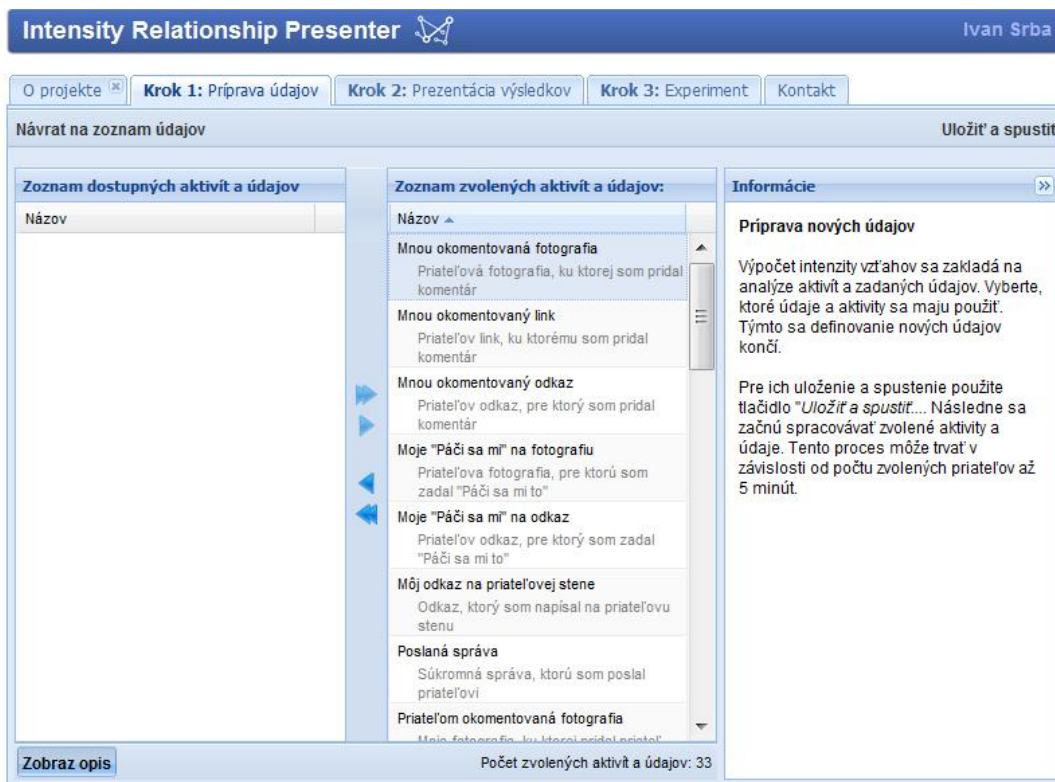
Tip: presun priateľov môžete realizovať:

- dvojklikom na vybraného priateľa
- označením priateľa, alebo skupiny priateľov a následným stlačením šípky v prostrednej lište
- označením priateľa, alebo skupiny priateľov a následným pretiahnutím zvoleného výberu do druhého okna

Tip: nechce sa Vám vyberať priateľov po jednom? Celú skupinu priateľov vyberiete stlačením tlačidla SHIFT a vybratím prvého a posledného priateľa. Všetci medzi nimi budú automaticky zvolení.

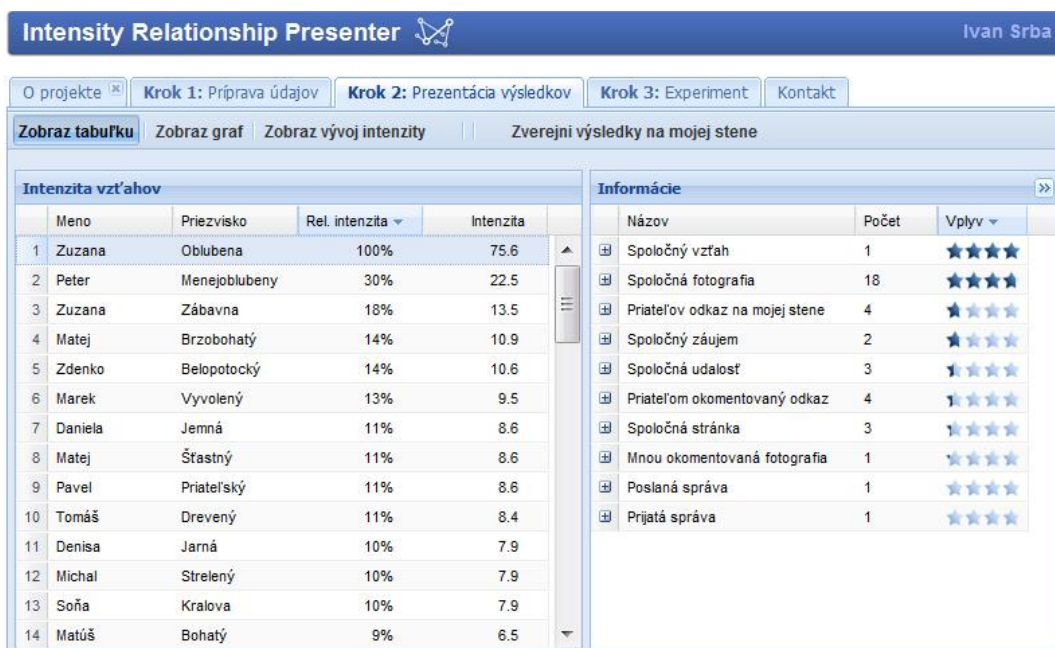
Počet zvolených priateľov: 51

Obr. E.3 Príprava nových údajov - výber priateľov



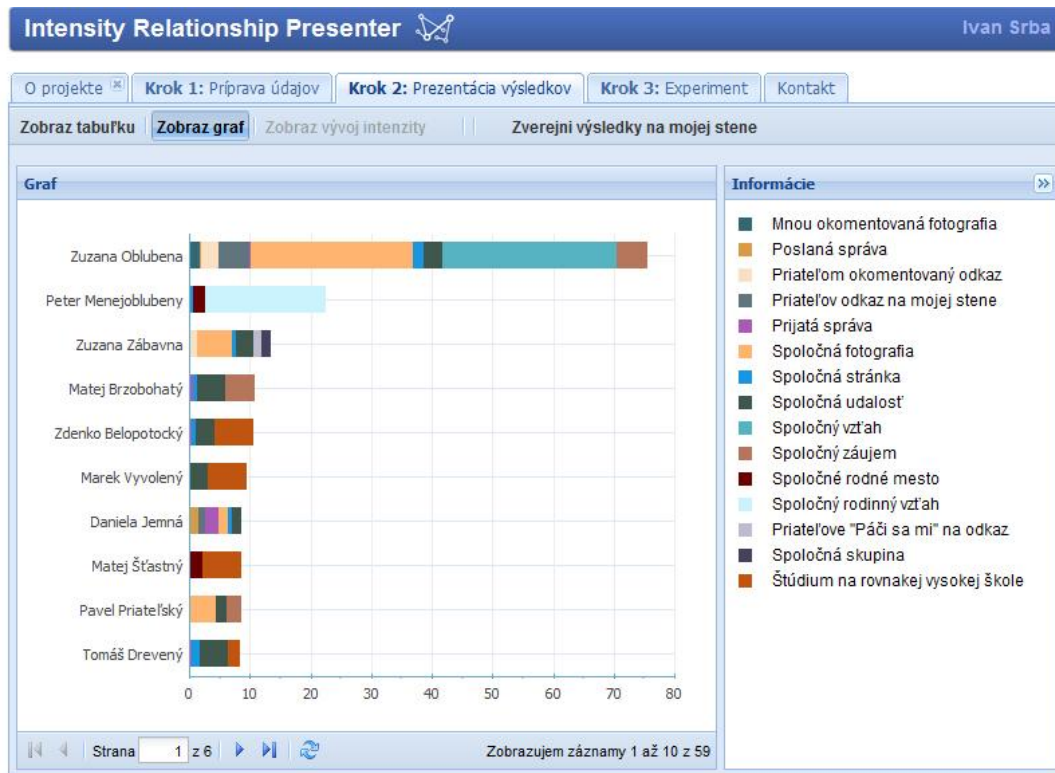
Obr. E.4 Príprava nových údajov - výber aktivít a údajov

V druhom kroku má používateľ možnosť prezerat' vypočítanú intenzitu na základe spracovaných údajov. Predvolene sa mu zobrazí prehľadová tabuľka vypočítaných intenzít vzťahov (obr. E.5). Kliknutím na jednotlivé riadky tabuľky sa v pravom rozširujúcom panely zobrazia doplňujúce informácie o zložení intenzity pre zvoleného používateľa. Jednotlivé riadky tejto doplňujúcej tabuľky je možné ďalej rozbaľiť a zobrazit' opis aktivity alebo údaju a v niektorých prípadoch aj konkrétne príklady danej aktivity, alebo údaju zo sociálnej siete.



Obr. D.5 Prezентация výsledkov – prehľadová tabuľka

Prezentáciu výsledkov je možné v hornej nástrojovej lište prepnúť na zobrazenie grafu (obr. E.6). Tu sa používateľovi zobrazí prehľad podielu jednotlivých aktivít a údajov na tvorbe celkovej intenzity každého vzťahu pre aktuálny čas.



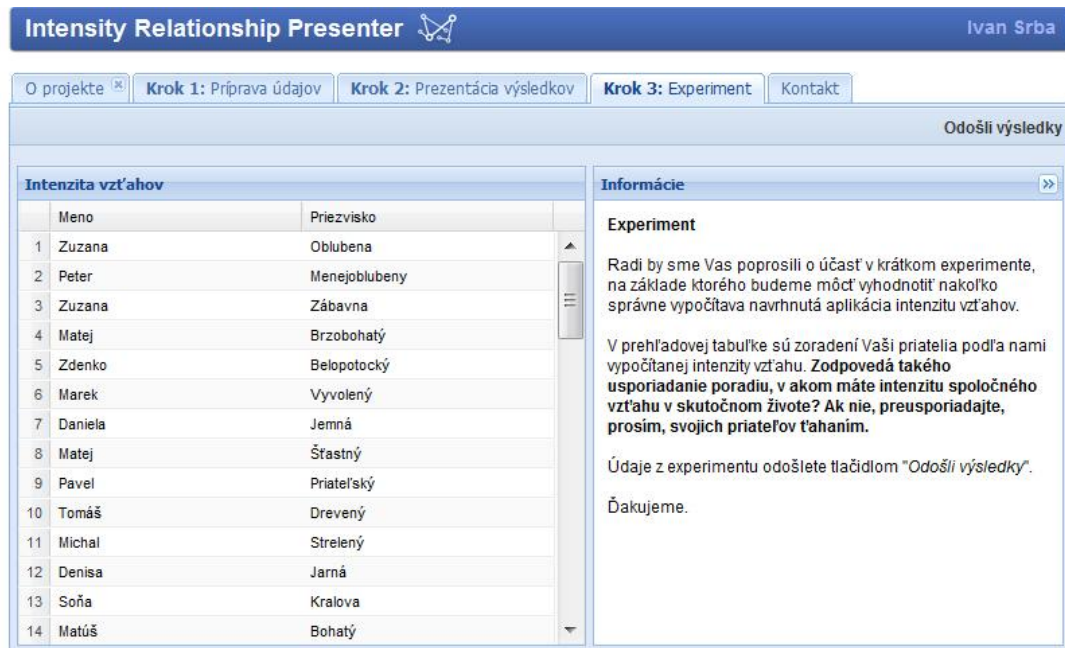
Obr. D.6 Prezentácia výsledkov – graf

Ďalšou formou prezentácie výsledkov je zobrazenie vývoja intenzity vzťahu v čase (obr. E.7)



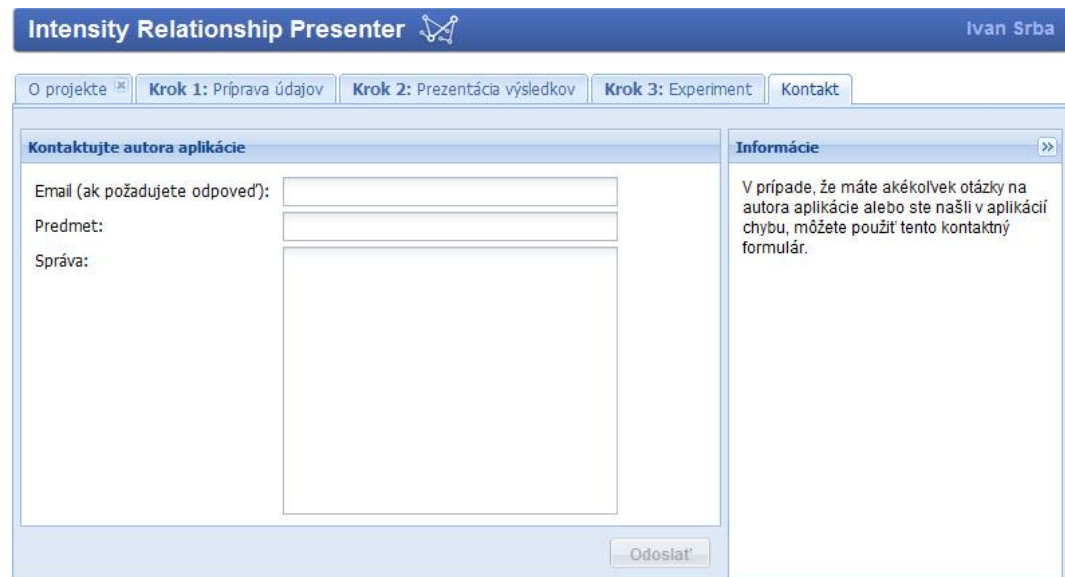
Obr. E.7 Prezentácia výsledkov – vývoj intenzity vzťahu v čase

Tretí krok pozostáva z vykonania experimentu (obr. E.8). V prehľadovej tabuľke sa zobrazia používateľovi mená zvolených priateľov, ktorí sú usporiadaní podľa intenzity vzájomného vzťahu od najsilnejšieho po najslabší.



Obr. E.8 Experiment

Aplikácia umožňuje používateľovi jednoducho poskytnúť explicitnú väzbu, a to prostredníctvom dostupného kontaktného formulára (obr. E.9).



Obr. E.9 Kontaktný formulár

Príloha F – Príspevok na konferenciu Web Intelligence

3rd Workshop on Collective Intelligence in Semantic Web and Social Networks

31. august – 3. september 2010

Toronto (Canada)

Zaslané dňa 23.04.2010

Tracing Strength of Relationships in Social Networks

Ivan Srba and Mária Bieliková
Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
srba07@student.fiit.stuba.sk, bielik@fiit.stuba.sk

Abstract

Current web is known as a space with constantly growing interactivity among its users. It is changing from a place for data storage to the social place, the place where people not only search interesting information, but also communicate and collaborate each with other. Obviously, the biggest places for common interaction are social networks considering relationships of people. We present a method for analysis of the strength and the evolution of relationships between users based on varied user's activities performed in social networks. We evaluate our approach employing Facebook social network.

1. Introduction and Related Work

Nowadays we witness rapid expansion of new generation of services known as the Web 2.0. One kind of these services are social networks where people can arrange and express several types of relationships. The strength of these relationships among users differs and can rapidly change in the time.

Knowledge on relationship intensity between two users is useful for many actions which depend on actual context. For example we can use information about intensity of relationship for recommendation based on an assumption that if one user is interested in some object the other user can be also interested in this same object if the users are somehow related [1]. Another example is web page adaptation based on preferences and behavior of a user with whom we have the strong relationship. Also the user can control the evolution of his relationships to friends for preserving the relationships considered by the user as important.

Recently, a lot of works focused on analyzing the structure and properties of large-scale social networks [2], [3], [4]. These studies cover social networks on high level and provide only general information about nodes and links between these nodes in traced social network. They do not

capture any information which is related to relationships between two concrete users.

For example the authors in [5] provide information on users of social portal YouTube and divide users into several groups following user's behavior. Even if this information is useful, it does not give us enough information about relationship between two users.

There are only several researches aiming to activities of particular users based on employing an *activity network*. The activity network is the network which is created by users who express their relationships using functionality provided by the social network. Chun et al. present work about activity network in Cyworld social portal [6]. The result of research was that activity network was really similar to social network. Another research [7] provides information that the activity network in the social portal Facebook is different from the graph of social network (considering graph properties).

The authors in [8] describe a method of user's relationships analysis based on Facebook's *wall* posts. The method has several limitations: it uses only wall posts and does not consider any other relationship expression (e.g., photo upload, messages), extraction was limited to subset of one region users (New Orleans Network), it does not use the access to all profiles, only to those publicly available. We consider as serious limitation also the fact that it is not universal and cannot be used with arbitrary social network.

2. Analysis of relationships intensity

Our method for analysis of relationships intensity primarily concerns a calculation of relationships intensity strength. In order to perform the calculation we need to guarantee access to social network's data and fulfill a precondition in terms of the input data.

2.1. Input data

Partial user's relationship intensity can be expressed by elementary activity between two users. Example of this activity is sending a message or uploading common photography. Such activity we denote as a *rate factor*. Rate

factor represents identifiable and countable expression of users' relationship. It can influence strength of a relationship in positive or negative way depending on effect of the activity to human relationships.

Rate factor can be shared among several *sources* of user's activities (social networks). Meanwhile for each source it can have different importance, which is represented numerically by a *weight*. The weight of rate factor expresses relative influence of the rate factor to the final relationship intensity. Weight for each source and rate factor is assigned experimentally (in this paper we provide weights estimated for Facebook social network). Examples of rate factors with weights determined according analysis and monitoring of users' behaviors are common photography (positive influence, weight 0.13), study at the same school (positive influence, weight 0.012), boyfriend/girlfriend relationship (positive influence, weight 0.95), not answering on messages for more than two days (negative influence, weight 0.063).

Not only weight but also the *count* of all appearances of the rate factor (not only in relationship of two users who are traced) influences final relationships strength. This fact assures that also the frequency of using the social network has effect to partial relationship intensity.

Time influences partial relationship intensity too. To include effect of time we differentiate rate factors of single activity, interval activity or unbounded activity.

Rate factor of single activity represents a single-shot expression of relationship for which we can identify the date and time of appearance. Furthermore we can estimate the duration of influence to the common relationship. Example is sending the message where we know the date and time of its sending and influence duration was experimentally estimated to two days.

Rate factor of interval activity represents expression of a relationship based on the activity for which we can identify the date and time of beginning and completion. As well as in previous type we can estimate the duration of influence. Example is study at the same school from September 2006 to June 2010, the duration of influence in this case is 365 days.

Rate factor of unbounded activity represents time unbounded activity for which we cannot define start and end date. Example is family relationship.

2.2. Calculation of the relationship intensity

As we have mentioned partial relationship intensity depends on weight, count of all appearances of rate factor and time:

$$I_f(k, j) = \frac{w_{kj} \sum_{i=1}^l f_i}{1 + \ln(1 + l_c)}$$

where I_f is partial relationship intensity for one rate factor, w_{kj} is weight of rate factor j for source k , l is count of appearances of rate factor in relationship of two traced users, l_c is count of all appearances of rate factor, and f_i is function expressing time influence.

Function expressing time influence depends on type of rate factor. We identified three variations of time influence functions according type of activity.

Function f_t for rate factor of single activity is:

$$f_t(t_{kj}, t_m, t_v) = \begin{cases} 1/(1 + \log_{\sqrt{t_{kj}}}(\max(1, t_m - t_v))) & t_m \geq t_v \\ 0 & t_m < t_v \end{cases}$$

where t_{kj} is time of influence duration in days, $\sqrt{t_{kj}}$ expresses the duration while intensity logarithmically decreases to the half of original value, t_m is date and time for which we calculate intensity, t_v is date and time of single activity, and $t_m - t_v$ is time difference in days between single activity and moment for which we calculate intensity.

Figure 1 illustrates how time function influences partial relationship intensity for single activity. In this case, the activity happened on second day, influence duration is set to 25 days so we can see that in the seventh day time function has value 0.5.

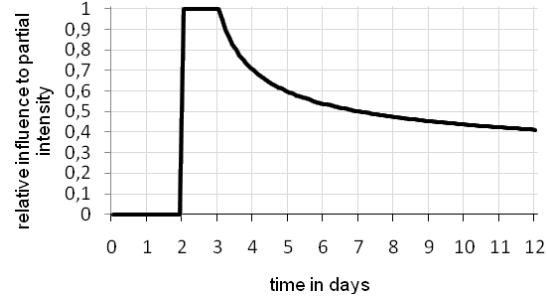


Figure 1. Time function for single activity.

Function f_i for rate factor of unbounded activity is:

$$f_i = \frac{1}{2}$$

Function f_i for *rate factor of interval activity* is adapted combination of previous two functions:

$$f_i(t_{kj}, t_m, t_z, t_k) = \begin{cases} 0 & t_m < t_z \\ 1 & t_m \in (t_z, t_k) \\ 1/(1 + \log_{\sqrt{t_{kj}}}(\max(1, t_m - t_k))) & t_m > t_k \end{cases}$$

where t_z is date and time of beginning interval activity, and t_k is date and time of ending interval activity.

Figure 2 illustrates how time function influences partial relationship intensity for

interval activity (influence duration is set to 25 days, interval activity starts on the second day and ends on the fifth day).

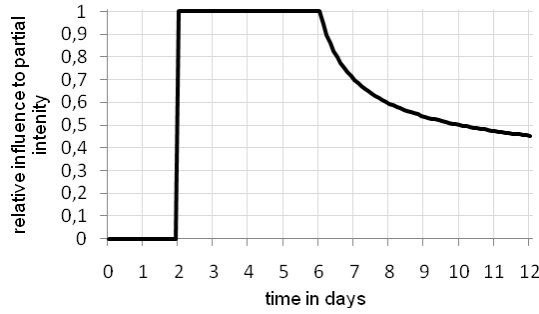


Figure 2. Time function for interval activity.

Partial relationship intensity for one source is a summation of all partial relationship intensity from rates factors. Final relationship intensity is computed as arithmetic average of partial relationship intensity from all sources.

3. Evaluation

We developed web-based application for computing the strength of relationships in social networks. Our design considers universality of the application, so we do not constrain neither the social network neither the rate factors. The application is divided into two main components: *Intensity Relationships Analyzer (IReA)* and *Intensity Relationships Presenter (IReP)*.

IReA uses wrappers to connect to social networks and to automatically acquire values for rate factors. For each new social network we need just to prepare a new wrapper.

IReA provides API available for the client component IReP. This API is used to start extraction (set extraction properties such as required sources, rate factors) and then for calculating and returning results.

For experimenting we have chosen popular social portal Facebook. We have also considered Flickr as another source but in this social portal only small part of users is source of majority actions so it is not appropriate social network for evaluation [9].

We evaluated our method in the experiment where we investigated (i) distribution of user interaction among friends, (ii) evolution of the calculated intensity of relationships in time, (iii) comparison of calculated results to the expected (for first ten best friends). Results in this paper are based on the activity of 238 users with 25,435 relationships.

3.1. User interaction among friends

At first we investigate the distribution of wall posts and all activities (rate factors of single

activity) among friends. We calculate cumulative distribution function (CDF) for this purpose (see Figure 3). About 82% of relationships have no wall activities. This result was expected and indicates that only small subset of users make majority of wall posts. About 50% of relationships have no occurrence of rate factor for single activity. It means that the graph representing activity network is really different from the graph representing social network. This result also shows that distribution of wall posts is quite similar but not equal to distribution of all activities.

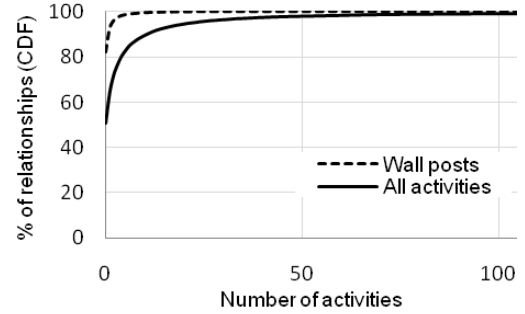


Figure 3. Distribution of user's activities among friends.

Second, we sum up for each user the intensity of single activities. Then we select three values (70%, 90% and 100%) of summed intensity and calculate which portion of relationships is involved in the required part of the intensity. Figure 4 presents cumulative distribution function for these percentages. This figure gives us a lot of useful information, e.g. information that for all users 90% amount of intensity is created only by 54% of relationships.

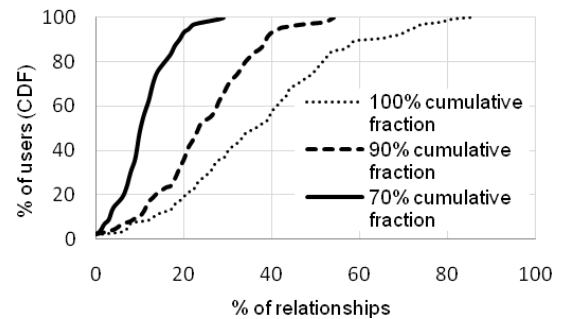


Figure 4. Distribution of relationships intensity among friends for three different amounts of intensity.

3.2. User interaction in time

To evaluate distribution of user interaction in time we calculate refresh rate of relationships (see Figure 5) for one year with month interval in the same way as described in [8]. Result is that only 31% of user pairs continue to interact in following

month. This result is almost the same as result 29% achieved in [8]. It proves that users regularly change friends who are involved in user's activities.

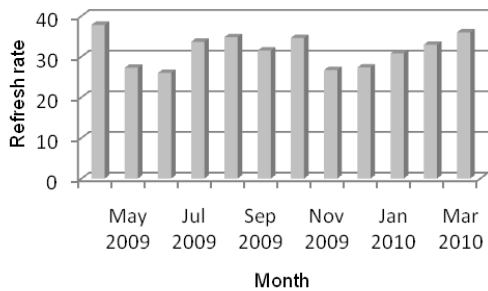


Figure 5. Refresh rate (one month interval).

To inspect dynamics of relationships in time we calculate relative change. We noticed big differences between more and less active relationships so we divided them into five groups according calculated intensity. The results are plotted in Figure 6. Average relative change of all relationships is 17%. Results show that highly coupled users do not change their relationships so often and dramatically as other users.

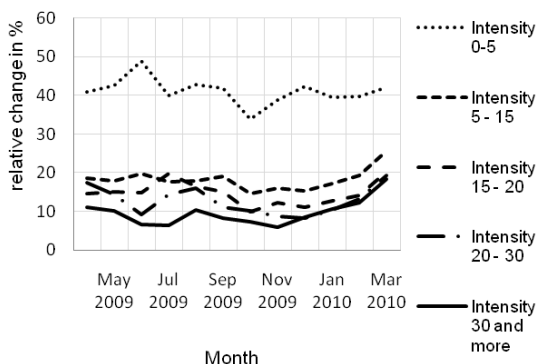


Figure 6. Relative intensity change for different groups of relationships.

3.3. User reception of calculated results

We also have got explicit feedback from users and have compared the list of top ten best friends sorted by users with the list of 10 friends with the highest calculated relationship intensity. Evaluating 95 acquired lists with friends sorted by the users we have found 88% overlap of the persons in both lists. The result is influenced also by the fact that default weights and times of duration for some rate factors are not appropriate for all traced users.

4. Conclusion

In this paper we propose and evaluate a method for analysis of relationship intensity. Approximation of user's relationships with other

users presents valued information on share the same interests, the same ways how to spend free time etc. Estimated relationships strengths can be used in many areas where we want to know more information about our users, for example on web pages which adopt their content according to current user's goals or characteristics.

5. References

- [1] Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: *Proc. of the 2000 ACM Conf. on Computer supported cooperative work*, Philadelphia (USA): ACM New York, 2000, pp. 241-250.
- [2] Hanneman, R., Riddle, M.: *Introduction to social network methods* [online], Riverside (California): University of California, 2005. <http://faculty.ucr.edu/~hanneman/nettext> (09. 02. 2010)
- [3] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B.: Measurement and Analysis of Online Social Networks. In: *Proc. of the 7th ACM SIGCOMM Conf.*, San Diego (California): ACM New York, 2007, pp. 29 - 42.
- [4] Ahn, Y., Han, S., Kwak, H., Moon, S., Jeong, H.: Analysis of topological characteristics of huge online social networking services. In: *Proc. of the Int. Conf. on WWW*, Alberta (Canada): ACM New York, 2007, pp. 835 - 844.
- [5] Maia, M., Alemida, J., Almeida, V.: Identifying user behavior in online social networks. In: *Proc. of the 1st workshop on Social network systems*, Glasgow (Scotland): ACM New York, 2008, pp. 1-6.
- [6] Chun, H., Ahn, Y., Eom, Y., Kwak, H., Moon, S., Jeong, H.: Comparison of online social relations in volume vs interaction: a case study of cyworld. In: *Proc. of the 8th ACM SIGCOMM Conf. on Internet Measurement*, Vouliagmeni (Greece): ACM New York, 2008, pp. 57-70.
- [7] Wilson, C., Boe, B., Sala, A.: User interactions in social networks and their implications. In: *Proc. of the 4th ACM European Conf. on Computer Systems*, Nuremberg (Germany): ACM New York, 2009, pp. 205-218.
- [8] Viswanath, B., Mislove, A., Cha, M., Gummadi, K.P.: On the evolution of user interaction in Facebook. In: *Proc. of the 2nd ACM Workshop on Online Social Networks*, Barcelona (Spain): ACM New York, 2009, pp. 37-42.
- [9] Valafar, M., Rejaie, R., Willinger, W.: Beyond friendship graphs: a study of user interactions in Flickr. In: *Proc. of the 2nd ACM Workshop on Online Social Networks*, Barcelona (Spain): ACM New York, 2009, pp. 25-30.

Príloha G – Obsah elektronického média

- *Bakalarska-praca* - elektronická verzia dokumentu k bakalárskej práci
- *Implementacia*
 - *Datovy-model* - schémy a DDL implementovaného dátového modelu
 - *Dokumentacia*
 - *IRA* - dokumentácia k zdrojovému kódu IReA
 - *IRP* - dokumentácia k zdrojovému kódu IReP
 - *GUI-aplikacie* - snímky grafického prostredia aplikácie
 - *Webova-sluzba* - opis webovej služby vo WSDL
 - *Zdrojovy-kod*
 - *IRA* - zdrojový kód serverovej aplikácie IReA
 - *IRP* - zdrojový kód klienta IReP
- *Konferencia-IIT.SRC* - článok a poster na konferenciu IIT.SRC 2010
- *Konferencia-WI* - článok na konferenciu WI 2010