

Slovenská technická univerzita v Bratislave  
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ  
ŠTUDIJNÝ PROGRAM: Softvérové inžinierstvo

Bc. Marek Tomša

**ADAPTÍVNA PODPORA  
NAVIGÁCIE V OTVORENÝCH  
PRIESTOROCH**

*Diplomová práca*

Vedúci diplomovej práce: prof. Ing. Mária Bieliková, PhD.

máj, 2008

## ZADANIE DIPLOMOVEJ PRÁCE

Meno študenta: **Bc. Marek Tomša**  
Študijný odbor: **SOFTVÉROVÉ INŽINIERSTVO**  
Študijný program: **Softvérové inžinierstvo**  
Názov projektu: **Adaptívna podpora navigácie v otvorených priestoroch**

### Zadanie:

Efektívne sprístupňovanie a navigácia v množstve dostupných informácií na webe je v súčasnosti aj napriek nespornému technologickému pokroku stále aktuálny. Problémy sú najmä v strate prehľadu o kontexte vyhľadávania ("okolie" práve prezeranej webovej stránky). Nelineárne prepojené informačné zdroje sme pomocou dnešných webových prehliadačov schopní prehliadať výlučne lineárnym prístupom. Analyzujte problematiku navigácie na webe a existujúce riešenia v tejto oblasti. Porovnajte špecifiká navigácie v rôznych typoch informačných priestorov vrátane priestoru metadát dôležitom v rámci iniciatívy webu so sémantikou a možnosťami personalizácie navigácie. Analyzujte možnosti využitia sociálnych aspektov navigácie na webe vrátane kolaboratívne vytváraných značiek k jednotlivým informačným zdrojom. Navrhňte metódu, ktorá umožní personalizovanú navigáciu v priestoroch reprezentovateľných grafom spolu s vizualizáciou týchto priestorov s ohľadom na ich potenciálne veľký rozsah. Navrhnuté riešenie experimentálne overte pre vybranú doménu dané prehliadaným informačným priestorom na základe vyvinutého softvérového prototypu.

### Odporúčaná literatúra:

Bieliková, M., Jemala, M.: Adaptive Incremental Browsing of Ontology Structure. In: Hypertext 2007, Proceedings of the Eighteenth ACM Conference on Hypertext and Hypermedia, Manchester, UK, Sept. 10-12, 2007. ACM Press, pp. 143-144.

Wu, H., Zubair, M., and Maly, K. 2006. Harvesting social knowledge from folksonomies. In Proceedings of the Seventeenth Conference on Hypertext and Hypermedia (Odense, Denmark, August 22 - 25, 2006). HYPERTEXT '06. ACM, New York, NY, 111-114.

Brusilovsky, P.: Social Information Access: The Other Side of the Social Web. In: V. Gefert et al. (eds.) Proceedings of SOFSEM 2008, 34th International Conference on Current Trends in Theory and Practice of Computer Science, High Tatras, Slovakia, January 19-25, 2008.

### Práca musí obsahovať:

- Anotáciu v slovenskom a anglickom jazyku
- Analýzu problému
- Opis riešenia
- Zhodnotenie
- Technickú dokumentáciu
- Zoznam použitej literatúry
- Výstupy celého diplomového projektu vrátane vlastnej diplomovej práce a vytvoreného softvéru (zdrojového kódu s dokumentáciou)

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

Termín odovzdania práce v letnom semestri: dňa 14. mája 2008

Bratislava, dňa 18. februára 2008



prof. Ing. Pavol Návrat, PhD.  
riaditeľ ÚISI

# Anotácia

---

Slovenská technická univerzita v Bratislave

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ

Študijný odbor: SOFTVÉROVÉ INŽINIERSTVO

Študijný program: Softvérové inžinierstvo

Autor: Bc. Marek Tomša

Diplomová práca: Adaptívna podpora navigácie v otvorených priestoroch

Vedenie diplomovej práce: prof. Ing. Mária Bieliková, PhD.

máj, 2008

Práca sa zaoberá témou navigácie na webe. Analyzuje problémy navigácie na webe a ich príčiny. Uvádza prehľad existujúcich metód a nástrojov pre podporu navigácie na webe a tiež prístupov pre personalizáciu a sociálne aspekty používania webu. V kontexte existujúcich služieb webu pre sociálne značkovanie diskutuje možnosti využitia týchto služieb pre podporu navigácie. Navrhuje všeobecne použiteľnú metódu adaptívnej podpory navigácie v priestoroch reprezentovateľných grafmi a jej špecializáciu vo forme metódy pre adaptívnu podporu navigácie na webe. Ďalej uvádza špecifikáciu a návrh softvérového systému využívajúceho navrhnutú metódu a overenie navrhnutého systému jeho implementáciou. Predstavuje výsledky overenia vlastností navrhnutej metódy a analýzu priestoru webu z hľadiska charakteristík relevantných pre fungovanie navrhnutej metódy. Práca ďalej obsahuje používateľskú štúdiu, ktorej cieľom bolo overiť užitočnosť implementovaného riešenia pre koncového používateľa, a diskutuje zistené závery.

# Annotation

---

Slovak University of Technology Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Degree Course: SOFTWARE ENGINEERING

Author: Bc. Marek Tomša

Thesis: Adaptive navigation support in open spaces

Supervisor: prof. Ing. Mária Bieliková, PhD.

2008, May

This work discusses the topic of Web navigation. We analyze problems with navigation on the Web and their reasons. We present an overview of existing methods and tools for navigation support and an overview of approaches to personalization and social aspects of web usage. In context of existing social bookmarking services, we discuss the possibility to use these services for navigation support. We propose a generally usable method for adaptive navigation support in information spaces which can be represented as graphs and a specialization of the proposed method to be used for adaptive navigation support on the Web. A specification and design of a system employing the proposed method and its evaluation by the implementation of the system is presented. We present results of evaluation of the proposed method and analysis of characteristics of the Web space relevant for the proposed method. Finally, a user study with the aim of evaluating the perceived usefulness of the proposed method and its implementation in a form of the web navigation support application is presented and results are discussed.

## **Prehlásenie**

Čestne prehlasujem, že som prácu vypracoval samostatne a použil som len uvedené zdroje.

Marek Tomša

## **Pod'akovanie**

Týmto sa chcem pod'akovať pani profesorke Márii Bielikovej za jej cenné rady pri vedení diplomovej práce a mojím rodičom za podporu a zázemie pri jej vypracovaní.

# Obsah

---

<b>1</b>	<b>ÚVOD.....</b>	<b>1</b>
<b>2</b>	<b>NAVIGÁCIA NA WEBE A WEBE SO SÉMANTIKOU .....</b>	<b>3</b>
2.1	NAVIGÁCIA V DOBRE DEFINOVANÝCH PRIESTOROCH .....	5
2.2	METÓDY A NÁSTROJE PRE PODPORU NAVIGÁCIE NA WEBE .....	6
2.3	ZHRNUTIE.....	12
<b>3</b>	<b>PERSONALIZÁCIA A SOCIÁLNE ASPEKTY NAVIGÁCIE .....</b>	<b>14</b>
3.1	PERSONALIZÁCIA A MODELOVANIE POUŽÍVATEĽA .....	14
3.2	ZNAČKOVANIE NA WEBE .....	15
3.3	ADAPTÍVNA SOCIÁLNA NAVIGÁCIA .....	18
<b>4</b>	<b>OTVORENÉ PROBLÉMY A CIELE PRÁCE .....</b>	<b>19</b>
<b>5</b>	<b>METÓDA ADAPTÍVNEJ NAVIGÁCIE V OTVORENOM PRIESTORE ..</b>	<b>21</b>
5.1	SCHÉMA METÓDY .....	21
5.2	APLIKÁCIA METÓDY PRE NAVIGÁCIU NA WEBE .....	22
5.2.1	Abstrakcia navigácie na webe na graf .....	22
5.2.2	Forma zobrazenia.....	22
5.2.3	Vyhodnotenie relevancie hrán .....	24
5.3	VÝHODY A OBMEDZENIA NAVRHNUTEJ METÓDY .....	26
<b>6</b>	<b>REALIZÁCIA ADAPTÍVNEJ PODPORY NAVIGÁCIE .....</b>	<b>28</b>
6.1	POŽIADAVKY.....	28
6.2	NÁVRH SYSTÉMU PRE PODPORU NAVIGÁCIE.....	30
6.3	IMPLEMENTÁCIA V PROSTREDÍ WEBU .....	37
6.3.1	Možnosti rozšírenia.....	38
6.3.2	Použité technológie.....	39
<b>7</b>	<b>EXPERIMENTÁLNE OVERENIE METÓDY .....</b>	<b>40</b>
7.1	OVERENIE VLASTNOSTÍ FUNKCIE PODOBNOSTI STRÁNOK NA ZÁKLADE SPOLOČNÝCH ZNAČIEK .....	40
7.2	ANALÝZA PRIESTORU WEBU .....	43
7.3	POUŽÍVATEĽSKÁ ŠTÚDIA .....	45
7.4	ZÁVERY EXPERIMENTOV .....	46
<b>8</b>	<b>ZÁVER .....</b>	<b>47</b>
	<b>LITERATÚRA .....</b>	<b>48</b>
	<b>PRÍLOHY .....</b>	<b>51</b>

<b>A</b>	<b>OBSAH ELEKTRONICKÉHO MÉDIA .....</b>	<b>52</b>
<b>B</b>	<b>TECHNICKÁ DOKUMENTÁCIA .....</b>	<b>53</b>
	B.1 PARALELNÝ MECHANIZMUS PRÍSTUPU K SLUŽBE DEL.ICIO.US .....	53
	B.2 ŠPECIALIZÁCIA TRIEDY VŠEOBECNÉHO NAVIGÁTORA.....	56
<b>C</b>	<b>INTEGRATION, INSTALLATION AND DEVELOPMENT MANUAL.....</b>	<b>59</b>
	C.1 BASIC INFORMATION.....	59
	C.1.1 Method Description .....	59
	C.1.2 Scenarios of Use .....	60
	C.1.3 External Links and Publications .....	60
	C.2 INTEGRATION MANUAL.....	60
	C.2.1 Dependencies .....	61
	C.2.2 Installation.....	61
	C.2.3 Configuration .....	61
	C.2.4 Integration Guide .....	62
	C.3 DEVELOPMENT MANUAL .....	64
	C.3.1 Tool Structure .....	64
	C.3.2 Method Implementation.....	64
	C.3.3 Enhancements and Optimization .....	65
	C.4 MANUAL FOR USE IN OTHER APPLICATION DOMAINS .....	66
	C.4.1 Configuration for use in Other Application Domains.....	66
	C.4.2 Dependencies .....	66
<b>D</b>	<b>PUBLIKÁCIE VZŤAHUJÚCE SA K VÝSLEDKOM DIPLOMOVEJ PRÁCE.....</b>	<b>67</b>
<b>E</b>	<b>NÁVRH ČLÁNKU .....</b>	<b>71</b>



# 1 Úvod

---

Navigácia na webe je problémom asi už od vzniku webu ako takého. Problém straty v hyperpriestore je dobre známy a identifikovaný už viac ako dvadsať rokov. Príchod vylepšení existujúceho webu vo forme idey webu so sémantikou dáva predpoklad pre priamočiarejšie riešenie viacerých problémov webu. Avšak web so sémantikou prináša momentálne prínos v prvom rade pre stroje. Pre bežného používateľa, ktorý môže mať problém vôbec pochopiť, aké problémy web so sémantikou adresuje, môže byť význam pokroku, ktorý sa v tejto oblasti robí, naozaj len marginálny. Avšak úplne každý používateľ webu pozná problém straty orientácie a prehľadu o bezprostrednej histórii prehliadania, čo bežne vyústí do stavu straty v hyperpriestore.

Existujúce nástroje pre podporu navigácie bývajú založené na explicitných formálne zapísaných znalostiach a aplikovateľné len na uzatvorené priestory. Web je však nielen z pohľadu navigácie neohraničený priestor (množstvo informácií na webe určite rastie rýchlejšie ako je schopný akýkoľvek človek spracúvať), ale v súčasnom stave ani nie je možné spoľahlivo definovať nad existujúcim webovým obsahom explicitnú sémantiku. Zároveň existujú na webe prúdy, založené na obsahu generovanom používateľmi a značkovaní webových zdrojov, bez vyžadovania explicitnej sémantiky. Takýto prístup sa zatiaľ ukázal v porovnaní s ideami webu so sémantikou ako pre používateľov jednoduchšie využiteľný a v konečnom dôsledku aj hodnotnejší. Zatiaľ čo existuje mnoho úspešných a veľmi široko rozšírených aplikácií využívajúcich jednoduchý princíp kolaboratívneho značkovania (označovaný tiež ako *folksonomy*), vytvárajúci v konečnom dôsledku „používateľmi generované slovníky“, zatiaľ nikto neprišiel s aplikáciou, ktorá by využívala možnosti silnej formálnej vyjadrovacej schopnosti technológií a jazykov webu so sémantikou, a najmä dokázala presvedčiť producentov webového obsahu, aby ju v širokom meradle využívali. Ťažko prekonateľný problém webu so sémantikou je tiež problém transformácie obrovského množstva existujúceho webového obsahu do ich reprezentácie zachytávajúcej sémantiku. Na druhej strane, je veľmi výhodné práve dnes, keď stojíme možno blízko prahu ozajstného nástupu webu so sémantikou, navrhovať riešenia tak, aby boli využiteľné pri aplikácii oboch prístupov.

V práci sa sústreďíme na problematiku navigácie a neobmedzujeme sa len na web ako taký, alebo výlučne technológie webu so sémantikou, ale snažíme sa o zachytenie všeobecného problému navigácie a jeho možných riešení v kontexte existujúcich webových technológií. Vychádzame z existujúcich široko rozšírených zdrojov metadát, akými sú napríklad služby poskytujúce možnosti sociálneho značkovania. Používatelia s pomocou takýchto služieb vytvárajú nad webom im samým zrozumiteľnú sémantiku, takže tento prístup sa javí ako veľmi vhodný pre podporu a prispôbenie navigácie týchto používateľov v im neznámom priestore webu.

V nasledujúcej kapitole analyzujeme problematiku navigácie na webe a webe so sémantikou, identifikujeme problémy navigácie, porovnávame navigáciu v otvorených a uzavretých priestoroch a analogicky porovnávame navigáciu a problémy navigácie na

webe a webe so sémantikou. Tiež uvádzame prehľad metód a nástrojov pre podporu navigácie na webe a prehľad techník pre adaptívnu podporu navigácie. V ďalšej kapitole (3) uvádzame prístupy k modelovaniu používateľa, sociálne aspekty navigácie a prehľad prostriedkov personalizácie, v ktorých sú využité. Tiež analyzujeme problematiku sociálneho značkovania na webe a diskutujeme rôzne zdroje značiek a možnosti usudzovania nad značkami pre modelovanie používateľa. Ďalej v kapitole 4 sumarizujeme analyzovanú problematiku, identifikujeme problémy a formulujeme ciele práce. V nasledujúcej kapitole (5) uvádzame návrh metódy adaptívnej podpory navigácie v priestoroch reprezentovateľných formou grafu. Metódu špecializujeme pre priestor webu a v nasledujúcej kapitole uvádzame špecifikáciu a návrh systému pre adaptívnu podporu navigácie na webe. V kapitole 6 uvádzame realizáciu systému pre podporu navigácie, v ktorom využívame navrhnutú metódu. Špecifikujeme požiadavky na takýto systém, navrhujeme jeho architektúru a jednotlivé časti. Uvádzame postup implementácie, fyzické rozloženie systému a použité technológie spolu s problémami, na ktoré sme pri implementácii narazili. Ďalej v kapitole 7 uvádzame výsledky overovania vlastností navrhutej metódy a prototypovania vybraných častí navrhnutého systému. Tiež uvádzame návrhy a výsledky experimentov, ktorými sme navrhnuté riešenie overili. Poslednou kapitolou je záver, kde sumarizujeme prácu a jej hlavné prínosy

Práca obsahuje 5 príloh. Technická dokumentácia ilustruje ukážky zaujímavých častí zdrojových kódov. Ďalšou prílohou je technická dokumentácia, vypracovaná podľa metodiky projektu NAZOU<sup>1</sup> v anglickom jazyku. Priložené sú tiež články, v ktorých boli publikované priebežné výsledky diplomovej práce na domácich aj celosvetových konferenciách a návrh článku sumarizujúceho výsledky diplomovej práce. Fyzickou prílohou práce je elektronické médium, ktoré obsahuje elektronickú verziu práce, zdrojové kódy implementovaného riešenia a prislúchajúcu technickú dokumentáciu.

---

<sup>1</sup> <http://nazou.fiit.stuba.sk>

## 2 Navigácia na webe a webe so sémantikou

---

*“Toto médium [web] je preplnené informáciami, zamorené redundantnými, chybnými a nekvalitnými informáciami. Postupuje smerom k neporiadku na základe princípu entropie a nemá žiadnu dobre definovanú štruktúru, ktorá by organizovala webové sídla. Navyše, používatelia nemajú žiadny globálny pohľad o celom webe, ktoré by im pomohli dostať sa k relevantným stránkam rýchlejšie.” (Abrams et al., 1997).*

Aj keď v poslednej dobe web technologicky veľmi rýchlo napreduje a dnes už ponúka mnoho aplikácií, ktoré majú používateľské rozhrania na úrovni, akú určite nikto ešte pred niekoľkými rokmi nepredpokladal, stále nemožno web považovať za médium, ktoré by bolo v stave takom, aký by umožňoval naozaj efektívne vyhľadávanie a prehliadanie informácií. Dnes sú už na webe úplne bežné bohaté klientské rozhrania, existujú štandardizované webové služby, ktoré môžu fungovať naprieč platformami a pridávajú k dátam webu vrstvu operácií, začína sa rozvíjať web so sémantikou a iné technologické inovácie, ktoré posúvajú možnosti v tejto oblasti dopredu, ale stále je pri navigácii na webe prítomný základný problém, a síce problém straty v hyperpriestore.

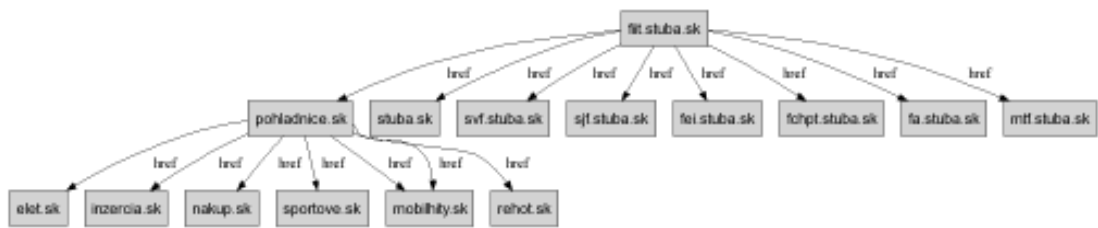
Tento problém na webe je perzistentný a známy už veľmi dlho. Už v roku 1987 identifikoval Conklin dva základné problémy s navigáciou v hypertexte: stratu orientácie a kognitívne preťaženie (Conklin, 1987). Strata orientácie vyplýva aj z nemožnosti zachytiť sémantiku odkazov pri nasledovaní hypertextových odkazov. Odkazy sú navyše jednosmerné. Používateľ pri nasledovaní odkazu okamžite stráca kontext stránky, z ktorej prišiel. Stránky sa zjavujú od „nikadiaľ“ a po ich opustení miznú „nikam“. Štruktúra ich vzájomných prepojení je pre používateľa neviditeľná (Kreutz, 1998).

Pri navigácii na webe rozlišujeme nasledovné typy hypertextových prepojení (Brunk, 1999):

- skok na iné miesto v rámci tej istej stránky,
- skok na inú stránku v rámci domény (relatívne k base URL) – hierarchické odkazy,
- skok na inú stránku (iné base URL) – krížové odkazy.

Povaha hypertextových prepojení, ktoré nezahŕňajú žiadnu sémantiku, ešte ďalej zvyšuje šancu, že používateľ sa pri prehliadaní webu stratí, alebo jednoducho dostane tam, kam nechcel. Hypertextové prepojenia na webe prepájajú častokrát sémanticky nepodobné stránky, a už dvomi kliknutiami sa používateľ môže dostať na (vzhľadom na jeho cieľ) úplne irelevantnú stránku.

Uvedené potvrdzuje vykonaná analýza prepojenia webových sídiel, kde sa ukazuje, že už na dve kliknutia sa dá zo stránky [www.fiit.stuba.sk](http://www.fiit.stuba.sk) dostať napríklad na stránku [www.rehot.sk](http://www.rehot.sk), ktorá s pôvodnou stránkou nemá naozaj nič spoločné. Časť grafu prepojení webových stránok ilustrujúca túto skutočnosť je zobrazená na obrázku 1.



Obrázok 1: Graf prepojení webových stránok.

Používateľ začína typicky sedenie na webe hľadaním. Často nevie čo presne chce nájsť (topologicky vzaté, ktoré odkazy má nasledovať, aby našiel čo potrebuje). Typicky vyberá odkazy z výsledkov poskytnutých fulltextovým vyhľadávačom, kliká na tlačidlá „späť“ a „vpred“, prípadne prezerá históriu prehľadávania. Kvôli obmedzenému doprednému prehľadu mu ostáva len hádať, ktorý odkaz je najvhodnejší, a či prehľadávajú web správnym smerom (Mat-Hassan et al., 2001). Možno sa prekliká cez 5 alebo 10 odkazov a skončí presne jeden odkaz predtým, než by našiel to, čo hľadal. Čo ak by niekto videl zaňho dopredu? Nasledoval zaňho odkazy a vedel by usúdiť či je daná informácia pre daného používateľa relevantná?

Prehľadávanie do hĺbky je pre tento účel nevhodné, pretože sa môže ľahko dostať do „čiernych dier“, z ktorých niet návratu. Prehľadávanie do šírky je vhodné len pre veľmi malú hĺbku, pretože počet uzlov rastie exponenciálne. Prehľadávanie metódou „najlepší najskôr“ (*best-first*) je možné, ale zlyháva v prípadoch, keď sa na trase medzi relevantnými stránkami nachádzajú iné, menej relevantné, uzly (Wheeldon, 2003). Riešenie podľa Wheeldona predstavuje algoritmus hľadania najlepších stôp (*best trails*). Jednotlivým prepojeniam priradzuje pravdepodobnosti, ktoré zodpovedajú počtu nasledovaní daného odkazu používateľom alebo skupinou používateľov, a tiež reprezentujú očakávanú užitočnosť nasledovania daného odkazu.

Ďalším problémom, ktorý identifikoval Brunk v (Brunk, 1999) je fakt, že výsledky vyhľadávania vo fulltextových vyhľadávačoch často ukazujú na podstránku konkrétneho webového sídla (nie na hlavnú stránku). Toto je výhodné ak hľadájúci používateľ priamo na danej podstránke nájde čo hľadal, ale často sa stáva, že k tomu čo hľadá sa nachádza v skutočnosti na inej podstránke, dostupnej napríklad z hlavnej stránky webového sídla, avšak k tejto nemá používateľ v momente, keď sa dostane na danú podstránku priamy prístup. Mnohé stránky toto riešia pridávaním odkazu na hlavnú stránku na každej podstránke. Pridanie navigačného menu alebo zavedenie navigačných prvkov založených na „odrobinkách“ (*breadcrumbs navigation*), kedy je na každej podstránke zrejماً postupnosť krokov z hlavnej stránky, je tiež čiastočným riešením tohto problému.

Princíp nasledovania odkazov a skokov medzi webovými zdrojmi ostáva rovnaký bez ohľadu na konkrétnu webovú stránku, avšak všetko ostatné závisí od špecifik dizajnu konkrétneho webového sídla a je premenlivé. Farby, rozloženie menu, celkové rozloženie stránky, animácie, formuláre a applety len diverzifikujú a komplikujú prehľadávanie na webe (Brunk, 1999). Vyhľadávače sú užitočné vo výbere potenciálne relevantnej množiny stránok a ich čiastočnom usporiadaní vzhľadom na používateľov dopyt, ale výber z výsledkov a v konečnom dôsledku samotné prehľadávanie ostáva na

používateľovi. Jediným spôsobom, ako si vybrať z výsledkov vyhľadávania fulltextového vyhľadávača, ktorý odkaz nasledovať, je bez využitia nástroja pre podporu navigácie rozhodnutie na základe názvu stránky, jej domény a krátkeho výseku okolia nájdeného dopytu (ktorý sa niekedy vo výsledku vôbec nenachádza).

Mat-Hassan a Mark Levene vykonali štúdiu, v ktorej hľadali odpoveď na otázku, či pridanie poloautomatického prvku pre podporu navigácie na webovom sídle zlepšuje vyhľadávanie z pohľadu používateľov (Mat-Hassan et al., 2001). Výsledky štúdie potvrdili hypotézu, že pri použití navigátora *NavZone* používatelia našli hľadané informácie zo sady testovacích úloh oproti použitiu fulltextového vyhľadávača (Google) a stránkovo špecifického (*site-specific*) vyhľadávača (Compass) rýchlejšie a presnejšie (našli správnu odpoveď na vyšší počet otázok).

## 2.1 Navigácia v dobre definovaných priestoroch

Za dobre definované priestory budeme považovať priestory reprezentovateľné grafom s jasne definovanou sémantikou prepojení. Na rozdiel od webu, kde sémantika prepojení typu hypertextový odkaz nie je vôbec definovaná, v prípade metadát webu so sémantikou má každé prepojenie definovaný význam. Sémantika jednotlivých prepojení je definovaná taktiež v uzavretých úzko špecifických systémoch, akými sú napríklad výučbové webové systémy, kde je možné odvodiť sémantiku prepojenia ako napríklad zobrazenie ďalšej položky kurzu alebo zobrazenie zoznamu položiek kurzu.

Podporou navigácie v hypermediálnych systémoch, ktoré môžeme považovať za dobre definované priestory, sa zaoberá napríklad Brusilovsky. V (Brusilovsky, 2007) identifikuje nasledovné techniky pre prispôsobovanie navigácie:

- **Usporiadanie odkazov** – odkazy na danej stránke sú zoradené na základe dát v modeli používateľa. Čím vyššie sa daný odkaz nachádza, tým relevantnejší je pre daného používateľa. Štúdia (Kaplan, 1993) ukázala, že usporiadanie odkazov má signifikantný vplyv na efektívnosť navigácie (rýchlosť nájdenia hľadaných informácií).
- **Skrývanie odkazov** – princíp spočíva v obmedzení navigačného priestoru skrytím alebo znefunkčnením odkazov na (momentálne) málo relevantné stránky. Tento prístup často používatelia pri experimentoch kritizujú, okrem prípadov, keď skrývanie prebiehalo dekrementálne, t.j. na začiatku skryté odkazy sa používateľom postupne odkrývali (Brusilovsky, 2007).
- **Anotácia odkazov** – odkazy sú obohatené anotáciou, prostredníctvom ktorej používateľ dostáva dodatočné informácie o tom, čo nasleduje po nasledovaní daného odkazu. Forma anotácie môže spočívať v zmene fontu, farby odkazov alebo pridaní dodatočných ikon.
- **Priame smerovanie (*direct guidance*)** – systém ponúka nasledujúci najvhodnejší uzol na základe odporúčacieho mechanizmu. Používateľ kliká na prepojenie „ďalšia stránka“ a systém rozhoduje zaňho, ktorá stránka sa mu zobrazí ako ďalšia v poradí.

- **Generovanie odkazov** – vytvára nové odkazy na stránke. Odkazy sa získavajú odvodením podobnosti medzi dokumentmi a pridajú sa do jednotlivých stránok natrvalo, alebo sa generujú na základe vypočítanej podobnosti stránok. Najzaujímavejší je prístup dynamického generovania odkazov na základe histórie navigácie a modelu používateľa.

Za dobre definovaný priestor môžeme považovať aj priestor metadát na webe so sémantikou, kde prepojenia predstavujú RDF trojice. Aj pri navigácii v priestore metadát na webe so sémantikou ostáva prítomný problém straty širšieho prehľadu o okolí, používateľ stále nemá ani pri použití sofistikovaných nástrojov pre prehliadanie webu so sémantikou, ako napríklad nástroj Tabulator (Berners-Lee, 2006), prehľad o bezprostrednej histórii navigácie. Aj keď navigačné kroky nemusia predstavovať nasledovanie odkazov, ale môžu byť reprezentované aj zadávaním explicitných dopytov (napr. vo forme SPARQL dopytov), stále ide z pohľadu používateľa o skoky v informačnom priestore, pri ktorých sa stráca prehľad o lokálnom okolí.

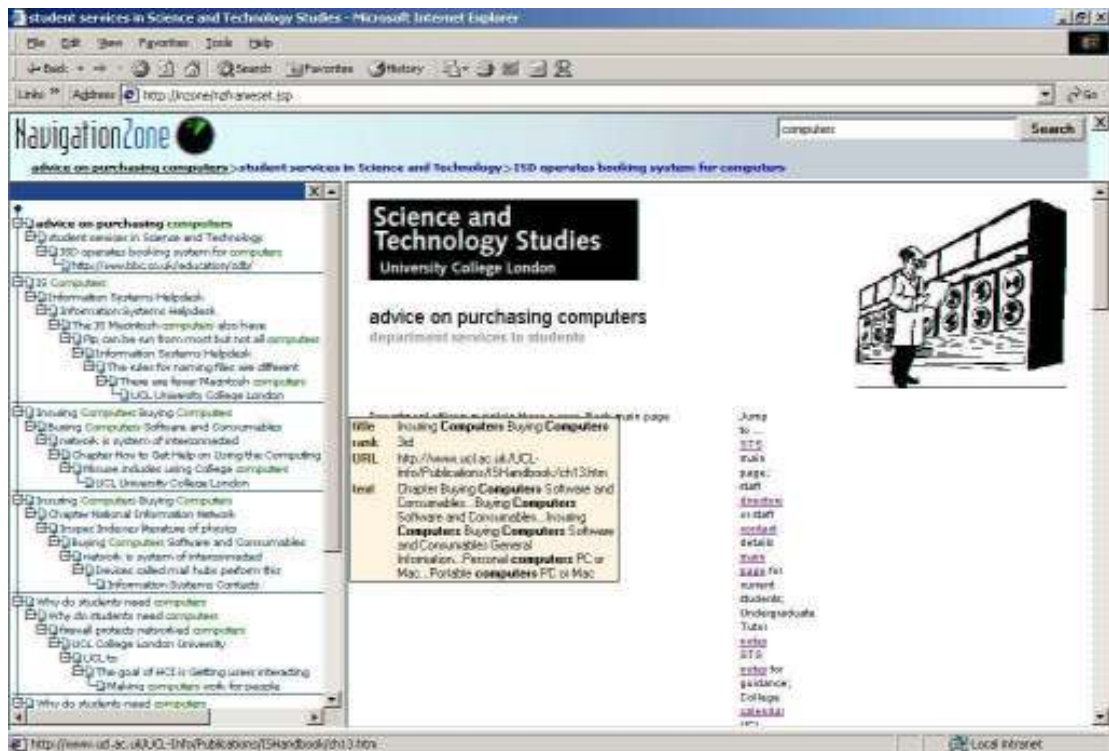
Niektoré webové systémy vyžadujú alebo svojou povahou poskytujú možnosť navigácie vychádzajúcej čisto z povahy doménových dát. Príkladom takéhoto systému môže byť digitálna knižnica. Na digitálnu knižnicu sa dá pozrieť tiež ako na graf (ako v prípade webu), kde uzly predstavujú publikácie, a hrany napr. ich vzájomné citácie. Ďalším možným pohľadom je pohľad cez autorstvo, kde hrany predstavujú publikácie, ktoré majú spoločného aspoň jedného autora.

Priestor webu ako ho poznáme dnes predstavuje otvorený priestor a definícia prepojení existuje len na syntaktickej úrovni hypertextových prepojení. Web a web so sémantikou sú si navzájom podobné v zmysle možnosti ich reprezentácie vo forme orientovaného grafu. Zatiaľ čo v priestore webu sú uzly webové stránky a explicitné hrany hypertextové prepojenia medzi nimi, v priestore metadát na webe so sémantikou sú uzlami zdroje a hranami ich relácie vo forme predikátov, ktoré zachytávajú sémantiku prepojení. Navigáciu v oboch prípadoch je možné zovšeobecniť na graf, kde vrcholy predstavujú navštívené stránky a hrany navigačné kroky, ktorými používateľ prechádza priestorom.

## 2.2 Metódy a nástroje pre podporu navigácie na webe

V tejto kapitole opisujeme vybrané metódy pre podporu navigácie na webe a nástroje, ktoré ich implementujú. Zamerali sme sa najmä na metódy zabezpečujúce podporu navigácie na základe štruktúry prepojení medzi webovými stránkami. Pre analýzu sme vybrali riešenia tak, aby pokrývali viacero prístupov k podpore navigácie. Uvádzame riešenia pre vizualizáciu odkazov (vo forme generátora máp stránok), pre odporúčanie odkazov, fasetový prehliadač s adaptívnymi črtami aj nástroje opierajúce sa o explicitne definované metadáta.

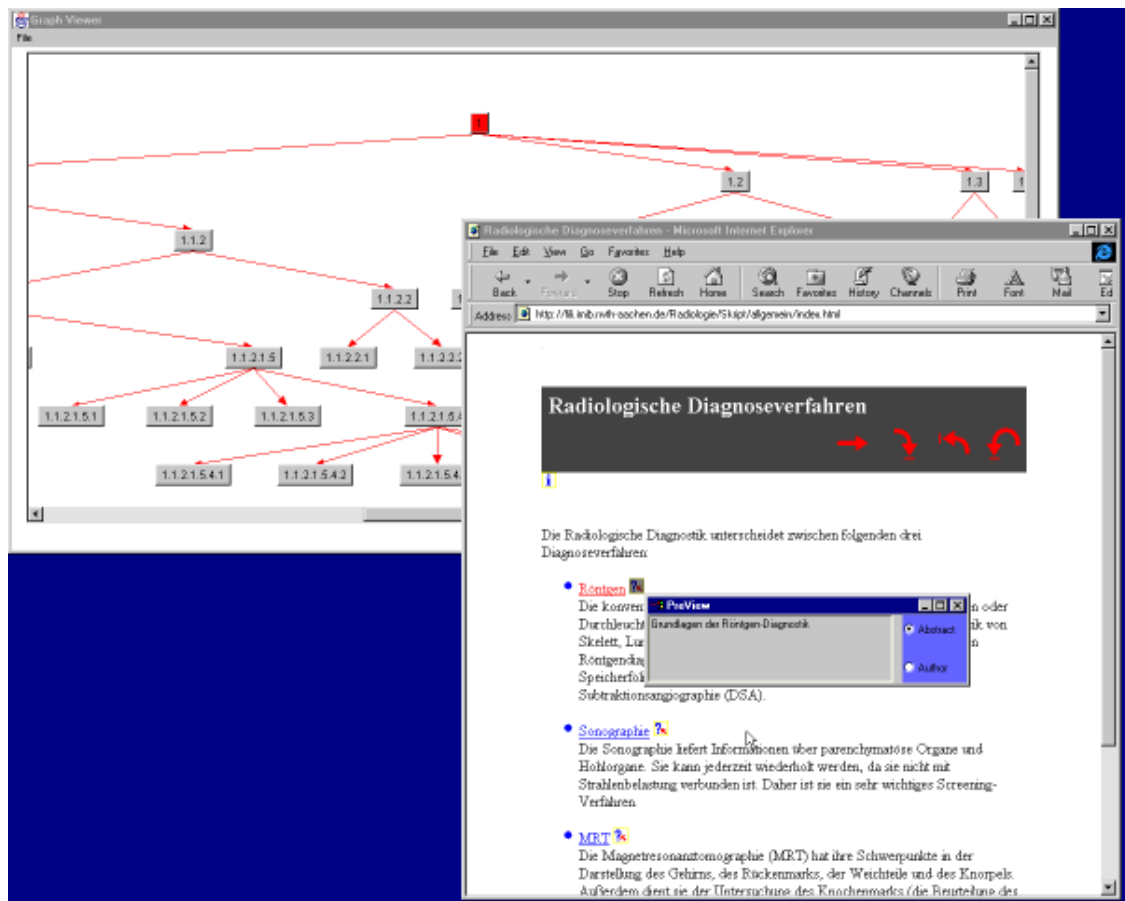
**Web Site Navigation Engine** využíva v navigátore NavZone algoritmus hľadania najlepšej stopy (*best trail algorithm*). Algoritmus je adaptívny zmysle, že dynamicky prehľadáva preferované stopy napodobňovaním správania používateľa a ohodnotením stop v závislosti od topológie webovej stránky (Levene, 2001).



Obrázok 2: Rozhranie nástroja Web Site Navigation Engine.

Rozhranie nástroja *NavZone* využívajúceho *Web Site Navigation Engine* je znázornené na obrázku 2. Navigačná lišta hore obsahuje postupnosť odkazov, ktorých nasledovaním sa používateľ dostal k aktuálnej stránke. Predstavuje to lineárnu bezprostrednú históriu daného sedenia. Aktuálna stránka je v postupnosti podčiarknutá a napravo od aktuálne navštíveného odkazu je zobrazené odporúčanie ďalšej trajektórie prehládavania. Vľavo sa nachádza zoznam ponúknutých trajektórií prehládavania, z ktorých používateľ vyberá kliknutím myšou. Pri zastavení myšou nad odkazom sa zobrazí vyskakovacie okno (*popup*) zobrazujúce názov stránky, jej URL a sumár obsahu danej stránky. Takýto prístup je v kontraste s prísne lineárnym princípom hľadania, aký je aplikovaný v prípade použitia vyhľadávačov ako napríklad Google a klasických webových prehliadačov.

**Graph Viewer.** Kreutz v (Kreutz, 2008) predstavuje návrh princípu prezentácie štruktúry webového dokumentu na princípe vytvárania *N-Context*-ov – množín stránok, ktoré môžu byť dosiahnuté z momentálnej stránky *N* krokmi. Výsledky zobrazuje ale v pomerne nešikovnej forme, ako znázorňuje Obrázok 3.



Obrázok 3: Rozhranie nástroja Graph Viewer pre generovanie štruktúry webového dokumentu.

Aj vzhľadom na technologickú nezrelosť sú možnosti praktického využitia nástroja veľmi obmedzené. Zaujímavá je však idea prehliadania priestoru do hĺbky.

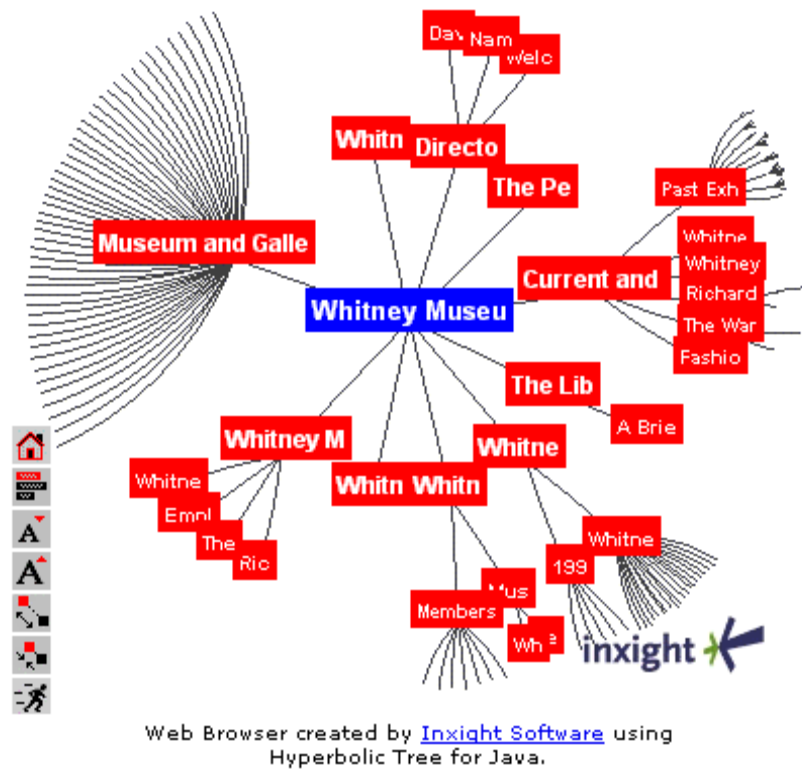
**Applet VizControls.** Brunk v (Brunk, 1999) opisuje nástroj, ktorý využíva pri zobrazovaní okolia stránky vizualizáciu formou hyperbolického stromu. Ide o takmer 10 rokov starý produkt, ktorý bol svojho času ponúkaný a predávaný administrátorom webových stránok. Títo museli manuálne vytvoriť mapu svojej stránky, ktorú potom tento nástroj prezentoval návštevníkovi stránky. Tento nástroj teda nefungoval ako univerzálna podpora navigácie na strane klienta, ale ako generátor máp stránok na strane servera.

Ako uvádza Brunk, hlavnou výhodou použitia hyperbolického stromu pre zobrazenie je, že umožňuje reprezentovať všetky podstránky veľmi veľkého webového sídla prehľadným a jednoducho navigovateľným spôsobom. Hlavnou nevýhodou nástroja bol okrem možnosti nasadenia len na serveri obsah vizualizácie uzlov. Ako je zrejmé z obrázka, nástroj neobohacuje vrcholy o žiadne metadáta a zobrazuje len opisy vrcholov, ktoré navyše v grafe nie sú pre navigáciu veľmi užitočné.

Ukážka rozhrania je zobrazená na obrázku 4. Vrcholy grafu predstavujú webové stránky. Dvojitým kliknutím na vrchol grafu sa zobrazí stránka v separátnom okne. Dnes už nástroj nie je na webe dostupný a spoločnosť ponúka produkty z úplne inej



oblasti. Nástroj je však zaujímavý formou zobrazenia a ideou vizualizácie štruktúry prepojení medzi stránkami. Jeho hlavnou nevýhodou je možnosť použitia len na strane servera a teda nemožnosť aplikovať ho ako podporu navigácie medzi webovými sídlami.



Obrázok 4: Applet VizControls zobrazujúci hyperbolický strom okolia stránky.

Žiadny zo spomínaných nástrojov, kladúcich si za cieľ podporu navigácie vizualizáciou prepojení medzi stránkami neposkytuje informáciu o spätnom odkazovaní medzi webovými sídlami. A pritom, vzhľadom na jednosmernú povahu hypertextového odkazu, môžu byť odkazy, ktoré na danú stránku odkazujú, rovnako relevantné ako tie, na ktoré odkazuje samotná stránka.

**Grouplab Internet History System.** Kaasten v (Kaasten, 2000) predstavuje rozšírenie mechanizmu histórie webového prehliadača, ktoré adresuje problém histórie založenej na zásobníku zabezpečenej pomocou tlačidiel *Back* a *Forward* v dnešných webových prehliadačoch. Rozšírenie predstavuje zavedenie zoznamy usporiadanej na základe času prístupu (*recency-ordered list*), s integrovanou funkcionalitou zabezpečujúcou lokálne ukladanie záložiek. História sedenia je zobrazená formou panelu nástrojov (*toolbar*) na ľavej strane okna prehliadača. Položky histórie nie sú reprezentované len názvom stránky, ale aj zmenšeným náhľadom zobrazujúcim vyrenderovanú podobu stránky. Ďalšou vlastnosťou nástroja je podpora pre vytváranie dynamických dopytov pre efektívne vyhľadávanie v záložkách.

**Tabulator.** Zatiaľ čo na webe sa naviguje medzi dokumentmi prostredníctvom odkazov, na webe so sémantikou navigácia prebieha medzi konceptmi prostredníctvom vzťahov (predikátov) (Berners-Lee, 2006). Nástroj Tabulator slúži pre prehľadávanie RDF dát na webe. Okrem generického tabulovaného zobrazenia poskytuje niekoľko doménovo špecifických zobrazení, napríklad geografické RDF dáta zobrazuje na mape, alebo časové dáta vie zobraziť vo forme kalendára. Umožňuje vytvárať dopyty pre hľadanie zadávaním príkladov, kedy používateľ označí grafový vzor dopytu klikaním v grafickom zobrazení, alebo zadávať priamo SPARQL dopyty.

Hlavné nedostatky nástroja Tabulator, ktoré identifikoval jeden z jeho autorov, Tim Berners Lee, sú nasledovné:

- žiadna možnosť vrátiť sa späť do predchádzajúceho stavu,
- neprítomnosť kognitívnej mapy,
- navigácia podobná tabuľkovému procesoru.

Ďalšou nevýhodou, ktorú sme identifikovali, je podpora výlučne pre RDF a nízke rozšírenie (nielen) tohto formátu a všeobecne webu so sémantikou pre to, aby bol nástroj v súčasnosti prakticky široko použiteľný. Pri pokusoch o vyskúšanie nástroja<sup>2</sup> sme narazili tiež na problémy s rýchlosťou odozvy a funkčnosťou, nástroj hlásil implementačné chyby<sup>3</sup>, čo však mohlo byť spôsobené nejakým dočasným problémom.

**Factic.** Nástroj Factic (Tvarožek et al., 2007) slúži pre prezentáciu priestorov reprezentovaných ontológiami. Je založený na fazetovej navigácii na základe pohľadov, kedy používateľ naviguje v rámci informačného priestoru výberom jedného alebo viacerých obmedzení z množiny dostupných faziet. Fazeta predstavuje rozmer informačného priestoru, typicky mapovaný na atribút informačného artefaktu. Výsledky hľadania predstavujú vybrané inštancie z informačného priestoru, ktoré spĺňajú vybrané obmedzenia.

Rozhranie nástroja Factic pre navigáciu v priestore pracovných ponúk (Návrat et al., 2006) je znázornené na obrázku 5. V ľavej časti rozhrania sú zobrazené momentálne dostupné fazety a obmedzenia. Používateľ môže fazety rozbaľovať a zbaľovať a použitím tlačidla s ikonou žiarovky vypnuté alebo zapnuté. Momentálne vybrané obmedzenia sú zobrazené vpravo hore.

---

<sup>2</sup> Nástroj Tabulator je dostupný na <http://dig.csail.mit.edu/2005/ajar/release/tabulator/0.8/tab.html>

<sup>3</sup> Vo webových prehliadačoch Microsoft Internet Explorer 7.0, Firefox 2.0 a Opera 9.25

Home Job Offers Factic-TopK Criteria Search Registration User registration About us Help

**Start date**

- August (2)
- December (1)
- November (7)
- October (36)

**Duty location**

- United States (46)

**Offered position**

- Computing professionals (46)

**Offered management level**

- team leader (1)
- worker (1)

**Acquisition date**

- Last 2 Years (6)
- Last 5 Years (6)

**Hours/week**

- 30.0 - 40.0 (14)

**Travelling involved**

- 

**Offered benefits**

- Complete array of benefits (2)
- Comprehensive package (1)
- Dental (1)
- Direct deposit (1)

**Current restrictions**

**Start date:** All > 2005 (46)

**Duty location:** All > World > America > North America (46)

**Offered position:** All > Professionals > Physical mathematical and engineering science professionals (46)

Sort by: Name | Salary | Organization | Region Item per page: 10 | 15 | 25 | 50 | 100

#	Name	Salary	Organization	Region	Rate it!
9,38	<a href="#">PeopleSoft Programmer</a>	53.81	Manpower Professional	Durham	☆☆☆☆☆
9,02	<a href="#">Programmer; VC++ Makefile generation</a>	30.0		Phoenix	☆☆☆☆☆
8,26	<a href="#">Genesis10, Lake Oswego, OR US</a>	35.0	Genesis10	Oregon	☆☆☆☆☆
8,05	<a href="#">Programmer Analyst Needed In Chicago!</a>	45000.0		Chicago	☆☆☆☆☆
8,00		60000.0		Erie	☆☆☆☆☆
7,95	<a href="#">Programmer</a>	29.7		Fort Lauderdale	☆☆☆☆☆
7,59	<a href="#">SAS Programmer</a>	35.0		Philadelphia	☆☆☆☆☆
7,43	<a href="#">Programmer</a>	25.0	Manpower Professional	Panama City	☆☆☆☆☆
7,36	<a href="#">Sr. Oracle P/A with Accounting Bkgrnd</a>	75000.0	Diversified Technical Solutions, Inc.	Summit	☆☆☆☆☆
7,02	<a href="#">Software Programmer; VC++ Makefile generation</a>	30.0	Manpower Professional	Phoenix	☆☆☆☆☆
7,00	<a href="#">Database Programmer- VB6/.net/SQL</a>	23.0	Manpower Professional	Seymour	☆☆☆☆☆
6,96	<a href="#">VB.Net Programmer Analyst</a>	55000.0		Columbia	☆☆☆☆☆
6,82	<a href="#">HMS Associates Of Tri-State Inc.</a>	55.0	HMS Associates Of Tri-State Inc.	New York	☆☆☆☆☆
6,74	<a href="#">ORACLE APPLICATION PROGRAMMER</a>		ATMI	Danbury	☆☆☆☆☆
6,72	<a href="#">HMS Associates Of Tri-State Inc.</a>	65.0	HMS Associates Of Tri-State Inc.	New Jersey	☆☆☆☆☆

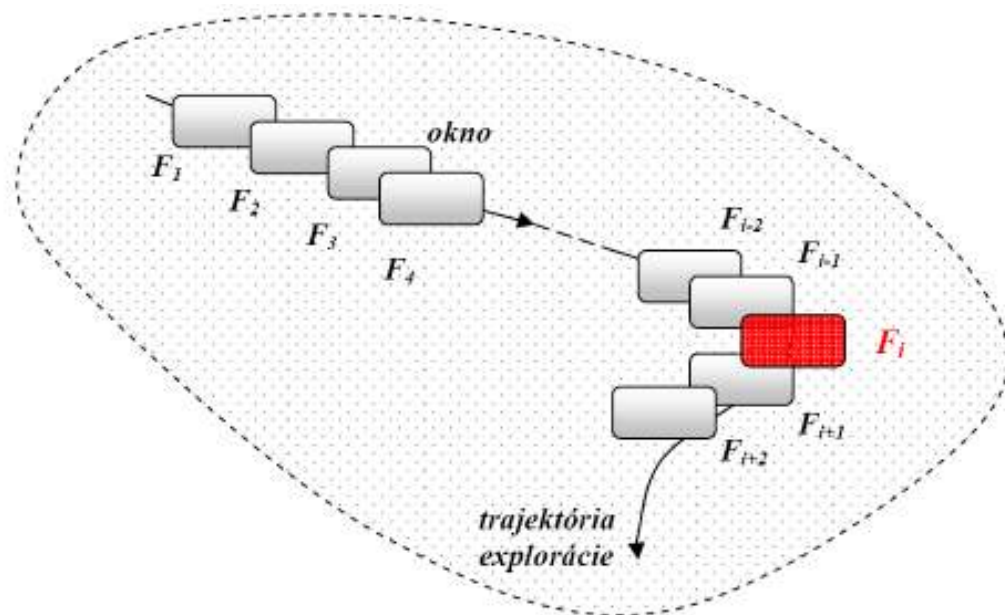
Total Matches: 46 on 4 pages.  
Page: 1 | 2 | 3 | 4 | Next >

Obrázok 5: Grafické rozhranie nástroja Factic.

Nasledovaním odkazov predstavujúcich výber obmedzení je možné obmedzenia zrušiť a rozšíriť množina zobrazených inštancií. Nástroj Factic ďalej zahŕňa integráciu s externými nástrojmi zabezpečujúcimi usporiadanie inštancií, ich hodnotenie a získavanie a odvodzovanie informácií o používateľovi.

Hlavnou výhodou fazetovej klasifikácie, aplikovanej v nástroji Factic, je vysoká efektivita pri vyhľadávaní informácií (Tvarožek, 2006). Dôležitou výhodou nástroja Factic je adaptívne prispôbovanie sa jednotlivému používateľovi výberom faziet a dynamické generovanie faziet. Nevýhodou prístupu je neposkytnutie prehľadu o celom informačnom priestore a z toho vyplývajúce ťažkosti s pochopením priestoru ako celku.

**SemViz.** Michal Jemala vo svojej diplomovej práci navrhuje metódu inkrementálneho prehľadávania, v ktorej je možné identifikovať a oddeliť od vizualizácie proces navigácie, daný prechodom trajektóriou exploračie (Jemala, 2006) (pozri obrázok 6).



Obrázok 6: Princíp inkrementálneho prehľadávania (Jemala, 2006).

Inkrementálne prehľadávanie je postavené na nasledovných tézach (Jemala, 2006):

- v danom momente sa vizualizuje vždy len časť priestoru,
- poskytuje sa viac pohľadov na vizualizovaný priestor,
- používateľ je podporovaný pri navigácii v priestore ontológie nielen na syntaktickej, ale aj na sémantickej úrovni.

Princíp inkrementálneho prehľadávania poskytuje v každom kroku prehľadávania možnosť lokálneho prispôsobenia. Proces riadenia navigácie predstavuje sledovanie a odporúčanie trajektórie explorácie (Jemala, 2006).

Metóda inkrementálneho vizuálneho prehľadávania poskytuje možnosť ohodnocovania konceptov v ontológii na základe významu, čo má mať za cieľ zlepšiť navigáciu. Ohodnocovanie sa deje na základe sledovania používania ontológie nástrojmi, ako aj používateľmi, prehľadávajúcimi štruktúru ontológie. Navigácia je pri metóde inkrementálneho prehľadávania chápaná ako kontinuálny proces, ktorého výstupmi sú jednotlivé odporúčania budúcej trajektórie explorácie (Jemala, 2006).

## 2.3 Zhrnutie

Ako hlavný problém navigácie na webe sa ukazuje povaha webu ako takého, s odkazmi prepájajúcimi nesúvisiace zdroje, bez sémantiky a možnosti udržať pri nasledovaní odkazu kontext. Existujúce nástroje podporujúce navigáciu na webe sú vždy použiteľné len pre podporu navigácie v rámci jedného webového sídla. Navigácia naprieč webovými sídlami je podporená len nástrojmi dostupnými vo webových prehliadačoch – tlačidlá *Back* a *Forward*, história prehliadania a záložky (*bookmarks*).

Paradigmou rozširujúcou lineárny model prehliadania webu sú webové prehliadače s podporou viacerých stránok zobrazených v záložkách (*tabbed browsing*). Ich

praktické používanie však často vedie k ešte väčšej strate prehľadu a kognitívnemu preťažaniu, pretože používateľ nestráca kontext v jednom lineárnom sedení, ale vo viacerých naraz. Možným riešením prevencie straty kontextu pri nasledovaní navigačných krokov je vizualizácia histórie sedenia a okolia momentálne zobrazenej stránky. Aj keď existujú návrhy prístupov pre vizualizáciu okolia momentálne zobrazenej stránky a pre podporu navigácie rozšírením modelu bezprostrednej histórie bežného v dnešných webových prehliadačoch, nepodarilo sa nám nájsť žiadne riešenie, ktoré by pozeralo na problém navigácie ako na celok a riešilo ho zo všetkých relevantných aspektov.

## 3 Personalizácia a sociálne aspekty navigácie

---

### 3.1 Personalizácia a modelovanie používateľa

Základné techniky prispôsobovania predstavujú zmenu obsahu, štruktúry alebo prezentácie tak, aby vyhovovali potrebám jednotlivého používateľa (Paterno, 1999). Z hľadiska uskutočnenia zmeny sa môžeme pozeráť na personalizáciu z dvoch pohľadov:

- zmenu vykonáva systém automaticky (adaptivita)
- používateľ si prispôsobuje zobrazenie sám (adaptabilita)

Pre automaticky vykonávanú personalizáciu je potrebné vytvárať modely používateľa. Model používateľa je možné odvodiť nasledujúcimi spôsobmi a ich kombináciou (Barla, 2006):

- vyplnenie formulárov
- analýza dokumentov
- sledovanie prístupu k zdrojom
- sledovanie správania používateľa

Výstupy dosiahnuté aplikáciou týchto prístupov sa líšia relevanciou a dôveryhodnosťou získaných dát. Zatiaľ čo o správnosti dát získaných sledovaním akcií používateľa nemôžeme mať istotu, a často nemáme pre ne ani interpretáciu v zmysle „na základe čoho bola daná charakteristika odvodená?“, pri explicitnom získaní charakteristík, napríklad formou vyplnenia formulárov, máme pre dáta vždy zdôvodnenie (používateľ odpovedal kladne na položenú otázku) a ich dôveryhodnosť je tiež vyššia. Dôležité je však vždy brať do úvahy kontext, v ktorom sú informácie získavané. Napríklad v prípade, že používatelia sú systémom nútení odpovedať často na otázky týkajúce sa hodnotenia obsahu, je možné, že veľa používateľov neodpovedá na otázky presne a pravdivo, pretože to pre nich predstavuje nechcenú dodatočnú záťaž. Ako výhodná sa ukazuje kombinácia jednoduchých odpovedí typu „Pochopil som“/“Nepochopil som“, „Páči sa mi“/“Nepáči sa mi“ s vynucovaním týchto odpovedí navigačnou schémou systému, ktorá používateľovi zobrazí ďalšiu položku až po zodpovedaní otázky.

Rozlišujeme dva základné typy modelov (Brusilovsky, 1996):

- stereotypný model – sú zadefinované skupiny a používateľ sa zaradí do stereotypu. Systém sa prispôsobuje na základe stereotypu.
- prekryvný model – sú modelované charakteristiky každého používateľa zvlášť, má svoju inštanciu modelu používateľa. Systém sa prispôsobuje na základe modelovaných charakteristík každému používateľovi zvlášť.

Pri použití stereotypov je možné akcie vedúce k prispôbeniu systému presne špecifikovať a definovať. Používateľ môže byť zaradený do stereotypu pomerne

jednoducho zodpovedaním niekoľkých otázok (prípadne sa môže do stereotypu zaradiť sám). Nevýhodou je neprispôsobovanie sa jednotlivému používateľovi.

Na druhej strane, pri použití prekryvného modelu je prispôsobenie do vyššej miery individuálne, avšak definícia akcií pre adaptáciu v systéme môže byť zložitejšia a ich reprezentácia nemusí byť jednoznačná. Inicializácia modelu tiež trvá dlhšie. V počiatočnej fáze, keď model ešte nie je dostatočne nainicializovaný, sa čelí tzv. *cold-start* problému. Preto sa v praxi aplikuje kombinácia oboch prístupov, kedy sa na začiatku aplikuje stereotypný model, a po získaní dostatočného množstva informácií o používateľovi sa adaptácia realizuje na základe prekryvného modelu (Barla, 2006).

V prípade navigácie v otvorených priestoroch je najlepšie použiteľný prístup pozorovania správania používateľa, respektíve skupiny používateľov. Množina prístupov založených na týchto princípoch sa nazýva sociálna navigácia a sociálne hľadanie.

### 3.2 Značkovanie na webe

S nástupom technológií a prístupov známych ako Web 2.0 sa spája okrem iného aj rozšírenie sociálneho kolaboratívneho značkovania a vytvárania verejne prístupných záložiek. Dnes existuje veľké množstvo stránok pre sociálne značkovanie<sup>4</sup>. Jedným z prvých projektov, a dnes možno najrozšírenejším, je služba del.icio.us (<http://del.icio.us>). Použitie značiek je rozšírené okrem aplikácií pre zdieľanie záložiek aj na stránky pre ukladanie fotografií (Flickr), videí (Youtube), organizáciu emailov (Gmail), katalógy hudby (last.fm), blogy a mnohé iné aplikácie.

Sociálne značkovanie je prístup pre ukladanie, organizovanie, hľadanie, správu a zdieľanie záložiek (*bookmarks*) webových zdrojov na webe s použitím metadát. Záložky sú obohatené o metadáta, typicky vo forme značiek, reprezentovaných slovami vybranými používateľmi. Uložené záložky sú zoskupované na základe asociovaných značiek, služby sociálneho značkovania zobrazujú počet použití značky, pri pridávaní záložiek ponúkajú značky najčastejšie asociované s danou URL inými používateľmi a z existujúcich záložiek.

Hlavným obmedzením použitia záložiek je ich sémantická dvojznačnosť vyplývajúca z ich povahy danej použitím slov prirodzeného jazyka. Pri značkách neexistuje žiadna štandardná množina používaných slov (kontrolovaný slovník) ani žiadne štandardy pre štruktúrovanie značiek (napríklad pre rozlíšenie singuláru/plurálu, použitie viacslovných značiek). S použitím značiek je spojené riziko nesprávne vyhláskovaných značiek, preklepov, a predovšetkým nemožnosť rozlíšenia homoným. Taktiež neexistuje možnosť usporadúvať značky do klasifikácie a vytvárať hierarchické vzťahy pre opis nadradenosti/podradenosti pojmov. Aj keď existujú návrhy ako spomínané problémy spojené so značkováním vyriešiť, najmä v podobe vizionárskych článkov (Gruber, 2007), pri dnešnom praktickom využití značiek sú to problémy stále aktuálne a nedoriešené.

---

<sup>4</sup> [http://en.wikipedia.org/wiki/List\\_of\\_social\\_software](http://en.wikipedia.org/wiki/List_of_social_software)

V porovnaní s riešeniami založenými na klasifikácii opierajúcej sa o metadáta reprezentované ontológiami sa môže praktická použiteľnosť značiek ukazovať ako silno obmedzená. Prináša so sebou však hlavnú výhodu, vyplývajúcu z mechanizmu ich vytvárania, kedy relevantné slová pre opis zdrojov vyberá človek, ktorý im na rozdiel od stroja rozumie (aj keď existujú riešenia pre automatizovanú extrakciu značiek z webových zdrojov (Zhou et al., 2008)) a organizuje ich spôsobom, ktorému intuitívne rozumie a vie v takto vybudovanej informačnej báze vyhľadávať.

V súvislosti s rozšírením značiek v aplikáciách pre organizáciu informácií, špeciálne v systémoch pre sociálne značkovanie, daným najmä orientáciou na človeka, ich jednoduchým použitím a prínosom, sa však momentálne ukazujú značky v porovnaní s klasifikáciami založenými na striktno formálnych ontológiách, ako vhodnejší zdroj pre získavanie metadát o webových zdrojoch.

Keďže pre značky chýba sémantická reprezentácia, odvodzovanie a usudzovanie nad značkami nie je formalizovateľné a značkám nie je možné dať formálnu interpretáciu ani pri aplikácii metód spracovania prirodzeného jazyka. Výber značiek jednotlivými používateľmi býva silne individuálny, a často aj iracionálny. Bežné je tiež používanie pragmaticky zvolených značiek pre vlastné použitie<sup>5</sup>, ktoré nemajú relevantnú objektívnu interpretáciu alebo hodnotu pre iných používateľov. Princíp značkovania je potrebné chápať predovšetkým ako orientovaný na koncového používateľa (na rozdiel od ontológií ako základu webu so sémantikou). V porovnaní s ontologickými reprezentáciami je však ich využitie napriek sémantickým obmedzeniam priamočiarejšie.

Značky, ktoré používateľ využíva v systémoch využívajúcich princíp značkovania, najmä v systémoch pre správu záložiek, je možné považovať za aproximáciu jeho záujmov. Vychádzajúc z predpokladu, že používateľ si medzi záložky ukladá najmä webové zdroje, ktoré sú pre neho dôležité a zaujímajú ho, je možné množinu značiek, ktoré používa pre ich opis, považovať za reprezentáciu modelu jeho záujmov. O zdrojoch, ktoré sú opísané rovnakými značkami, je potom možné usudzovať, že sú pre daného používateľa zaujímavejšie ako zdroje popísané inými značkami. Ďalej je možné z použitých značiek odvodiť aj ďalšie charakteristiky používateľa, ako mieru participácie, vedomosti v rámci témy označovaného zdroja, mieru organizovanosti, záujem o danú tému, či kreativitu (Carmagnola, 2007). Carmagnola navrhuje usudzovanie nad značkami na základe ich kategorizácie na základe zdroja, z ktorého pochádzajú (značky vytvorené používateľom, ponúknuté systémom, použité viacerými používateľmi, synonymá). Vzhľadom na kontextovú subjektívnu reprezentáciu značiek však pravdepodobne nebude nikdy možné strojovo ich spoľahlivo interpretovať ani s použitím vyspelých technológií spracovania prirodzeného jazyka a pri využití značiek

---

<sup>5</sup> Napríklad verejné záložky autora diplomovej práce vzťahujúce sa k počiatočnému výskumu zdrojov na internete označené značkou „diplomovka“ (prístupné na <http://del.icio.us/462735/diplomovka>) a verejné záložky všetkých používateľov označené danou značkou (prístupné na <http://del.icio.us/tag/diplomovka>) nemajú kvôli pragmatickému výberu značky „diplomovka“ a jej relatívnej interpretácii objektívne nič spoločné.



pre usudzovanie treba brať do úvahy ich relatívne obmedzenú spoľahlivosť a vypovedaciu hodnotu.

### Programové rozhrania služby del.icio.us

Spomedzi obrovského množstva služieb, ponúkajúcich správu záložiek s podporou ich organizácie formou zdieľaných značiek, sme vybrali službu del.icio.us, ktorá je pravdepodobne najstaršou a momentálne najpoužívanejšou službou pre sociálne značkovanie. Analyzovali sme programové rozhrania služby, ich formy a možnosti, ktoré poskytujú.

Služba del.icio.us ponúka niekoľko aplikačných programových rozhraní (API) v nasledovných formách:

- RSS – *feedy* pre históriu záložiek pre dané URL, použitie jednotlivých značiek, značky jednotlivých používateľov
- API – programové rozhranie služby prístupné cez HTTP protokol pre manipuláciu so značkami
- HTML – rozhranie služby pre pridávanie obsahu z del.icio.us priamo do webových stránok
- JSON – rozhranie prístupné cez javascript pre manipuláciu so značkami

Základnou formou je služba nazývaná „del.icio.us API“, prístupná cez HTTP protokol. Poskytuje nasledovné metódy:

<b>Názov metódy</b>	<b>Opis</b>
---------------------	-------------

<i>Update</i>	Vráti čas poslednej zmeny v záložkách
---------------	---------------------------------------

<i>Tags</i>	Vráti zoznam použitých značiek s počtom ich výskytov v záložkách prihláseného používateľa, umožňuje ich premenovať
-------------	--

<i>Posts</i>	Vráti zoznam všetkých záložiek, umožňuje manipuláciu so záložkami (pridávanie, mazanie)
--------------	---

<i>Bundles</i>	Vráti zoznam definovaných zväzkov značiek ( <i>tag bundles</i> ), umožňuje manipuláciu s nimi
----------------	---

Jednotlivé formy prístupu cez API neposkytujú jednotnú množinu metód. JSON rozhranie poskytuje navyše nasledovné metódy:

<b>Názov metódy</b>	<b>Opis</b>
---------------------	-------------

<i>URL</i>	Vráti detaily pre konkrétne URL
------------	---------------------------------

<i>Network</i>	Vráti zoznam mien pridaných do siete aktuálneho používateľa
----------------	---

<i>Fans</i>	Vráti zoznam používateľov, ktorí si aktuálneho používateľa pridali do siete
-------------	---

Z metód prístupných cez JSON je zvlášť zaujímavá metóda URL, poskytujúca okrem iného zoznam najčastejších značiek, ktorými bolo dané URL všetkými používateľmi označené, spolu s počtom použití danej značky. Potenciálne zaujímavé môžu byť aj metódy *Network* a *Fans*, s pomocou ktorých sa dajú získať informácie o sociálnych sieťach.

Vďaka opísaným rozhraniám sa služba del.icio.us ukazuje ako vhodná pre využitie pre získavanie dodatočných metadát vo forme značiek generovaných priamo používateľmi. Výsledky analýzy pokrytia priestoru webových stránok značkami na del.icio.us uvádzame v kapitole 7.2.

### 3.3 Adaptívna sociálna navigácia

Dnes sú už pomerne efektívne zvládnuté prístupy k podpore navigácie v hypermediách založené na uzavretom korpuse (*closed corpus*). Priestor dokumentov v systémoch, pre ktoré sú navrhnuté metódy podpory navigácie určené, je uzavretá ohraničená množina, navyše obohatená doménovými expertmi o rôzne druhy metadát. Klasické adaptívne hypermediálne systémy nie sú použiteľné pre otvorený korpus (*open corpus*), akým je napríklad web (Brusilovsky, 2007), prípadne sú použiteľné len s obmedzeniami, napríklad v systémoch *KBS-Hyperbook* alebo *SIGUE* je potrebné všetky pridávané zdroje manuálne indexovať. V projekte *Knowledge Sea* (Brusilovsky, 2002) autori navrhujú metódu automatického zoskupovania podobných dokumentov na webe z domény výučbových kurzov v jazyku C s využitím Kohonenových samoorganizujúcich sa neurónových sietí. Jednotlivé zdroje sú zoskupené na mriežke 8x8 na základe výskytu kľúčových slov a jednotlivé bunky v mriežke sú navyše zafarbené v závislosti od počtu zdrojov v jednotlivých bunkách mriežky. Z povahy použitej neurónovej siete vyplýva, že podobné zdroje sa nachádzajú v mriežke blízko seba, čo má podporovať navigáciu. Nevýhodou systému, ktorú identifikoval Brusilovsky v (Brusilovsky, 2004) je, že v jednotlivých bunkách sa zoskupili zdroje s rôznou kvalitou a systém nepodporoval navigáciu v rámci buniek, teda medzi kvalitnými a menej hodnotnými zdrojmi nemohli používatelia jednoducho rozlíšiť.

Riešenie, ktoré navrhuje Brusilovsky, nazval sociálna adaptívna podpora navigácie. Prístup sa spolieha na kolektívnu znalosť veľkého množstva používateľov. Vychádza zo známeho sociálneho fenoménu: ľudia majú tendenciu pri navigácii sledovať stopy iných ľudí (Brusilovsky, 2004). Tento fenomén sa využíva v systéme *Knowledge Sea II*, ktorý rozširuje prístupy v systéme *Knowledge Sea* o vizualizáciu počtu návštev jednotlivých buniek obsahujúcich podobné dokumenty. Sýtosť farby bunky predstavuje počet návštev používateľmi v rovnakej používateľskej skupine ako je skupina, do ktorej patrí aktuálny používateľ. Zároveň bol aplikovaný princíp anotácie odkazov informáciou o počte návštev daného uzla aktuálnym používateľom. Takto si používateľ vytvára prehľad o tom, ktoré časti priestoru má preskúmané dobre, a ktoré ešte nevidel alebo videl malý počet krát.

Tento prístup môže byť aplikovaný aj v prípade webu, ak vychádzame z predpokladu, že stránky, ktoré sú hodnotnejšie, majú vyššiu návštevnosť. Výhody tohto prístupu sú, že je samoorganizujúci, teda nie je potrebný zásah expertov ani anotácia alebo iná príprava informačného priestoru a je použiteľný pre otvorené priestory, akým je web.

## 4 Otvorené problémy a ciele práce

---

V predchádzajúcich kapitolách sme identifikovali hlavné problémy navigácie na webe a ich príčiny. Problém „straty v hyperpriestore“, prejavujúci sa stratou prehľadu o okolí momentálne zobrazenej stránky, a najmä rýchlou stratou prehľadu o navštívenej histórii, je spomínaný v mnohých výskumných prácach. Autori v (Theng, 2007) dokonca poukazujú na matematický dôkaz nevyhnutnosti existencie problému straty v hyperpriestore analógiou s turingovým strojom. Ukazujú, že funkcia navigácie v hypertexte je nevypočítateľná, a z toho dedukujú, že strácanie sa v hyperpriestore nie je ľudský psychologický problém, ale že v hyperpriestore by sa stratil akýkoľvek stroj fungujúci na algoritmickom princípe. Web prirovnávajú k dátovej štruktúre, avšak poukazujú na fakt, že nad dátovou štruktúrou webu nie sú dobre definované navigačné operácie (Theng, 2007). Používateľovi, ktorý nemá žiadnu podporu rozhodovania pri klikaní na odkazy, neostáva iná možnosť, ako „klikat’ tak, ako ho práve napadne“, čo má skôr či neskôr za následok stratu prehľadu.

Existujúce v praxi bežne používané prístupy aplikované návrhármi webových sídiel sa logicky sústreďa len na zlepšenie navigácie v rámci jedného webového sídla. Problém straty prehľadu býva riešený pridaním navigačných prvkov do stránky, s pomocou ktorých sa môže používateľ v prípade, že má pocit stratenia sa, vrátiť na dobre definované miesto webového sídla. Podpora navigácie je teda efektívne riešená len na úrovni jednotlivých webových sídiel. Bežne však navigácia prebieha cez hranice webových sídiel. Pre navigáciu naprieč rôznymi webovými sídlami neexistuje pri použití analyzovaných prístupov adekvátne podpora.

Stratu prehľadu o bezprostrednej histórii navštívených stránok riešia webové prehliadače čiastočne možnosťou vracať sa v krokoch s použitím tlačidla späť. Rozšírením je ukladanie histórie navštívených stránok. Zaznamenaná história je však použiteľná najmä pre archivačné, nie však navigačné účely.

Podpora RDF dát sa ukazuje dnes pre vyhľadávanie informácií na webe v súčasnom stave webu pre všeobecné použitie ako málo využiteľná, vzhľadom na momentálne nízke množstvo stránok, ktoré by publikovali metadáta vo forme RDF.

Dopredné prehľadávanie informačného priestoru spolu s ohodnocovaním relevancie trás sa potvrdzuje aj na základe výsledkov štúdie (Mat-Hassan, 2001) ako efektívny prístup. Ďalším prístupom sú metódy pre odporúčanie odkazov, podporujúce navigáciu v hypermediálnych systémoch (Romero, 2007), vždy sú však problémovo orientované, a sú zamerané na dobre definované a ohraničené priestory.

Aj keď existuje niekoľko rozličných prístupov pre adaptívnu podporu navigácie v hypermediách založených na usporiadaní, anotácii, skrývaní a generovaní odkazov (Brusilovsky, 2007), všetky zdieľajú základný princíp adaptácie prezentácie odkazov obsiahnutých priamo v hypertextovej stránke. Takáto forma prispôsobenia má za cieľ brať do úvahy ciele, znalosti a preferencie jednotlivých používateľov, avšak vyžaduje dobre známy informačný priestor a väzbu na dobre opísanú problémovú doménu.

Taktiež sa predpokladá explicitný model používateľa a jeho naplnenie. Riešenia, podporujúce navigáciu v otvorených priestoroch, sa spoliehajú na sociálnu navigáciu a sociálne hľadanie (Brusilovsky, 2004) a bývajú založené na sledovaní správania sa skupiny používateľov. Takýto prístup má niekoľko výhod, ako najdôležitejšiu v prípade jeho aplikácie pre navigáciu na webe sme identifikovali vlastnosť samoorganizácie, kedy je systém schopný prispôbovať sa len na základe sledovania správania používateľov.

Na základe identifikovaných problémov a výsledkov analýzy sme definovali nasledovné ciele práce. Prvoradým cieľom je navrhnúť metódu podpory navigácie v otvorenom priestore webu, použiteľnú pri prehliadaní webu naprieč rôznymi webovými sídlami, s prihliadnutím na nasledovné aspekty:

1. Zohľadnenie štruktúry a vlastností prepojení medzi webovými stránkami.
2. Podpora zachovania kontextu pri nasledovaní odkazov na webe.
3. Adaptívne zohľadnenie sociálneho správania sa používateľov na webe s využitím existujúcich globálnych zdrojov metadát (napríklad verejne prístupných označovaných záložiek) pre prispôbovanie.

Za účelom overenia navrhutej metódy vyvineme nástroj pre podporu navigácie v otvorenom priestore webu, ktorý má tieto vlastnosti:

1. Efektívne zobrazenie okolia momentálneho uzla grafu (webovej stránky), vyhodnotenie a prispôbenie zobrazenia s využitím navrhutej metódy adaptívnej podpory navigácie
2. Podpora navigácie obohatením zobrazenia o bezprostrednú históriu prehliadania
3. Použiteľnosť nástroja aj pre prehliadanie iných priestorov reprezentovaných grafmi.

## 5 Metóda adaptívnej navigácie v otvorenom priestore

---

V tejto kapitole predstavujeme návrh metódy vychádzajúcej z metódy inkrementálneho prehľadávania, ktorá slúži pre navigáciu v otvorenom priestore definovanom grafom. Navigácia je založená na postupnom rozvíjaní uzlov grafu, teda nie je potrebné dopredu poznať celý graf, čo by ani v prípade aplikácie metódy na navigáciu na webe nebolo prakticky možné. Rozvíjanie je dané funkciou susednosti vrcholov v prehliadanom grafe. Pre prípad webu môže byť triviálna forma tejto funkcie definovaná množinou všetkých odkazov, na ktoré sa momentálna stránka odkazuje, respektíve ktoré sa odkazujú na danú stránku. Rozvinuté uzly (možné nasledovníky) predstavujú usporiadanú množinu. Usporiadanie sa vykoná na základe funkcie počítajúcej relevanciu hrán vychádzajúcich z aktuálneho vrcholu k jeho susedom. Reprezentáciou hodnoty relevancie je potenciálna zaujímavosť danej stránky pre momentálneho používateľa vzhľadom na momentálne zobrazenú stránku.

V nasledujúcej časti uvádzame všeobecnú schému navrhovanej metódy pre navigáciu v ľubovoľnom grafe s predpokladom dodefinovania metrik a konštánt, a ďalej špecializáciu metódy pre navigáciu v priestore webu.

### 5.1 Schéma metódy

Prehliadaný graf nie je dopredu známy. Je definovaný susednosťou uzlov, a táto sa generuje inkrementálne. Na začiatku navigácie sa špecifikuje počiatkový uzol grafu. V každom kroku navigácie sa generujú možné nasledovníky aktuálneho uzla. Vyhodnotia sa metriky ohodnocujúce relevanciu hrán. Váženou sumou je hodnota metrik agregovaná pre každý uzol. Uzly sa usporiadajú a filtrujú na základe vypočítanej hodnoty relevancie. Pri počítaní relevancie môžu byť brané do úvahy metriky, ktoré určujú hodnotu relevancie aj s ohľadom na momentálneho používateľa.

Nasleduje schéma metódy zapísaná v pseudokóde:

```

Navigácia(e : hrana)
  Vstup: hrana vybraná používateľom e = (v1, v2)
  1. Pridaj v1 do histórie
  2. H = Generuj okolie(v2)
  3. Pre všetky hrany h v H: Ohodnoť(h)
  4. H = Top-n(H)
  5. Usporiadaj(H)
  Výstup: nové zobrazenie
```

Funkcia generovania okolia vráti zoznam hrán, ktoré spájajú momentálny uzol so susediacimi uzlami v prehliadanom grafe:

```

Generuj okolie(v : vrchol)

  Vstup: vrchol v

  Výstup: množina hrán vychádzajúcich vrcholu v
```

Funkcia ohodnotenia hrán vypočíta relevanciu pre každú hranu na základe hodnoty definovaných metrík:

```
Ohodnotenie hrán(h : množina hrán)
  Vstup: množina hrán
  Pre všetky hrany: ohodnotenie(h) = vážená suma definovaných
metrík
  Výstup: usporiadaný zoznam hrán
```

Funkcia filtrovania odstráni z množiny hrán tie hrany, ktoré boli ohodnotené relevanciou nižšou, ako je hraničná hodnota:

```
Top-n(H : množina hrán)
  Vstup: množina hrán pre filtrovanie
  Výstup: n hrán s najvyšším ohodnotením
```

## 5.2 Aplikácia metódy pre navigáciu na webe

Zaoberali sme sa aplikáciou navrhutej metódy prehliadania grafu pre navigáciu na webe, kde navrhujeme formu zobrazenia v prípade použitia pre prehliadanie webových stránok, abstrakciu procesu navigácie na graf a špecializáciu metódy návrhom funkcie generovania okolia susedných vrcholov v grafe a ohodnocovania ich relevancie.

### 5.2.1 Abstrakcia navigácie na webe na graf

Navigáciu na webe je možné zovšeobecniť na graf, v ktorom hrany predstavujú navigačné kroky konkrétneho používateľa a vrcholy webové zdroje, ktoré navštívil. Vzhľadom na povahu prepojení na webe, kedy stránky, ktoré sa odkazujú na danú stránku, môžu byť rovnako relevantné ako stránky, na ktoré sa odkazuje, je vhodné brať pri generovaní možných nasledovníkov v grafe do úvahy aj prichádzajúce odkazy. Identifikovali sme nasledovné typy vrcholov a hrán:

Vrcholy: webové stránky

Hrany:

- Odchádzajúce odkazy (href)
- Prichádzajúce odkazy (href)
- Odporúčané hrany z dopytu google použitím operátora related<sup>6</sup>

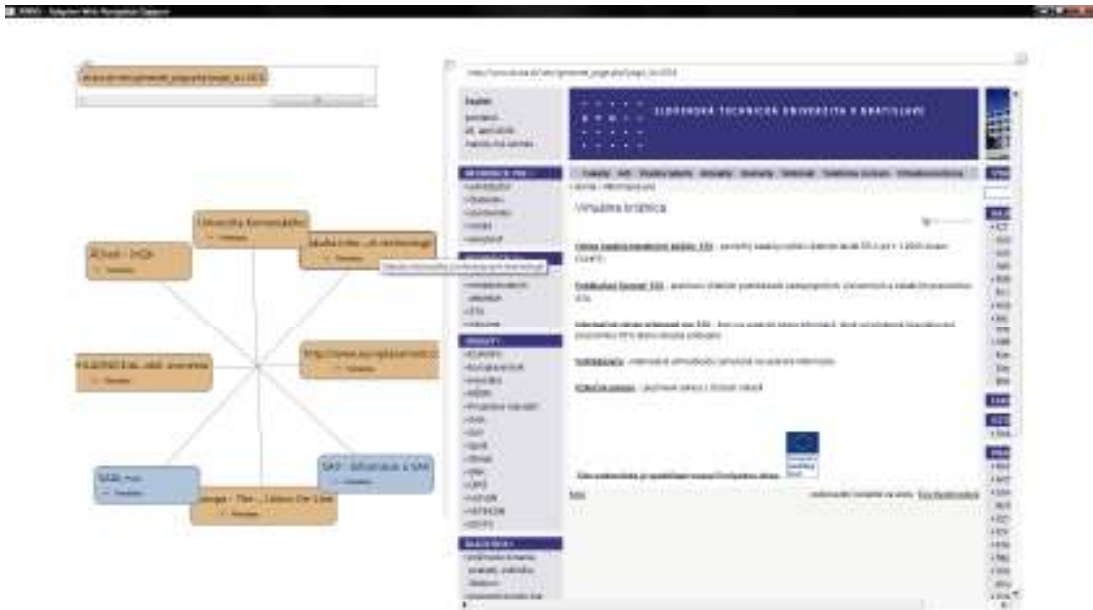
### 5.2.2 Forma zobrazenia

Na obrázku 7 sa nachádza návrh obrazovky rozhrania pre aplikáciu metódy pre navigáciu na webe. V ľavej časti je zobrazený graf, obsahujúci vrchol predstavujúci momentálne zobrazenú stránku a jeho okolie. Tiež zachytáva časť histórie

---

<sup>6</sup> „related:“ je kľúčové slovo vyhľadávača Google Search, ktoré zabezpečí vrátenie stránok podobných stránke špecifikovanej ako argument.

momentálneho sedenia vo forme grafu predstavujúceho postupnosť navštívených stránok.



Obrázok 7: Navrhnuté rozhranie demonštrujúce princíp podpory navigácie.

Pravá horná časť obsahuje riadok s adresou, kde používateľ špecifikuje počiatočnú stránku navigácie. V pravej časti sa nachádza zobrazenie momentálnej stránky, zatiaľ čo v ľavej časti je zobrazený graf predstavujúci jej okolie (vrcholy generované funkciou generovania okolia a hrany, ktoré ich spájajú). Ponúka sa zobrazenie okolia momentálne zobrazenej stránky. Ďalej môže ponúkať medzi nasledovníkmi aj stránky získané z väčšej hĺbky v grafe, v prípade, že ich vyhodnotí ako dostatočne relevantné. Kontext je zachytený vizualizáciou histórie aktuálneho sedenia formou zoznamu usporiadaného na základe času poslednej návštevy, ktorý sa v grafickom rozhraní nachádza vľavo hore.

Používateľ môže prehliadať web buď spôsobom bežným vo webových prehliadačoch, klikaním na odkazy priamo v obsahu zobrazenej stránky, alebo klikaním do grafu vľavo. Výberom položky z histórie sedenia sa môže vrátiť v grafe histórie prehliadania na stránku, ktorú v danom sedení navštívil jeden alebo viac navigačných krokov dozadu.

Pri prehliadaní webu sú uzly grafu vzhľadom na technologické obmedzenia externých služieb (služby del.icio.us, ktorá nepovoľuje vykonanie požiadavky častejšie ako jeden krát za sekundy) pridávané inkrementálne. V zobrazení sa nachádza maximálne špecifikovaný počet uzlov (8), pričom v procese inkrementálneho ohodnocovania sú uzly s nižšími hodnotami relevancie postupne zamenené za uzly ohodnotené vyššie.

Paralelne s vyhodnotením relevancie odkazov sú získavané metadáta o uzloch reprezentujúcich webové stránky (počet návštev daného uzla momentálnym používateľom a celkovo, značky, ktorými je dané URL označené na del.icio.us). Taktiež sú inkrementálne získavané názvy stránok (HTML title) a použité ako textové nadpisy

uzlov, ktoré jednotlivé stránky reprezentujú. Dlhé názvy stránok sú skrátené odstránením prefixu a sufixu obsahujúceho názov domény (napríklad “SME.SK | Veľmi dlhý názov stránky obsahujúci naozaj veľa slov” sa zmení na “Veľmi dlhý... veľa slov”). Pri zotrvaní myšou nad uzlom je zobrazený celý názov stránky.

Uzly sú rôznych farieb, v závislosti od typu hrany, ktorá do nich vedie. Uzly získané z dopytu služby Google Search „related:“ sú zobrazené ako svetlo modré, odchádzajúce uzly hnedé a prichádzajúce svetlo hnedé.

### 5.2.3 Vyhodnotenie relevancie hrán

Pre vyhodnotenie relevancie hrán pri podpore navigácie na webe sme navrhli niekoľko metrík. Pri ich návrhu vychádzame z predpokladu dostupnosti značiek popisujúcich webové stránky. Ďalšie navrhnuté metriky využívajú princíp sociálnej navigácie a predpoklady o vyššej návštevnosti všeobecne kvalitnejších a pre používateľa potenciálne zaujímavejších stránok.

#### Relevancia na základe podobnosti stránok

Vychádzame z predpokladu, že na zabránenie straty prehľadu o okolí nečakanými kontextovými skokmi, by používateľ mal uprednostňovať také navigačné kroky, ktoré ho zavedú na stránky podobné momentálne navštívenej stránke. Za predpokladu, že momentálna stránka je preňho zaujímavá, budú preňho pravdepodobne zaujímavé aj podobné stránky. Taktiež ak používateľ sleduje prehliadaním webu nejaký konkrétny cieľ, je vyššia pravdepodobnosť jeho dosiahnutia, ak sa bude pohybovať medzi podobnými stránkami. Preto jednou z metrík relevancie označíme podobnosť stránok.

Relevancia(hrana(url1, url2)) ~ podobnosť(url1, url2)

Definujme metriku pre určenie podobnosti dvoch stránok:

```
Podobnosť(url1, url2) : <0..1>
```

```
Nech t1 = značky(url1), t2 = značky(url2)
```

```
S = |t1| + |t2|
```

```
H = 0
```

```
Pre všetky spoločné značky sz = prienik(t1, t2)
```

```
H += (Počet výskytov(sz, t1) + Počet výskytov(sz, t2))
```

```
Podobnosť = H / S
```

Takto definovaná metrika berie do úvahy počet použítí jednotlivých značiek asociovaných s dvojicou webových zdrojov. Predpokladáme globálne dostupný zdroj značiek vytvorený veľkou množinou používateľov a vychádzame pritom z predpokladu, že značka, ktorú použije viac ľudí, je dôležitejšia ako značky použité menším počtom ľudí. V prípade, že sa dve URL líšia len v značkách, ktorými je jedna z nich označená malý počet krát, hodnota podobnosti ostáva vysoká. Podobnosť vyjadruje pomer spoločných značiek k celkovému počtu použítí značiek pre označenie daných dvoch URL a nadobúda hodnotu od 0 do 1.



## Relevancia na základe záujmov používateľa

Ďalším aspektom, ktorý je potrebné brať do úvahy, sú záujmy používateľa. Záujmy používateľa môžeme definovať množinou značiek, ktoré používa pri ukladaní svojich záložiek. Potom stránky, ktoré sú označené značkami, ktoré používateľ používa často, sú pravdepodobne preňho zaujímavejšie, ako stránky označené inými značkami. Preto jednu z metrik označíme zaujímavosť stránky, na ktorú vedie ohodnocovaná hrana.

Relevancia(hrana(url1, url2)) ~ zaujímavosť(url2)

Definujme metriku pre odvodenie zaujímavosti stránky na základe značiek:

```
Zaujímavosť(url) : <0..1>  
  
Nech p = množina všetkých používateľových značiek  
t = značky(url)  
S = celkový počet (t)  
H = 0  
  
Pre všetky spoločné značky sz = prienik(p, t)  
    H = H + Počet výskytov(sz, t)  
  
Zaujímavosť = H / S
```

Takto definovaná metrika vyjadruje pomer značiek. Ako najzaujímavejšiu ohodnotí URL, ktorá je označená len značkami, ktoré používateľ použil vo svojich záložkách. Pritom berie do úvahy počet použití značiek a nadobúda hodnotu od 0 do 1.

## Relevancia na základe sociálnych stôp

Vychádzajúc z princípov sociálnej navigácie (Brusilovsky, 2004) predpokladáme, že pri navigácii na webe sú viac relevantné tie odkazy, ktoré nasledoval vyšší počet používateľov. Preto ako jednu z metrik použijeme počet uskutočnených prechodov danou hranou všetkými používateľmi.

Relevancia(hrana(url1, url2)) ~ počet prechodov(url1, url2)

Definujme metriku sociálnej stopy pre danú hranu:

```
Sociálna stopa(url1, url2) : <0..1>  
  
Nech p = všetky zaznamenané prechody (url1, _x)  
S = celkový počet(p)  
H = 0  
  
Pre všetky prechody sp = (u1, u2) v p  
    Ak url2 = u2  
        H++  
  
Sociálna stopa = H / S
```

Podobne ako v prípade predchádzajúcej metriky, aj tu definujeme hodnotu metriky na základe pomeru počtu uskutočnených prechodov z url1 na url2 k celkovému počtu hrán vychádzajúcich z url1. Pre začiatok vyslovujeme predpoklad, že zohľadnenie celej množiny používateľov bude dávať pre túto metriku dostatočne relevantné výsledky.

V prípade, že overenie ukáže nedostatky takéhoto prístupu, môže byť metrika rozšírená o podobnosť medzi používateľmi.

### **Relevancia na základe návštevnosti**

Podobne, ako v prípade predchádzajúcej metriky, predpokladáme, že pri navigácii na webe sú viac relevantné tie stránky, ktoré videl vyšší počet používateľov. Ako jednu z metrik definujeme počet zobrazení stránky, na ktorú vedie ohodnocovaná hrana.

Relevancia(hrana(url1, url2)) ~ počet zobrazení(url2)

Uvedené metriky pre určovanie relevancie sa spoliehajú na dáta generované používaním metódy. Ide jednak dáta spojené priamo so sledovaním správania používateľov (sociálne stopy a návštevnosť stránok), a tiež metadáta generované používateľmi vo forme značiek asociovaných so záložkami, ktoré používatelia medzi sebou zdieľajú.

Ďalšie možné metriky pre určovanie podobnosti dvoch stránok môžu byť definované na základe porovnania textu dvoch stránok ako dokumentov. Pritom je možné využiť rôzne metódy dolovania z textov, kategorizácie na základe textov, či extrakcie kľúčových slov (Matsuo et al., 2003).

## **5.3 Výhody a obmedzenia navrhutej metódy**

Navrhovaná metóda navigácie na webe má za cieľ adresovať čiastočné riešenie navigačného problému, spočívajúceho v strate orientácie. Z vlastností navrhovanej metódy vyplýva výhoda získania prehľadu o relevancii odchádzajúcich uzlov. Táto je navyše prispôbená aktuálnemu používateľovi, čo ešte viac zvyšuje hodnotu odporúčaní. Aplikáciou metód sociálnej navigácie a sociálneho značkovania sme eliminovali nutnosť použiť pre navigáciu uzavretý dobre opísaný priestor.

Výhodami vyplývajúcimi z povahy grafickej reprezentácie navrhovanej metódy v prípade navigácie na webe je možnosť integrovať metódu do procesu navigácie, reprezentovaného použitím webového prehliadača, neinvazívnym spôsobom. Ako výhodu použitia navrhovanej metódy sme identifikovali aj poskytnutie prehľadného zobrazenia histórie, umožňujúceho návrat na ľubovoľný krok momentálneho sedenia, čo zvyšuje praktickú použiteľnosť navigátora a má za cieľ eliminovať problém straty prehľadu.

Ohraničenia metódy spočívajú najmä v použití existujúcich značiek na webe. Ďalším ohraničením, dôležitým najmä ak chceme metódu experimentálne overiť, je nutnosť mať k dispozícii dáta o používaní webu. Tieto zbiera nástroj počas jeho použitia, ale pri riadených experimentoch môže byť problém inicializovať dáta o používaní rozumnými hodnotami. Aspekt kolaboratívneho filtrovania, zabezpečený použitím metrik opierajúcich sa o dáta o používaní systému, sa môže prejaviť až po istej dobe používania systému netriviálne veľkou množinou používateľov.

Metóda pravdepodobne nebude vhodná pre podporu navigácie z každej stránky. Pre stránky z veľkým počtom rozličných odkazov smerujúcich na podstránky tej istej stránky (napríklad rôzne stránky s novinkami alebo diskusné fóra), môže byť množstvo

odkazov (aj tých dostatočne relevantných) neúmerne vysoké pre praktické použitie podpory navigácie.

Pri realizácii metódy treba uvažovať aj technologické obmedzenia súčasného webu, dané možnosťami získať zoznam stránok, na ktoré sa daná stránka odkazuje (jednoduché v HTML, zložitejšie v prípade stránok využívajúcich napríklad technológiu Flash, ešte zložitejšie v prípade, že stránka využíva technológie ako napríklad AJAX).

## 6 Realizácia adaptívnej podpory navigácie

---

Za účelom overenia navrhnutej metódy sme navrhli a vyvinuli softvérový systém pre adaptívnu podporu navigácie. Formulovali sme požiadavky na funkcionality, navrhli sme architektúru a jednotlivé komponenty systému a systém sme pre účely overenia návrhu implementovali.

### 6.1 Požiadavky

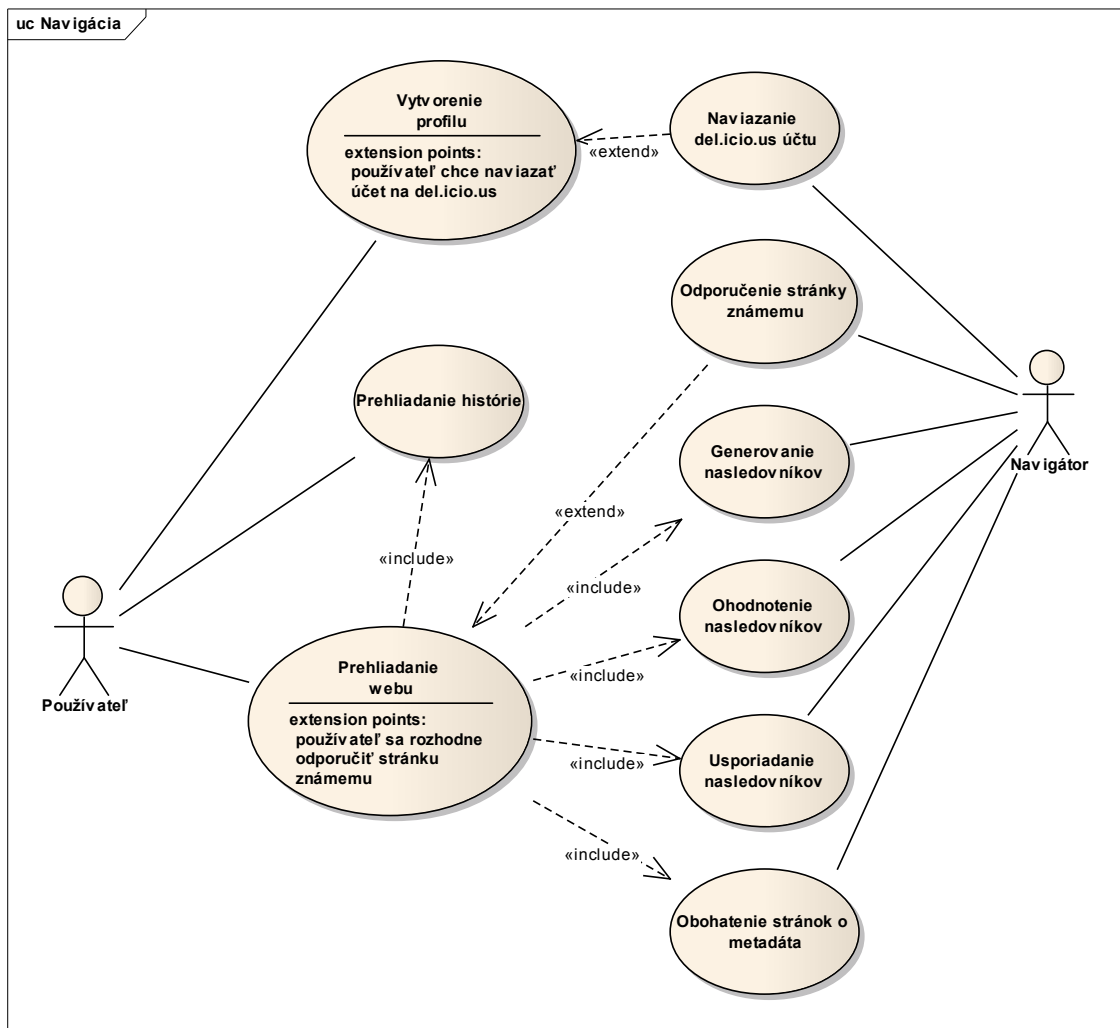
#### Scenár použitia

Používateľ zapína webový prehliadač. Naviguje na stránku, z ktorej sa mu otvára separátne okno, reprezentujúce klientskú aplikáciu prehliadača. Zobrazí sa mu naposledy navštívená stránka a história posledného sedenia. Používateľ vidí graf alebo usporiadaný zoznam, v ktorom je zobrazených niekoľko posledných stránok, ktoré navštívil pri poslednom sedení. Stránky sú anotované zobrazením počtu návštev daným používateľom aj ostatnými používateľmi systému, posledný čas návštevy a značky, ktorými je daná stránka označená (v našom prípade na del.icio.us). Používateľ vidí, že jeho kolega mu odporučil na prezretie zaujímavú stránku, ktorú mu systém farebne zvýraznil.

Používateľovi sa počas navigácie zobrazuje v ľavej časti prehliadača vždy okolie momentálnej stránky (vybrané odkazy, na ktoré sa daná stránka odvoláva a stránky, ktoré sa odvolávajú na danú stránku) a príbuzné stránky, usporiadané podľa relevancie. Zároveň je zobrazenie obohatené o vizualizáciu bezprostrednej histórie prehliadania, a používateľ má možnosť nelineárnym spôsobom skokovo prehliadať históriu navštívených stránok, pri zachovaní prehľadu o predchádzajúcej postupnosti krokov navigácie.

#### Prípady použitia

Základné prípady použitia systému sú znázornené v diagrame na obrázku 8.



Obrázok 8: Prípady použitia.

Opis najdôležitejších prípadov použitia:

**Vytvorenie profilu** – používateľ si vytvára profil v systéme. Má možnosť naviazať ho na svoj účet služby del.icio.us. V takomto prípade zadáva prihlasovacie meno a heslo do služby del.icio.us.

**Naviazanie del.icio.us účtu** – v prípade, že má používateľ účet na službe del.icio.us, môže so svojim profilom zviazať

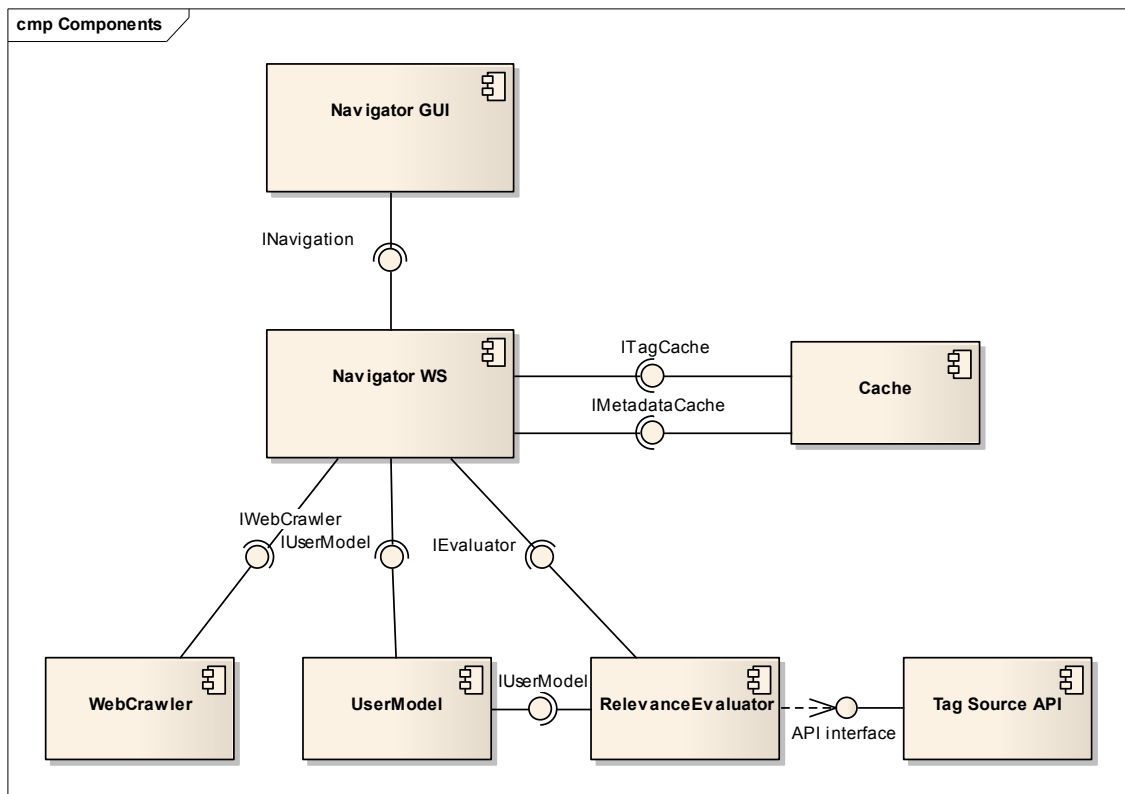
**Prehliadanie webu** – používateľ zadá v rozhraní navigátora začiatočnú stránku navigácie, systém generuje okolie stránky a odporúčania na základe zaznamenaného správania ostatných používateľov. Usporiadáva a filtruje možné nasledovníky momentálnej stránky na základe vypočítanej relevancie. Medzi nasledovníky zahŕňa odchádzajúce odkazy, prichádzajúce odkazy a odporúčania získané dopytom related:url zo služby Google Search.

**Odporúčanie stránky známemu** – používateľ môže vzhľadom na kontext zobrazenej stránky odporučiť stránky pre pozretie známemu. Tieto sú mu následne systémom odporúčané v procese prehliadania webu.

**Prehliadanie histórie** – systém zobrazuje momentálnu históriu sedenia formou grafu reprezentujúceho nasledované prepojenia medzi stránkami. Používateľ sa môže kedykoľvek vrátiť ku ktorémukoľvek uzlu v bezprostrednej histórii. Po ukončení navigácie sa ukladá posledný stav navigácie a pri opätovnom spustení používateľ začína navigáciu tam, kde naposledy skončil.

## 6.2 Návrh systému pre podporu navigácie

Systém je navrhnutý s využitím technológie webových služieb. Diagram znázorňujúci architektúru systému sa nachádza na obrázku 9. Jadrom je webová služba navigátora, ktorá poskytuje základ funkcionality navigácie a funguje ako fasáda pre ostatné komponenty systému. Je prepojená s komponentmi *UserModel*, *WebCrawler* a *RelevanceEvaluator*. Tiež obsahuje prepojenie na *Cache*, kde pre zefektívnenie obsluhy požiadaviek a zníženie sieťovej záťaže ukladá dáta pre ďalšie použitie.



Obrázok 9: Architektúra systému.

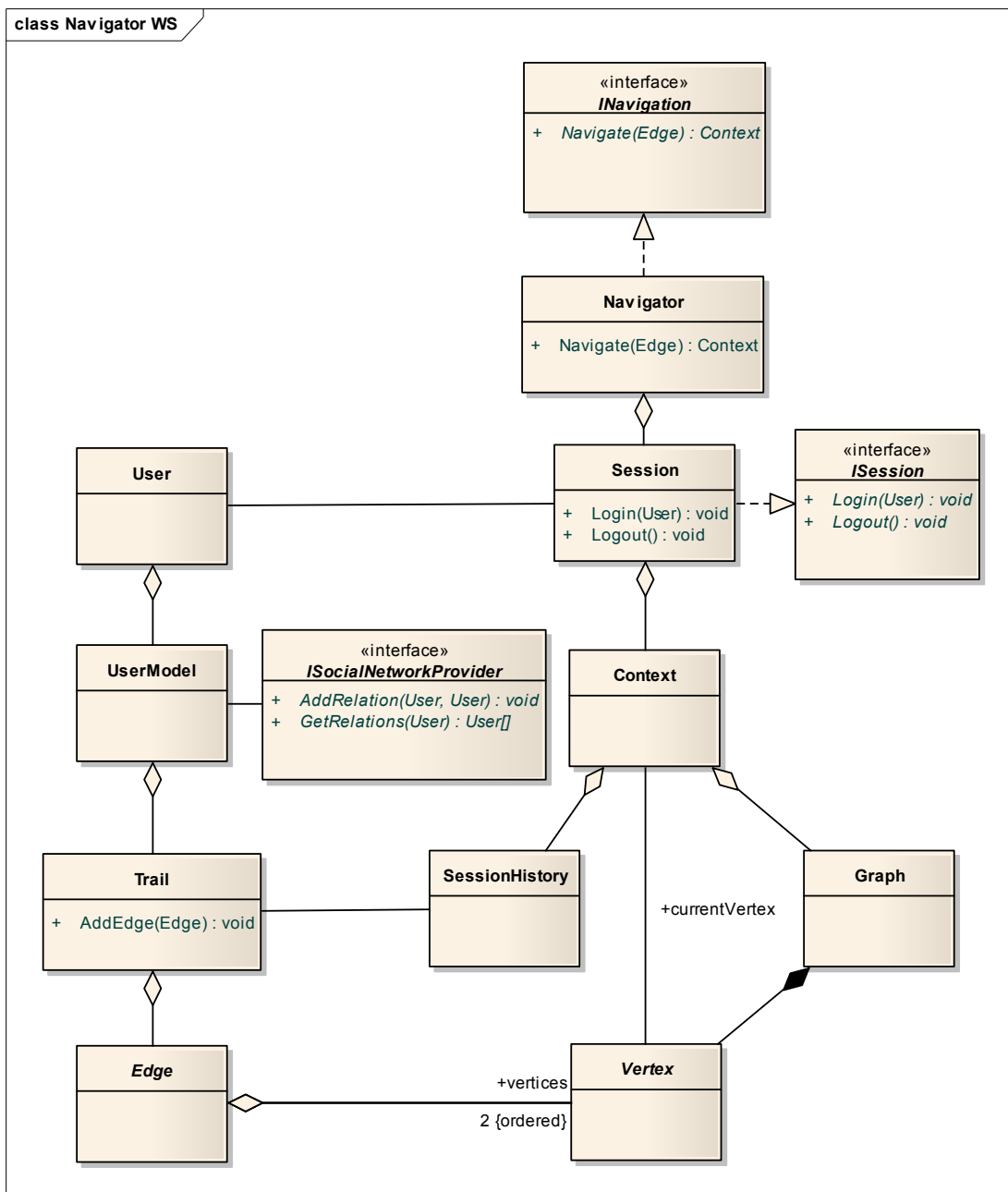
Používateľská aplikácia je navrhnutá vo forme tenkého klienta a poskytuje najmä grafické používateľské rozhranie. Tiež zabezpečuje zachytenie klientských udalostí prehliadača a ich zasielanie službe navigátora pre ďalšie spracovanie.

Služba *Navigation* je prístupná ako stavová webová služba. Rozhranie služby predstavuje rozhranie *INavigation*, ponúkajúce základnú metódu *Navigate(Edge)*. Vracia nový kontext (inštanciu triedy *Context*), zahŕňajúci momentálne zobrazené okolie a históriu prehliadania.

Komponent *UserModel* slúži pre správu modelov používateľa a ukladanie trás prehliadania. *RelevanceEvaluator* využíva dáta uložené v komponente *UserModel* pre vyhodnocovanie relevancie a odporúčanie navigačných krokov v procese navigácie a rozhranie *ITagSource* pre získanie značiek pre jednotlivé URL.

*WebCrawler* zabezpečuje získavanie dát o štruktúre prepojení webových stránok. Získava zdrojový HTML text stránok a z nich extrahuje odkazy na ostatné stránky. Tiež obsahuje rozhranie pre prístup k službe Google Search, ktorá poskytuje informáciu o spätnom odkazovaní. Disponuje vyrovnávacou pamäťou pre zníženie sieťovej záťaže.

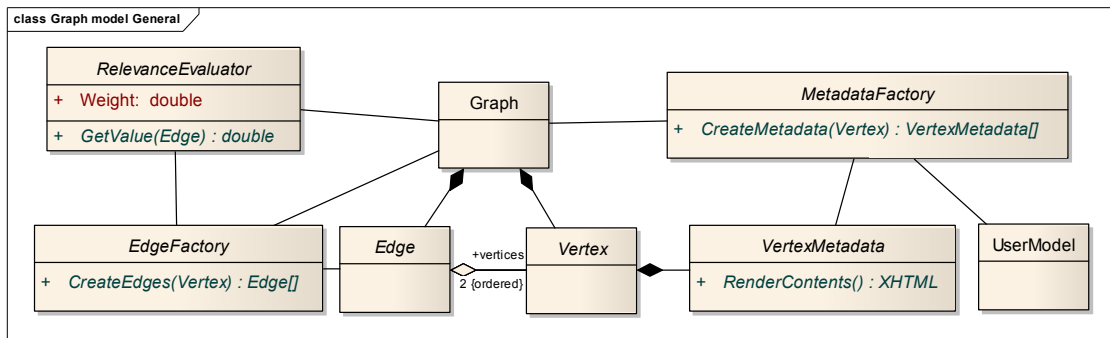
Diagram tried na obrázku 10 znázorňuje triedy, ktoré zabezpečujú infraštruktúru navigačnej služby a jadro metódy. Prihlásenie a správu používateľov zabezpečuje trieda *Session*. Je asociovaná s prihláseným používateľom. Používateľ je modelovaný triedou *UserModel*, ktorá zachytáva navigačné stopy jednotlivých používateľov. Obsahuje tiež odkaz na rozhranie *ISocialNetworkProvider* (prístupom obrátenej kontroly (*Inversion of Control*)), ktorého implementáciou sa pridá možnosť modelovať v modeli používateľa sociálne vzťahy medzi používateľmi. Navigačná stopa (*Trail*) predstavuje postupnosť hrán, obsahujúcich vždy dva vrcholy patriace do grafu (trieda *Graph*).



Obrázok 10: Služba Navigation.

Na obrázku 11 je znázornený všeobecný model grafu, ktorého realizáciou sa systém prispôsobuje pre navigáciu v rôznych doménach, a to na dátovej úrovni špecializáciou tried *Edge* a *Vertex* pre konkrétnu doménu, a na funkcionálnej úrovni (predstavujúcej generovanie lokálneho okolia uzlov v grafe) špecializáciou tried zodpovedných za vytváranie hrán (*EdgeFactory*) prípadne generovanie metadát uzlov (*MetadataFactory*).



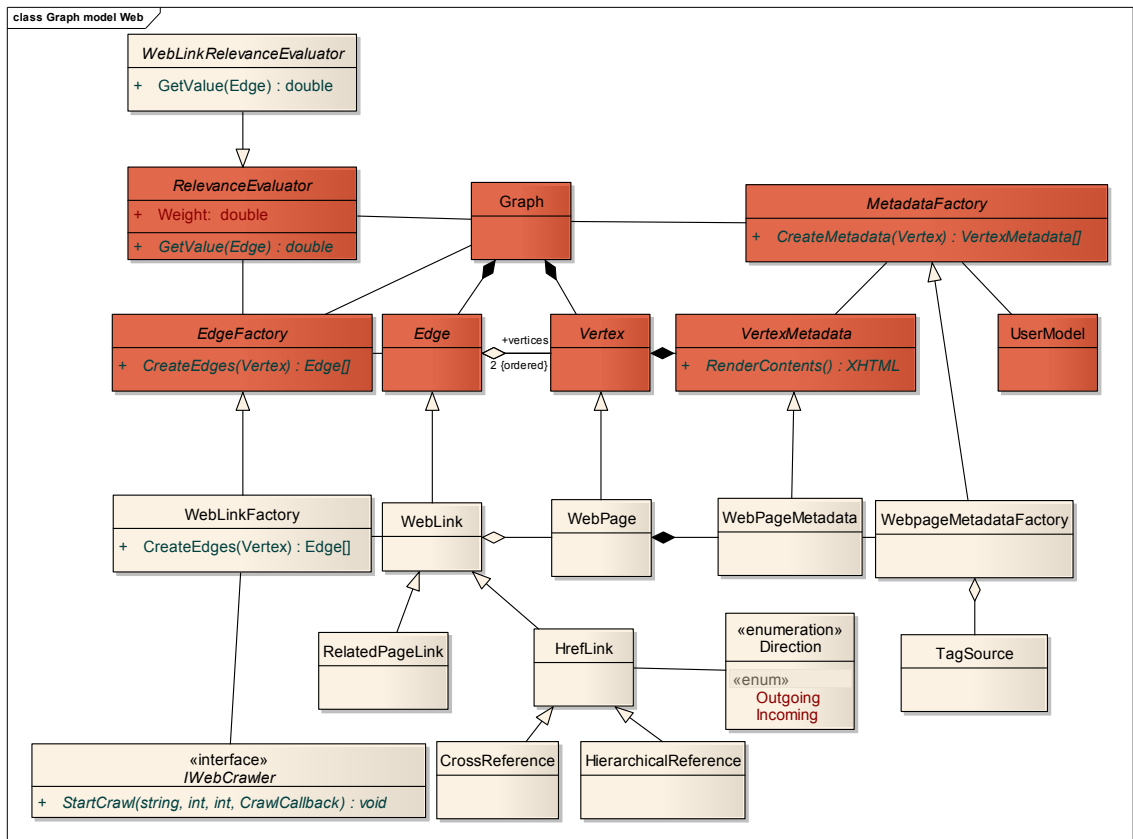


Obrázok 11: Všeobecný model prehliadaného grafu.

Konkrétna realizácia všeobecného modelu prehliadaného grafu pre doménu navigácie na webe je znázornená na obrázku 12. Triedy všeobecného modelu sú v diagrame zvýraznené, zvyšok sú špecializované triedy pre doménu navigácie na webe.

Hrany sú modelované triedou *WebLink*, respektíve jej konkrétnymi podtriedami *CrossReference*, *HierarchicalReference* (predstavujú a *RelatedPageLink*. Vrcholy v grafe predstavujú inštanície triedy *WebPage*, s ktorou sú asociované metadáta (*WebPageMetadata*).

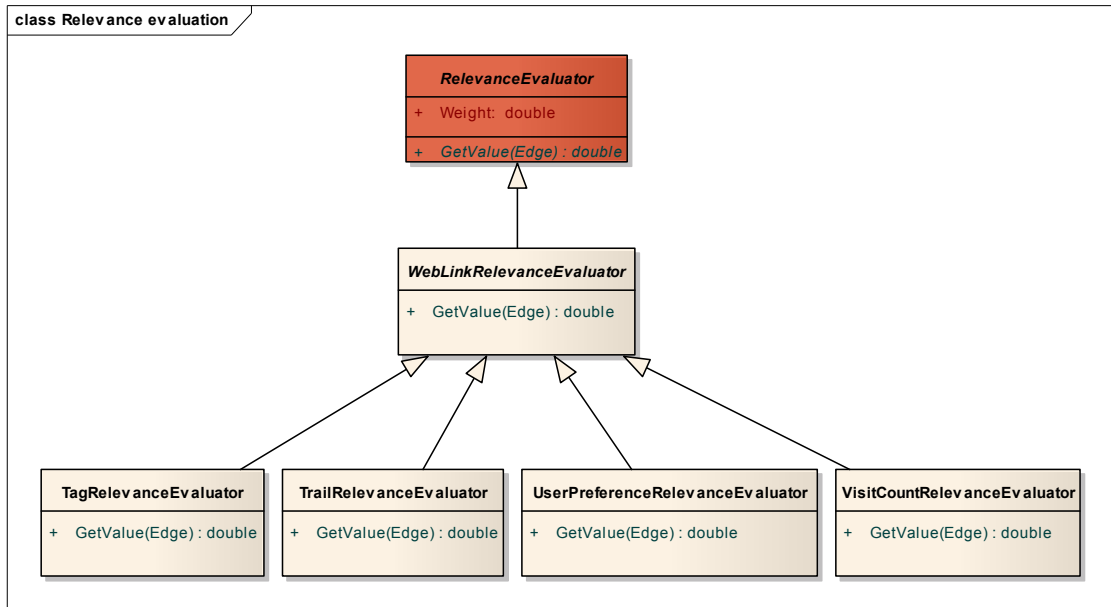
Za vytváranie sú zodpovedné triedy *WebLinkFactory* (generuje okolie momentálneho uzla), ktorá využíva realizáciu *crawlera* (implementáciu *IWebCrawler*) a *WebPageMetadataFactory*, ktorá využíva pre získavanie značiek asociovaných s webovými zdrojmi triedu *TagSource*. Štruktúra tried *WebLinkFactory* a *WebPageMetadataFactory* je realizáciou návrhového vzoru abstraktná továreň (*Abstract Factory*) (Gamma et al., 1995).



Obrázok 12: Model grafu v prípade domény webu. Všeobecná časť modelu je zvýraznená.

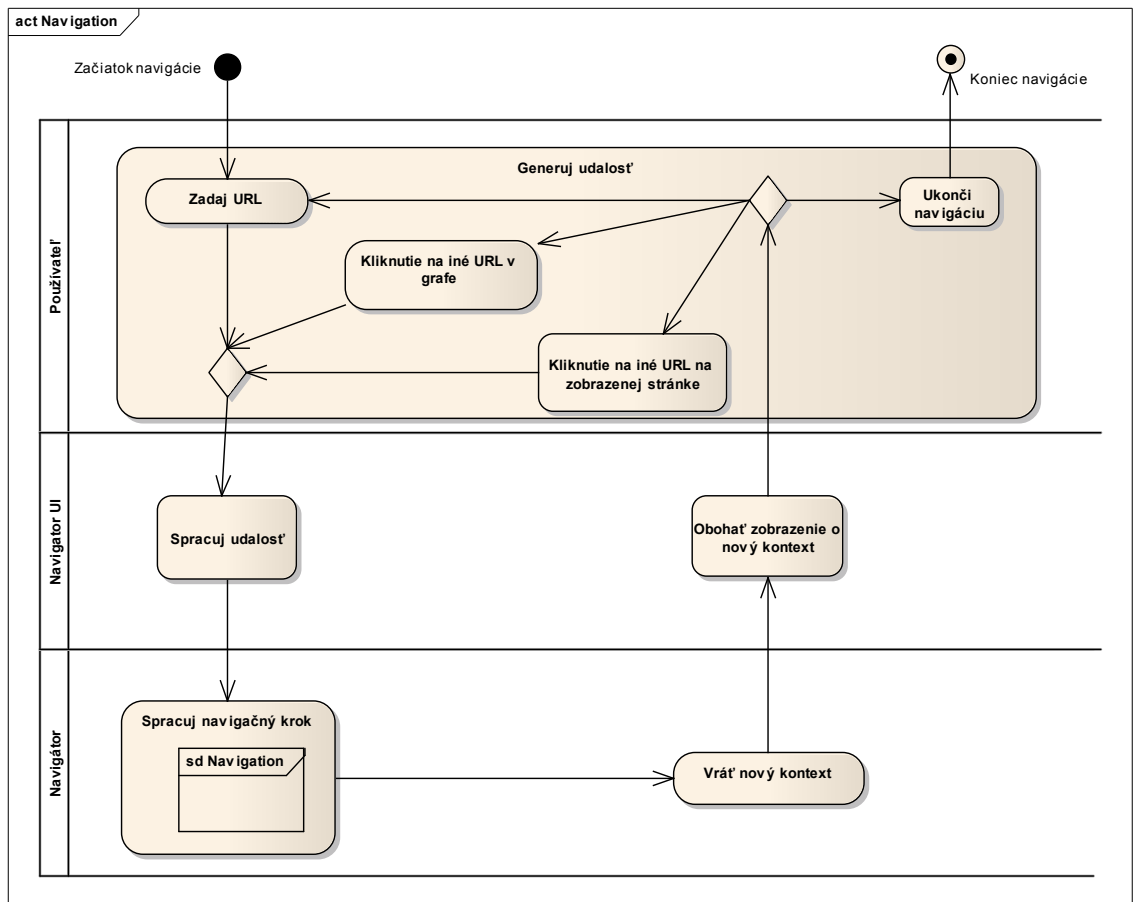
Štruktúra tried zodpovedných za vyhodnocovanie relevancie hrán v prípade použitia pre navigáciu na webe je znázornená na obrázku 13.

Triedy *TagRelevanceEvaluator*, *TrailRelevanceEvaluator*, *VisitCountRelevanceEvaluator* a *UserPreferenceRelevanceEvaluator* vyhodnocujú relevanciu daného navigačného kroku implementáciou metrick opísaných v kapitole 5.2.



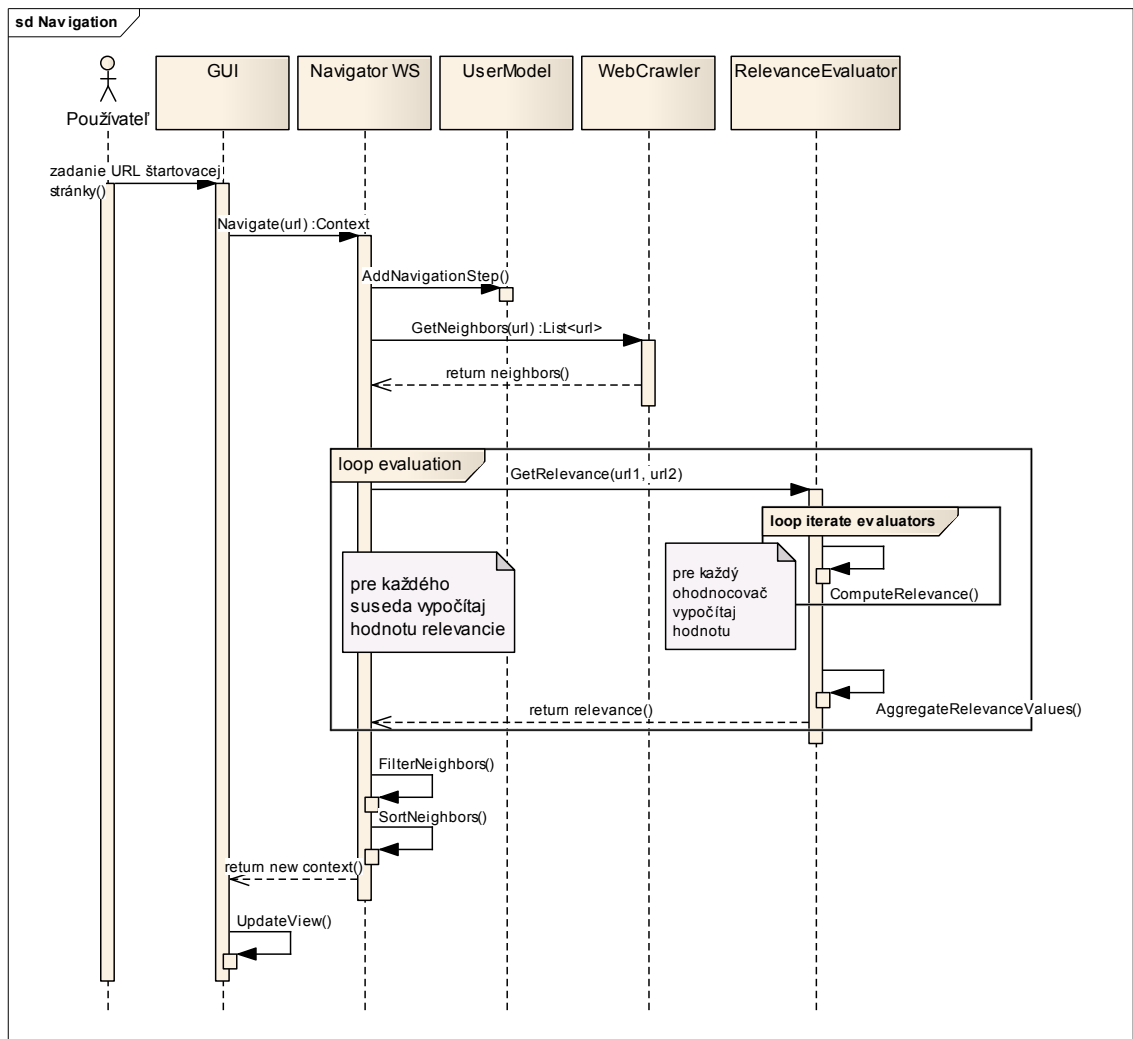
Obrázok 13: Vyhodnocovače relevancie v prípade domény webu.

Na obrázku 14 je znázornený proces navigácie formou diagramu aktivít. Navigácia začína zadaním adresy (URL) počiatočnej stránky. Grafické rozhranie spracuje udalosť a odošle požiadavku na webovú službu navigátora. Tá spracuje navigačný krok (pozri obrázok 15 a opis diagramu ďalej v texte) a vráti nový kontext pre zobrazenie. Grafické rozhranie navigátora spracuje vrátený kontext a obohatí zobrazenie o nové vrcholy a hrany spolu s prípadnými asociovanými metadátami. Nasleduje akcia používateľa, ktorý sa môže rozhodnúť či vykoná navigačný krok kliknutím do zobrazeného grafu, alebo priamo nasledovaním odkazu na zobrazenej stránke. Taktiež môže zadať novú adresu do riadku s adresou a začať navigáciu znova. Po prijatí akcie používateľa grafickým rozhraním sa proces opakuje.



Obrázok 14: Diagram aktivít zachytávajúci schému navigácie.

Podrobný proces spracovania jedného kroku navigácie je znázornený v diagrame na obrázku 15.



Obrázok 15: Sekvenčný diagram schémy navigácie.

Po zachytení akcie používateľa grafickým rozhraním navigátora je zavolaná metóda *Navigate* webovej služby navigátora. Tento zaznamená daný navigačný krok v modeli používateľa (volaním metódy *AddNavigationStep*) a s pomocou triedy *WebCrawler* získa okolie daného URL. Pre získané odkazy vytvorí dvojice predstavujúce hrany prepojené medzi webovými stránkami (*WebLink*) a pre jednotlivé vyhodnocovače (podtriedy *RelevanceEvaluator*) získa agregovanú hodnotu relevancie jednotlivých hrán. Usporiada a vyfiltruje vygenerovaných možných nasledovníkov. Grafickému rozhraniu odovzdá inštanciu nového aktuálneho kontextu a ten obohatí zobrazenie o nové vrcholy a hrany.

### 6.3 Implementácia v prostredí webu

S cieľom overenia navrhutej metódy sme implementovali nasledovné časti systému:

- *AWNS.ServiceLibrary*, knižnica obsahujúca hlavné komponenty jadra systému, reprezentáciu priestoru grafom, implementáciu všeobecnej metódy navigácie v grafoch, triedy umožňujúce spúšťanie a beh . Z dôvodu jednoduchšej

implementácie a výhod použitej knižnice pre implementáciu webových služieb je knižnica referencovaná na strane klienta aj servera

- *AWNS.ServiceLibrary.Host.Web*, konzolová aplikácia, v ktorej môže bežať webová služba navigátora
- *AWNS.Web*, klientská aplikácia, obsahujúca grafické rozhranie a implementáciu asynchrónneho inkrementálneho mechanizmu konštrukcie a ohodnocovania grafu navigácie
- Podporné nástroje pre analýzu priestoru webu, ktoré sme použili pri analýzach, ktorých výsledky uvádzame v kapitole 7.2.

Webová služba navigátora môže bežať buď v konzolovej aplikácii *AWNS.ServiceLibrary.Host.Web*, alebo priamo v klientskej aplikácii. Takáto možnosť bola realizovaná kvôli obmedzeniam služby *del.icio.us*, ktorá neumožňuje spracovanie požiadavky z jednej IP adresy častejšie ako jeden krát za sekundu a v prípade, že by bolo k službe navigátora pripojených viacero klientov naraz, čas pre spracovanie požiadaviek by priamoúmerne narastal z ich počtom, a veľmi skoro by sa predĺžil na neúmernú hranicu. Preto môže byť webová služba umiestnená priamo v procese klientskej aplikácie, a dáta, ktoré služba používa (modely používateľov, štatistiky o vykonaných navigačných krokoch a návštevnosti jednotlivých stránok), sú centralizované v relačnej databáze, ktorá zároveň slúži ako globálna vyrovnávacia pamäť.

### 6.3.1 Možnosti rozšírenia

Realizovaná implementácia metódy podpory navigácie je univerzálna z pohľadu aplikačnej domény, a je možné ju aplikovať pre navigáciu v ľubovoľných priestoroch, pre ktoré existuje reprezentácia vo forme grafu, a kde môže byť špecifikovaná funkcia ohodnocovania hrán.

Pre adaptáciu implementovaného riešenia je potrebné vykonať nasledovné kroky:

- Špecifikovať funkciu generovania susedných hrán
- Definovať funkciu ohodnocovania hrán
- V prípade potreby je možné implementovať triedy reprezentujúce hrany a vrcholy grafu a metadát k hranám (v opačnom prípade je možné použiť všeobecné triedy, ktoré sú súčasťou jadra systému)
- V prípade potreby môže byť deklarované rozhranie obsahujúce metódy špecifické pre navigáciu v danej doméne
- Deklarovať triedu odvodenú z triedy *NavigationBase<TEdge, TVertex>* a v prípade potreby implementovať deklarované špecifické rozhranie

Službu pozostávajúcu z výstupov krokov popísaných vyššie je možné inštancovať a hosťovať v rámci infraštruktúry WCF.

Pre demonštračné účely sme implementovali službu navigátora pre doménu publikácií. Grafová reprezentácia obsahovala vrcholy predstavované publikáciami, a hrany, ktoré reprezentovali spoluautorstvo, citácie a ostatné publikácie autora danej publikácie. Ako

zdroj dát sme využili ontológiu publikácií v projekte Mapekus<sup>7</sup>. Pre ohodnocovanie hrán boli použité funkcie opierajúce sa o počet prezretí daného vrcholu a sociálne stopy, tak ako boli popísané vyššie.

Zdrojové kódy k implementácii v doméne publikácií sa nachádzajú na priloženom elektronickom médiu.

### 6.3.2 Použité technológie

Webová služba využíva rámec Windows Communication Foundation. Grafické rozhranie je implementované s využitím Windows Presentation Foundation. Pre prístup k relačnej databáze využívame dopytovanie integrované v jazyku C# 3.0 a objektovo relačný mapper DLINQ. Celé riešenie je implementované v jazyku C# s využitím rámca .NET 3.5 vo vývojovom prostredí Microsoft Visual Studio 2008.

Pre zobrazenie grafu bola použitá knižnica *J832.Wpf.BagOTricksLib*<sup>8</sup>, implementujúca dynamický pružinový model rozmiestnenia uzlov v grafe. Pre prístup k službe del.icio.us využívame knižnicu *Del.icio.us.NET*<sup>9</sup>, využívajúcu API služby del.icio.us, ktorú sme rozšírili o dodatočnú funkcionálnosť a pre mierne upravili dosiahnutie vyššej spoľahlivosti.

---

<sup>7</sup> <http://mapekus.fiit.stuba.sk>

<sup>8</sup> Prístupná na <http://j832.com/bagotricks/>

<sup>9</sup> Prístupná na <http://www.codeplex.com/NDelicious>

## 7 Experimentálne overenie metódy

---

Navrhnutú metódu sme sa rozhodli overiť vykonaním niekoľkých experimentov. Analyzovali sme priestor webu z hľadiska štruktúry prepojení stránok a tiež z hľadiska špecifických vlastností, ktoré predpokladá navrhnutá metóda adaptívnej podpory navigácie. Overili sme vlastností navrhnutej a implementovanej funkcie podobnosti na základe spoločných značiek. Navrhli a vykonali sme experiment formou používateľskej štúdie, ktorého cieľom bolo overiť vnímanie vlastností systému používateľmi a mieru jeho využiteľnosti pre použitie pri prieskumnom prehľadávaní.

### 7.1 Overenie vlastností funkcie podobnosti stránok na základe spoločných značiek

Prvý experiment, ktorý sme uskutočnili, mal za cieľ overiť predpoklad, že podobné stránky sú na stránkach zaoberajúcich sa sociálnym značkováním značkované podobne, a teda že metrika podobnosti stránok na základe spoločných značiek bude dávať relevantné výsledky.

Boli vybrané nasledovné skupiny stránok:

#### **Skupina 1 Stránky o aktuálnych udalostiach**

- <http://news.google.com/>
- <http://news.bbc.co.uk/>

#### **Skupina 2 Stránky s hardvérovými novinkami**

- <http://www.tomshardware.com/>
- <http://www.anandtech.com/>
- <http://www.xbitlabs.com/>

#### **Skupina 3 Stránky leteckých spoločností**

- <http://www.ryanair.com/>
- <http://www.skyeurope.com/>

#### **Skupina 4 Slovenské stránky s technologickými novinkami**

- <http://www.dsl.sk/>
- <http://www.zive.sk/>

#### **Skupina 5 Stránky zoberajúce sa fotografovaním**

- <http://www.dpreview.com/>



- <http://photo.net/>

### Skupina 6 Stránky s populárno-vedeckou tematikou

- <http://www.popsci.com/>
- <http://www.physorg.com/>
- <http://www.dailytech.com/>
- <http://www.tgdaily.com/>

### Skupina 7 Stránky veľkých softvérových firiem

- <http://www.apple.com/>
- <http://www.microsoft.com/>

Hypotéza predpokladá pre stránky v identifikovaných skupinách vzájomne vysokú mieru podobnosti vypočítanú na základe metriky podobnosti značiek na del.icio.us. Taktiež predpokladáme, že stránky v skupinách, ktoré sú si navzájom podobné, budú mať vyššiu mieru podobnosti. Tabuľka 1 zachytáva predpokladané podobnosti medzi dvojicami skupín stránok. Čísla v záhlaví predstavujú čísla vybraných skupín stránok.

Tabuľka 1: Identifikovaná podobnosť medzi dvojicami skupín stránok.

	1	2	3	4	5	6	7
1	vysoká						
2	stredná	vysoká					
3	nulová	nulová	vysoká				
4	veľmi nízka	stredná	nulová	vysoká			
5	nulová	nízka	nulová	veľmi nízka	vysoká		
6	stredná	vysoká	nulová	stredná	nulová	vysoká	
7	nulová	nízka	nulová	veľmi nízka	nulová	nulová	vysoká

S využitím implementovaného nástroja bola pre každú dvojicu stránok vypočítaná hodnota funkcie podobnosti. Výsledky experimentu sú znázornené v tabuľke 2. Jednotlivé stránky sú v tabuľke uvedené v poradí, v akom boli uvedené v rozdelení vyššie (záhlavia tabuľky zobrazujú z priestorových dôvodov vždy len niekoľko prvých písmen každého názvu). Odtieň pozadia určuje podobnosť danej dvojice stránok (čím tmavšie pozadie, tým podobnejšie). Hodnota bunky predstavuje hodnotu funkcie podobnosti pre danú dvojicu stránok.

Tabuľka 2: Podobnosť dvojíc stránok v jednotlivých skupinách.

	news.	news.	tomsh	ananc	xbitla	ryana	skyeu	dsl.sk	zive.sl	dprevi	photo	pops	physo	dailyt	tgdaily	apple	micro
news.	<b>1,00</b>																
news.	<b>0,70</b>	<b>1,00</b>															
tomsh	0,35	0,40	<b>1,00</b>														
ananc	0,36	0,41	<b>0,99</b>	<b>1,00</b>													
xbitla	<b>0,55</b>	<b>0,53</b>	<b>0,98</b>	<b>0,98</b>	<b>1,00</b>												
ryana	0,03	0,02	0,02	0,00	0,02	<b>1,00</b>											
skyeu	0,03	0,02	0,01	0,00	0,01	<b>0,95</b>	<b>1,00</b>										
dsl.sk	0,48	0,48	0,19	0,23	0,19	0,00	0,00	<b>1,00</b>									
zive.sl	0,48	0,48	0,30	0,30	0,29	0,00	0,00	<b>0,75</b>	<b>1,00</b>								
dprevi	0,17	0,24	0,27	0,27	0,25	0,00	0,00	0,02	0,02	<b>1,00</b>							
photo	0,00	0,00	0,13	0,13	0,11	0,00	0,00	0,00	0,00	<b>0,76</b>	<b>1,00</b>						
pops	0,48	0,47	0,30	0,35	0,30	0,00	0,00	0,26	0,26	0,02	0,00	<b>1,00</b>					
physo	0,47	0,48	0,34	0,37	0,41	0,00	0,00	0,25	0,20	0,05	0,00	<b>0,68</b>	<b>1,00</b>				
dailyt	<b>0,55</b>	<b>0,53</b>	<b>0,78</b>	<b>0,79</b>	<b>0,81</b>	0,00	0,00	0,39	0,36	0,03	0,00	<b>0,59</b>	0,50	<b>1,00</b>			
tgdaily	<b>0,56</b>	<b>0,53</b>	<b>0,93</b>	<b>0,94</b>	<b>0,93</b>	0,00	0,00	0,35	0,38	0,17	0,08	0,48	0,45	<b>0,97</b>	<b>1,00</b>		
apple	0,02	0,02	0,38	0,35	0,29	0,00	0,00	0,00	0,18	0,00	0,00	0,06	0,08	0,30	0,30	<b>1,00</b>	
micro	0,05	0,03	0,29	0,27	0,23	0,04	0,03	0,00	0,12	0,00	0,00	0,06	0,12	0,35	0,36	0,30	<b>1,00</b>

Ako je zrejme z tabuľky 2, v súlade s predpokladom je medzi podobnými stránkami (v našom prípade stránkami v identifikovaných skupinách stránok a podobných skupinách stránok) hodnota metriky podobnosti vyššia ako medzi nepodobnými stránkami. Skupiny stránok, medzi ktorými sa ukázala vysoká hodnota vzájomnej podobnosti, sme v tabuľke zvýraznili orámovaním.

Pre položky, ktoré sa nachádzajú v tabuľke na diagonále, je zrejme vysoká hodnota vzájomnej podobnosti. Z tabuľky je tiež zrejme relatívne vysoká podobnosť medzi stránkami skupín, ktoré sme identifikovali ako vzájomne podobné (1, 2 a 6). Predpoklad sa nenaplnil pre dvojicu skupín 2 a 4, kde bola odmeraná len relatívne nízka vzájomná podobnosť (0,19-0,30). Toto môže byť spôsobené jednak rozdielom v jazyku skupín stránok (skupina 4 sú slovenské stránky) a tiež nízkym pokrytím slovenských stránok značkami vo všeobecnosti.

V tabuľke si je možné všimnúť izolovanú pozíciu stránok zo skupiny 3 (letecké spoločnosti), ktoré majú veľmi vysokú hodnotu vzájomnej podobnosti (0,95) a v podstate nulovú hodnotu podobnosti s ostatnými stránkami. Najnižšie hodnoty funkcie podobnosti vyskytujúce sa v tabuľke sú zapríčinené výskytom značiek generovaných rôznymi nástrojmi pre správu značiek (napríklad značka „imported“, ktorú pridáva nástroj pre export záložiek z webového prehliadača). Toto bude potrebné zohľadniť pri implementácii a identifikované irelevantné značky ignorovať. Je možné, že vzhľadom na frekvenciu výskytu takýchto značiek a vlastnosti funkcií ohodnocujúcich relevanciu dvojíc stránok na základe značiek ich systém odfiltruje bez vonkajšieho explicitného zásahu.

Stránky zo skupiny 5 (fotografické stránky) majú taktiež vzájomne vysokú mieru podobnosti (0,76) a s ostatnými nízku (hardvérové novinky) až nulovú podobnosť.

Na základe výsledkov vykonaného experimentu môžeme považovať vlastnosti navrhutej metriky pre množinu stránok, pre ktorú sme experiment vykonali, za vyhovujúce, a vyslovujeme predpoklad, že má takéto vlastnosti má aj vo všeobecnosti,

teda že dáva pre podobné stránky vyššie hodnoty ako pre menej podobné dvojice stránok.

## 7.2 Analýza priestoru webu

Po overení vlastností funkcie relevancie sme sa rozhodli predpokladať, na ktorých je založené fungovanie metódy adaptívnej navigácie, overiť analýzou priestoru webu. Pre vybrané časti webu vychádzajúc z niekoľkých webových stránok sme vytvorili graf prepojení webových stránok hypertextovými odkazmi. Výsledky analýzy uvádzame v tabuľkách.

Pre graf prepojení webových stránok sme zisťovali nasledovné charakteristiky:

- Priemerný počet odchádzajúcich odkazov
- Miera pokrytia značkami na del.icio.us
  - priemerný počet rôznych značiek na URL (vzhľadom na obmedzenia služby del.icio.us je tento údaj skreslený<sup>10</sup>)
  - podiel označovaných stránok

Pri analýze sa ukázalo, že pre isté stránky môže byť navrhovaná metóda prakticky použiteľná len v obmedzenej miere, pretože obsahujú prílišné množstvo krížových odkazov na stránky (iné podstránky daného webového sídla), pre ktoré existuje veľmi malé pokrytie značkami na del.icio.us. V týchto prípadoch bude užitočnejšie spoliehať sa viac na záznamy trás prehliadania ostatných (podobných) používateľov. K stránkam tohto typu patria rôzne fóra alebo stránky publikujúce pravidelné novinky, ako napríklad stránka dpreview.com, kde pri prehľadávaní do šírky crawler po prehľadaní 10.000 odkazov ostával stále na doméne dpreview.com, pričom jednotlivé podstránky na danej doméne boli značkované len veľmi málo. Miera označovania týchto stránok bola menej ako 15%.

Analýzu sme vykonávali vždy po častiach pre malé časti webu (maximálny počet uzlov 200-400). Toto bolo determinované aj technologickými obmedzeniami API služby del.icio.us, ktorá obmedzuje použitie verejných rozhraní a pri ich nadmernom využití odmieta z danej IP adresy ďalšie dopyty.

Experimenty sme vykonali so zohľadnením prepojení len medzi koreňovými stránkami. Výsledky všetkých uskutočnených experimentov v elektronickej forme spolu so zoznamami URL všetkých zahrnutých stránok sa nachádzajú v prílohe tohto dokumentu na elektronickej stránke. Na tomto mieste uvádzame prehľad vybraných experimentov.

<b>Počiatočná stránka</b>	<b>http://web.mit.edu</b>
Počet vrcholov	421
Počet hrán	400

---

<sup>10</sup> Služba del.icio.us poskytuje pre každé URL maximálne 11 najčastejšie použitých značiek.

Priemerný počet odchádzajúcich odkazov	12,1
Pokrytie značkami – vrcholy	58,4%
Pokrytie značkami – hrany	56,5%

<b>Počiatková stránka</b>	<b><a href="http://news.google.com">http://news.google.com</a></b>
Počet vrcholov	301
Počet hrán	300
Priemerný počet odchádzajúcich odkazov	12,5
Pokrytie značkami – vrcholy	61,5%
Pokrytie značkami – hrany	58%

<b>Počiatková stránka</b>	<b><a href="http://www.sav.sk/">http://www.sav.sk/</a></b>
Počet vrcholov	234
Počet hrán	200
Priemerný počet odchádzajúcich odkazov	5,3
Pokrytie značkami – vrcholy	44%
Pokrytie značkami – hrany	31%

Výsledky, zachytávajúce sumu a priemernú hodnotu zisťovaných charakteristík zo všetkých vykonaných experimentov, sa nachádzajú v nasledujúcej tabuľke:

<b>Agregovaná štatistika – počiatkové stránky</b>	<b><a href="http://www.tomshardware.com">http://www.tomshardware.com</a> <a href="http://www.sav.sk/">http://www.sav.sk/</a> <a href="http://www.fiit.stuba.sk">http://www.fiit.stuba.sk</a> <a href="http://www.azet.sk">http://www.azet.sk</a> <a href="http://www.amazon.com">http://www.amazon.com</a> <a href="http://www.acm.org/">http://www.acm.org/</a> <a href="http://news.google.com">http://news.google.com</a></b>
Počet vrcholov	1623
Počet hrán	1538
Priemerný počet odchádzajúcich odkazov	10,4

odkazov	
Pokrytie značkami – vrcholy	49%
Pokrytie značkami – hrany	42%

Agregované hodnoty sledovaných charakteristík ukazujú relatívne dobré pokrytie značkami pre stránky (polovica pokrytých stránok) aj pre dvojice na seba odkazujúcich stránok. Toto je spôsobené najmä faktom, že sme brali do úvahy len prepojenia medzi koreňovými stránkami a ignorovali sme podstránky. Miera pokrytia značkami pre koreňové stránky je vyššia ako pokrytie podstránok.

Počas testovania a ladenia implementovaného nástroja, ako aj pri vykonávaní ostatných experimentov, bežal nástroj v režime, kedy ukladal do vyrovnávacej pamäte značky zobrazených a ohodnocovaných stránok (prepojených hranami so zobrazenými stránkami). Tieto údaje sme použili pre vyhodnotenie reálneho pokrytia stránok značkami na del.icio.us. Výsledky zachytáva tabuľka 3. Pokrytie cca. 1/5 všetkých navštívených a ohodnocovaných URL sa zdá ako relatívne dobrá hodnota, ale môže sa ukázať ako obmedzenie pre ohodnotenie relevancie hrán pri reálnom použití. Ide o externé obmedzenie dané dostupnosťou dát na súčasnom webe.

Tabuľka 3: Pomer označovaných stránok - výsledky z reálneho použitia.

Celkový počet URL	1391
Počet označovaných URL	311
Pomer značkovania	22,36%

### 7.3 Používateľská štúdia

Za účelom overenia vlastností navrhnutej metódy sme implementované riešenie overili vykonaním experimentov formou používateľskej štúdie. Ako základ experimentu sme postavili nasledovné hypotézy:

- Pri použití aplikácie implementujúcej navrhnutú metódu používateľ počas prieskumného prehľadávania (*exploratory search*) objaví nové pre neho relevantné a zaujímavé webové zdroje.
- Vizualizácia kontextu momentálnej stránky je vnímaná ako užitočná, a používateľ ju s úspechom používa pre navigáciu namiesto nasledovania odkazov nachádzajúcich sa priamo v obsahu stránky.
- Používateľ nemá pocit, že stráca kontext, je preňho ľahké vrátiť sa na predtým navštívené stránky a udržiavať si predstavu o priebehu sedenia. Rozhranie považuje za jednoduché.

Experiment sme vykonali na jednom počítači, kde bola nainštalovaná implementovaná aplikácia. Používateľskej štúdie sa zúčastnilo 9 osôb, všetci boli študenti vo veku 20-24 rokov. Účastníci po oboznámení sa s ovládaním aplikácie dostali úlohu prehliadať web podľa ľubovôle s tým že môžu si prezerať čokoľvek ich zaujme počas 15 minút.

Aplikácia bežala v režime, kedy sledovala pomer kliknutí na odkazy priamo v tele zobrazenej stránky a na odporúčané odkazy v ľavej časti. Po uplynutí stanovenej doby používatelia vyplnili dotazník, kde vyjadrili súhlas s tvrdeniami na stupnici 1-9 (1=nesúhlasím, 9=úplne súhlasím). Výsledky boli spriemerované a sú zachytené v tabuľke 4.

Tabuľka 4: Výsledky používateľskej štúdie.

Vizualizácia okolia stránky mi uľahčila prehliadanie webu	7
Objavil som zaujímavé stránky, ktoré som predtým nepoznal	7,888889
Mal som pri prehliadaní webu pocit že strácam kontext	3,777778
Rozhranie považujem za jednoduché	7,777778
Riešenie sa mi páči	7,888889

Ako je zrejmé z tabuľky, ako hlavnú výhodu je možné označiť objavenie nových zaujímavých stránok. Používatelia skôr nesúhlasili s tvrdením, že strácajú pri prehliadaní webu kontext, čo bolo v súlade s hypotézou.

Zistená miera využitia podpory navigácie (oproti klikaniu priamo na odkazy v zobrazenej stránke) bola 54%. Toto číslo je vyššie ako sme očakávali, a môže byť spôsobené aj tým, že používatelia videli aplikáciu prvý krát a na odkazy zobrazené v grafe klikali zo zvedavosti.

## 7.4 Závery experimentov

Overili sme predpokladané vlastnosti navrhutej funkcie podobnosti stránok. Výsledky experimentu potvrdili predpoklad o vlastnostiach overovanej funkcie ohodnocujúcej podobnosť stránok na základe značiek. Ďalej sme zisťovali charakteristiky priestoru webu z hľadiska štruktúry prepojení webových stránok a diskutovali sme o výsledkoch analýzy.

Výsledky používateľskej štúdie potvrdili hypotézu, a ukazuje sa že testovaná vzorka používateľov riešenie vníma ako užitočné a jednoduché.

Overenie vlastností metódy vychádzajúcich z dát o správaní používateľov bolo obmedzené dostupnosťou takýchto dát. Návrh metrík pre ohodnotenie relevancie strán, spoliehajúci sa na sociálne stopy a návštevnosť stránok, predpokladá dlhodobé používanie systému relevantne veľkou množinou používateľov, ktoré sme v čase experimentu nemohli zabezpečiť. Ďalším obmedzením bola dostupnosť značiek na dnešnom webe, ktorá je dnes pomerne nízka. Možným riešením do budúcnosti je integrácia viacerých zdrojov značiek, prípadne aplikácia metód založených na ohodnotením podobnosti dvojíc stránok na základe ich obsahu.

## 8 Záver

---

V práci sme na základe analýzy problematiky navigácie na webe, vychádzajúc z identifikovaných problémov navigácie, navrhli metódu inkrementálnej adaptívnej podpory navigácie v priestoroch reprezentovateľných grafmi. Metóda je adaptívna v zmysle, že pri ohodnocovaní relevancie nasledovníkov pre odporúčanie vychádza z preferencií používateľa. Definovali sme špecializáciu navrhnutej metódy pre navigáciu v otvorenom priestore webu a overili sme predpokladané vlastnosti navrhnutej metódy. Tiež sme vykonali analýzu vybraných malých častí webu, kde sme skúmali charakteristiky dôležité pre fungovanie navrhnutej metódy. Špecifikovali sme požiadavky a navrhli sme systém pre adaptívnu podporu navigácie v priestore webu. Návrh systému sme overili jeho implementáciou. Navrhnutá metóda aj implementované riešenie sú doménovo nezávislé a použiteľné pre podporu navigácie v priestoroch definovaných grafom. Možnosti rozšírenia zahŕňajú prenos do inej domény, dodefinovanie iných typov ohodnocovacích metrík, spôsobov generovania grafu ako aj formy zobrazenia na strane klienta.

Vlastnosti navrhnutých metrík sme overili experimentmi nad dátami získanými z webu. Taktiež sme vykonali používateľskú štúdiu s cieľom overiť vnímanie užitočnosti navrhnutej metódy a jej implementácie vo forme aplikácie pre podporu navigácie na webe používateľmi. Výsledky používateľskej štúdie potvrdili stanovené hypotézy a používatelia vnímali riešenie ako jednoduché a užitočné. Ako hlavnú výhodu identifikovali objavenie nových stránok, ktoré predtým nepoznali. Tiež mali pocit, že použitie aplikácie implementujúcej navrhnutú metódu im nespôsobuje stratu kontextu pri nasledovaní odkazov.

Ako možné smery, ktorými by sa mohlo riešenie v budúcnosti ďalej vyvíjať, sme identifikovali uskutočnenie používateľskej štúdie veľkého rozsahu, kde by sme overili fungovanie metrík postavených na sociálnych stopách a návštevnosti stránok a vyhodnotenie efektívnosti týchto metrík v kombinácii s využitím značiek pre ohodnocovanie relevancie webových zdrojov s ohľadom na konkrétneho používateľa. Hlavným obmedzením riešenia, vyplývajúcim z momentálneho stavu webu, je dostupnosť značiek popisujúcich webové stránky. Možnosti usudzovania nad značkami sú momentálne tiež obmedzené, aj keď v tejto oblasti prebieha vo svete intenzívny výskum (Nakamoto et al., 2008).

## Literatúra

---

1. ABRAMS, D, BAECKER, R.: How people use WWW bookmarks, In: *CHI '97 extended abstracts on Human factors in computing systems: looking to the future*, Atlanta, Georgia, 1997.
2. BARLA, M.: Zachytenie záujmov používateľa na webe (Diplomová práca), Slovenská technická univerzita, 2006.
3. BERNERS-LEE, T. ET. AL.: Tabulator: Exploring and Analyzing linked data on the Semantic Web, *The 3rd International Semantic Web User Interaction Workshop*, Athens, Georgia, 2006.
4. BIELIKOVÁ, M.: Presentation of Adaptive Hypermedia on the Web. In Popelínský, L., ed.: *DATAKON 2003*, Brno, ČR, pp. 72–91, 2003.
5. BRUNK, B.: Overview and Preview Tools for Navigating the World-Wide Web, *SILS Technical Report TR-1999-03*, 1999.
6. BRUSILOVSKY, P.: Methods and techniques of adaptive hypermedia, In: *User Modeling and User-Adapted Interaction*, vol.6 no.2-3, pp. 87-129, 1996.
7. BRUSILOVSKY, P.: Adaptive Navigation Support, In: *The Adaptive Web*, Peter Brusilovsky, Alfred Kobsa, Wolfgang Nejdl (Ed.), Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, LNCS 4321, pp. 263-290, 2007.
8. BRUSILOVSKY, P., RIZZO, R.: Using maps and landmarks for navigation between closed and open corpus hyperspace in Web-based education. *The New Review of Hypermedia and Multimedia* 9, pp. 59-82, 2002.
9. BRUSILOVSKY P., CHAVAN, G., FARZAN, R.: Social Adaptive Navigation Support for Open Corpus Electronic Textbooks, In: *Adaptive Hypermedia and Adaptive Web-Based Systems*, Paul De Bra, Wolfgang Nejdl (Eds.), Third International Conference, AH 2004, Eindhoven, The Netherlands, LNCS 3137, pp. 24-33, 2004.
10. CARMAGNOLA, F, CENA, F., CORTASSA, O., GENA, C., TORRE, I.: Towards a Tag-Based User Model: How Can User Model Benefit from Tags?, In: *User Modelling 2007*, J. G. Carbonell, J. Siekmann (Eds.), 11th International Conference, UM 2007, Corfu, Greece, LNCS 4511, pp. 445-449, 2007.
11. CONKLIN, J.: Hypertext: an introduction and survey, *Computer*, vol.20 no.9, pp. 17-41, 1987.
12. GAMMA ET AL.: Design patterns: elements of reusable object-oriented software, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 1995.
13. GRUBER, T.: Ontology of Folksonomy: A Mash-up of Apples and Oranges, In: *Int'l Journal on Semantic Web & Information Systems*, vol.3 no.2, 2007.
14. HAGEN, P., MANNING, H., PAUL, Y.: Forrester survey: Must search stink?, *The Forrester report*, Forrester, 2000.



15. JEMALA, M.: Vizuálne prehľadávanie RDF dokumentov (Diplomová práca), Slovenská technická univerzita, 2006.
16. KAASTEN, S., GREENBERG, S.: Integrating Back, History and Bookmarks in Web Browsers, In: *Extended Abstracts of the ACM Conference of Human Factors in Computing Systems*, CHI'01, ACM Press, 2000.
17. KAPLAN, C., FENWICK, J., CHEN, J.: Adaptive hypertext navigation based on user goals and context. *User Modeling and User-Adapted Interaction*, pp. 193-220, 1993.
18. KATIFORI, A. ET AL.: A Comparative Study of Four Ontology Visualization Techniques in Protégé: Experiment Setup and Preliminary Results, In: *Proceedings of the Information Visualization*, IEEE CS DL, 2006.
19. KREUTZ, R. ET AL.: Improved Visual Navigation in Web-Documents, in Proc. *AACE ED-Media '98*, Freiburg, Germany, pp. 755-760, 1998.
20. LEVENE, M., WHEELDON, R.: A Web site navigation engine, In: *Poster Proceedings of International World Wide Web Conference*, Hong Kong, 2001.
21. LEVENE, M., WHEELDON, R.: Navigating the world-wide web. In: *Web Dynamics*, M. Levene and A. Poulovassilis, (Eds.), Springer-Verlag, 2003.
22. MAT-HASSAN, M., LEVENE, M.: Can navigational assistance improve search experience? A user study. *First Monday*, vol.6, no.9 (September 2001). [http://www.firstmonday.org/issues/issue6\\_9/mat/index.html](http://www.firstmonday.org/issues/issue6_9/mat/index.html) (online 20.11.2007).
23. MATSUO, Y., ISHIZUKA, M.: Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information, In: *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*, St. Augustine, Florida, USA, 2003.
24. NAKAMOTO, R., NAKAJIMA, S., MIYAZAKI, J., UEMURA, S.: Tag-Based Contextual Collaborative Filtering, In: *IAENG International Journal of Computer Science*, vol.34, no.2, 2008.
25. NÁVRAT, P., BARTOŠ, P., BIELIKOVÁ, M., HLUCHÝ, L., VOJTÁŠ, P.: *Tools for Acquisition, Organization and Presenting of Information and Knowledge*, Research Project Workshop, Bystrá Dolina, Low Tatras, Slovakia, 256 pages, 2006.
26. PATERNO, F., MANCINI, C. (1999). Designing web interfaces adaptable to different types of use. In: *Proc. Of the Workshop Museums an the Web*.
27. ROMERO, C., VENTURA, S., DELGADO J. A., DE BRA, P.: Personalized Links Recommendation Based on Data Mining in Adaptive Educational Hypermedia Systems, *LNCS 4753*, Springer Verlag, pp. 292-306, 2007.
28. SVÁTEK, V: Ontologie a WWW, *Datakon 2002*, Brno, pp.1-35, 2002.
29. THENG, Y. L., JONES, M., THIMBLEBY, H.: Lost in Hyperspace: Psychological problem or bad design?, <http://www.cs.waikato.ac.nz/oldcontent/mattj/lostinhyperspace.pdf> (online 8.12.2007).

30. USHOLD, M., GRUNINGER, M.: Ontologies: Principles, Methods and Applications. *The Knowledge Engineering Review*, vol.11 no.2, pp.93-136, 2006.
31. WHEELDON R.: The Best Trail Algorithm for Assisted Navigation of Web Sites, In: *Proceedings of the 1st Latin American Web Congress (LA-WEB'03)*, Santiago, Chile, 2003.
32. ZHOU, D., BIAN, J., ZHENG, SH., ZHA, H., GILES, C.L.: Exploring Social Annotations for Information Retrieval, *Proceedings of the 17th International World Wide Web Conference (WWW2008)*, April 21-25 2008, Beijing, China, 2008.

## Prílohy

---

## **A Obsah elektronického média**

---

/readme.txt – obsah elektronického média

/doc/dp-tomsa.doc - dokument vo formáte MS Office 2003

/doc/dp-tomsa.pdf - PDF verzia dokumentu

/data/experiments/ - výstupy vykonaných analýz štruktúry prepojení na webe vo fulltextovom a XML tvare

/data/tagmetricdata.xls - tabuľka s hodnotami z overenia funkcie podobnosti na základe značiek

/src/ - zdrojové kódy k prototypu a analyzačným nástrojom

/src/doc/ - technická dokumentácia k zdrojovým kódom

## B Technická dokumentácia

---

Technická dokumentácia vo forme opisu tried a ich metód sa nachádza v elektronickej podobe na priloženom elektronickej médiu. Tu uvádzame ukážky vybraných častí zdrojových kódov, kde ilustrujeme zaujímavé princípy a použitie vytvoreného kódu a formáty.

### B.1 Paralelný mechanizmus prístupu k službe del.icio.us

Získavanie grafu prepojení webových stránok nástrojom LinkCrawler je časovo náročný proces. Tak isto je kvôli obmedzeniam služby del.icio.us získavanie zoznamov značiek pre konkrétne URL v prípade väčšieho množstva vrcholov v grafe časovo náročné, pretože služba je nútená dodržať minimálne sekundový interval volaní del.icio.us API.

Preto sme implementovali paralelný mechanizmus prístupu k del.icio.us API, ktorý zabezpečuje, že značky sú získavane už počas behu crawlera priebežne pre URL, z ktorých sa vytvára výsledný graf.

Hlavné vlákno, zabezpečujúce výber úloh pre spracovanie a ich vykonanie. Úlohy môžu byť zaradené pre vykonanie v čase keď nie sú žiadne iné úlohy, alebo prioritne najskôr ako to bude možné.

Počas sťahovania grafu prepojení na webe sa zaraďujú úlohy do spracovania na pozadí a po skončení sa zaraďujú nové úlohy prioritne a tie, pre ktoré boli získané značky dopredu, sú prístupné cez vyrovnávaciu pamäť.

Metóda hlavného vlákna:

```
private void Run()
{
    DateTime lastCacheSave = DateTime.MinValue;

    while (run)
    {
        try
        {
            if (DateTime.Now.Subtract(lastCacheSave).TotalSeconds
                > CacheAutosaveInterval)
            {
                API.SaveCache();
                lastCacheSave = DateTime.Now;
            }
        }
        catch (Exception ex)
        {
            Console.WriteLine("error while autosaving cache: " +
                ex.ToString());
        }

        try
        {
            lock (aLock)
            {
```

```

        if (this.urls.Count > 0)
        {
            string url = this.urls.Dequeue();
            //stores into cache
            this.GetTagsInternal(url);
        }
    }
    Thread.Sleep(1);
    //na chvilku uvolnim lock nech si tam niekto zapise
    lock (aLock)
    {
        //vyberiem z idle vzdy len jednu
        if (this.urls.Count == 0 &&this.idleUrls.Count
> 0)
        {
            string url = this.idleUrls.Dequeue();
            this.idleSet.Remove(url);
            this.GetTagsInternal(url);
        }
    }
    Thread.Sleep(1);
}
catch
{
    throw;
}
}
}

```

Zápis do fronty úloh spracovaných na pozadí:

```

public void EnqueueIdle(string url)
{
    lock (aLock)
    {
        if (!cache.ContainsKey(url))
        {
            if (!idleSet.ContainsKey(url))
            {
                this.idleUrls.Enqueue(url);
                this.idleSet.Add(url, url);
            }
        }
    }
}
}

```

Prioritné spracovanie úlohy spolu s jej bezprostredným vrátením:

```

public List<Tag> GetTags(String url)
{
    //overenie spravnosti uri
    if (!Uri.IsWellFormedUriString(url, UriKind.Absolute))
    {
        return new List<Tag>();
    }
    url = new Uri(url).ToString();
    //zaradim do fronty
    lock (this.aLock)
    {
        if (cache.ContainsKey(url))
        {

```

```

        return cache[url];
    }
    else
    {
        this.urls.Enqueue(url);
    }
}

try
{
    int maxWait = 5000;
    int wait = 0;
    bool contains = false;
    //cakam kym bude dostupny vysledok, s timeoutom
    while (!contains && ++wait < maxWait)
    {
        lock (this.aLock)
        {
            List<Tag> res = null;
            contains = cache.TryGetValue(url, out res);
            if (contains)
            {
                return res;
            }
        }
        Thread.Sleep(2);
    }

    //po uplynuti timeoutu si to vezmem sam
    return this.GetTagsInternal(url);
}
catch
{
    return new List<Tag>();
}
return new List<Tag>();
}

```

Samotné volanie rozhrania služby del.icio.us:

```

private List<Tag> GetTagsInternal(String url)
{
    if (string.IsNullOrEmpty(url))
    {
        return new List<Tag>();
    }
    Uri uri = new Uri(url);
    url = uri.ToString();
    if (this.UseCache && API.cache.ContainsKey(url))
    {
        //aby nam nealiasovali
        return new List<Tag>(API.cache[url]);
    }

    string hashString = ComputeHash(url);

    //medzi dvoma volaniami dodrzim interval
    while (DateTime.Now.Subtract(lastQuery).TotalMilliseconds <
this.RequestInterval)
    {
        Thread.Sleep(250);
    }
}

```

```

    }

    //poziadavka na del.icio.us api
    try
    {
        Console.WriteLine("[{0}], {1}: making del.icio.us request for
{2}: ", Thread.CurrentThread.ManagedThreadId,
DateTime.Now.ToLongTimeString(), url);
        //metoda URL
        navigator.Navigate(urlBase + hashString.ToLower());
    }
    catch (Exception)
    {
        throw;
    }
    finally
    {
        lastQuery = DateTime.Now;
    }
    string content = navigator.Content;
    int idx = content.IndexOf(url);
    List<Tag> res = new List<Tag>();
    if (idx >= 0)
    {
        //extrakcia znaciek z JSON response sluzby
        res = ParseTags(content);
    }

    //ulozim do cache
    if (this.UseCache)
    {
        API.cache[url] = res;
    }

    Console.WriteLine("{0} tags", res.Count);

    return res;}

```

Rozhranie pre prístup k del.icio.us tiež implementuje vyrovnávaciu pamäť cache.

Vyrovnávacia pamäť je reprezentovaná dátovou štruktúrou `Dictionary<string, List<Tag>>`, kde kľúčom je URL, hodnotou je zoznam značiek, ktoré vrátila služba del.icio.us. Práca s vyrovnávacou pamäťou je naznačená v ukážke kódu vyššie

## B.2 Špecializácia triedy všeobecného navigátora

Implementované riešenie umožňuje priamočiare použitie triedy implementujúcej schému navigácie pre inú doménu. Tento princíp ilustruje nasledujúca ukážka zdrojového kódu, kde je odvodená trieda pre navigáciu na webe.

```

[ServiceBehavior(ConcurrencyMode=ConcurrencyMode.Multiple,
InstanceContextMode=InstanceContextMode.Single,
UseSynchronizationContext=false)]
public class WebNavigation : NavigationBase<WebLink, WebPage>,
IWebNavigation//Navigator<WebLink, WebPage>, IWebNavigation
{
    #region WebNavigation Singleton
    public static WebNavigation Instance

```



```

    {
        get
        {
            return NestedWebNavigation.instance;
        }
    }

    class NestedWebNavigation
    {
        // Explicit static constructor to tell C# compiler
        // not to mark type as beforefieldinit
        static NestedWebNavigation()
        {
        }

        internal static readonly WebNavigation instance = new
WebNavigation();
    }
    #endregion

    private Graph<WebLink, WebPage> graph;

    private ILogger logger;

    private WebNavigation() //: base(InitGraph())
    {
        this.graph = InitGraph();
        Trace.Listeners.Add(new
TextWriterTraceListener(Console.Out));
        Trace.AutoFlush = true;
        Trace.UseGlobalLock = true;
        this.logger = Logger.Instance;
    }

    private static Graph<WebLink, WebPage> InitGraph()
    {
        return new Graph<WebLink, WebPage>(new WebLinkFactory());
    }

    public void CleanUp()
    {
        try
        {
            Delicious.JSON.API.Instance.Stop();
        }
        catch { }
    }

    public WebContext Navigate(WebLink e)
    {
        Session<WebLink, WebPage> session = this.CurrentSession;
        Trace.WriteLine(string.Format("{0}: making navigation step
{1}", session.Owner.Name, e.ToString()));

        Context<WebLink, WebPage> result =
session.Navigator.Navigate(e);

        this.logger.LogNavigationStep(e.To.Url.ToString());

        return WebContext.Create(result);
    }
}

```

```

public double GetEvaluation(WebLink edge)
{
    Session<WebLink, WebPage> session = this.CurrentSession;

    Trace.WriteLine(string.Format("{0}: GetEvaluation {1}",
session.Owner.Name, edge.ToString()));
    return session.Navigator.GetEvaluation(edge);
}

public WebPageMetadata GetMetadata(WebPage vertex)
{
    Session<WebLink, WebPage> session = this.CurrentSession;

    Trace.WriteLine(string.Format("{0}: GetMetadata {1}",
session.Owner.Name, vertex.ToString()));
    return
session.Navigator.GetMetadata<WebPageMetadata>(vertex);
}

public override Session<WebLink, WebPage>
CreateSession(UserModel owner)
{
    if (owner == null)
    {
        throw new ArgumentNullException("owner");
    }

    return new Session<WebLink, WebPage>(owner, new
Navigator<WebLink, WebPage>(this.graph, new WebRelevanceEvaluator(),
WebBrowsingHistory.Instance));
}

public void SetDeliciousCredentials(string deliciousName, string
deliciousPassword)
{
    UserModel.CurrentUser.SetDeliciousName(deliciousName);

    UserModel.CurrentUser.SetDeliciousPassword(deliciousPassword);
}
}

```

# C Integration, installation and development manual

---

## C.1 Basic Information

The problem of being “lost in hyperspace” is persistent and well known for long time. Already in 1987, Conklin has identified two basic problems of hyperspace navigation: the loss of orientation and cognitive overhead. The loss of orientation occurs, in addition to other causes, as a result of inability to capture the link semantics while following the hypertext links. Furthermore, the links as we know them from the Web are single directional, so users lose their context immediately after following a link. The web pages appear from „nowhere“ and disappear somewhere again after reading.

The adaptive web navigation support tool (AWNS) offers a support in form of visual representation of current page neighborhood as well as immediate session history. Navigation is represented as a process of graph discovery, where edges are constructed incrementally as neighborhood of the actual vertex and are evaluated by using various metrics. Metrics are taking specific user preferences (represented as user model) into account, making the solution adaptive.

### C.1.1 Method Description

*AWNS* is based on incremental browsing paradigm, where the graph representing the information space is visualized incrementally employing a sliding window specifying subset of the information space to be visualized.

The user specifies the starting node (the web page URL). In every step, the neighborhood of the current node is generated. Metrics for edge relevance are evaluated and the values are aggregated using weighted sum for every edge. Edges are sorted and filtered based on the computed value of relevance.

The outline of the navigation method follows.

#### **Method 1 Web navigation**

```
Specify the node v to be presented (manual select by the user based or  
select the node connected with previously displayed node by a link with  
the highest importance)
```

```
Generate a vicinity of the node v:  
GenerateNeighborhood (v)
```

```
Add generated entities to the current window w
```

```
For each edge e in w: Evaluate(e)
```

```
Apply filters and sort: Sort(Filter(w))
```

### **C.1.2 Scenarios of Use**

While *AWNS*, as a tool presented in this document, is a specialization of a generic navigation method for web navigation, the method itself can be used for navigation in any information space that can be represented as a graph. Hence, *AWNS* (the web implementation of the navigation support method) can be used in the following scenarios:

- Users browse the information space (part of the Web) which is previously unknown to them
- Users are performing a collaborative effort in an exploratory search scenario

*AWNS* should not be used in following cases:

- There is no graph representation of the navigated space or the graph is very simple (a path)
- The graph is very dense (has high average vertex degree) and a good evaluation function determining edge relevance cannot be specified

### **C.1.3 External Links and Publications**

WCF Configuration schema - <http://msdn2.microsoft.com/en-us/library/ms731734.aspx>

WCF Configuration editor tool - <http://msdn2.microsoft.com/en-us/library/ms732009.aspx>

## **C.2 Integration Manual**

*AWNS* is developed in C# 3.0 programming language using Visual Studio 2008 and .NET Framework 3.5 as a WPF (Windows Presentation Foundation) application on the client side and WCF service hosted in a (Windows Communication Foundation) is used for communication. The distribution of *AWNS* consists of these parts:

- *AWNS.ServiceLibrary*, library with core components
- *AWNS.ServiceLibrary.Host.Web*, a console application hosting the navigation service
- *AWNS.Web*, the WPF client application
- A set of *app.config* configuration files for each executable (both the hosting application and the client application)
- A script to generate *AWNSdb*, SQL server database (Microsoft SQL Server 2005) for caching and as data store for selected metrics

The navigation service can be hosted in the WPF client application or as a separate. Client is configured to connect to either its local navigation service or a navigation service deployed on the Web. Currently, due to restrictions of the del.icio.us public API, it is recommended to configure *AWNS* navigation service to run in the client process. A SQL server database can be configured to act as a global cache. Caching behavior is configured in the *app.config* configuration file for the application hosting the navigation service.

## C.2.1 Dependencies

AWNS uses these external tools and libraries:

- *.NET Framework 3.5*, selected assemblies
- *Del.icio.us.NET*, a library providing access to del.icio.us public API
- *SemanticLog* logging service (optional)
- SQL Server 2005 (optional)

## C.2.2 Installation

There are two basic models of installation, depending on where the navigation service is deployed. It can be deployed either as a stand-alone service (not recommended due to performance reasons implied by del.icio.us API) or as a service hosted in the client application.

Before deploying AWNS the following prerequisites should be met:

- .NET Framework 3.5 must be installed on both client and server machine
- SQL Server 2005 can be installed (optional)

In case that navigation service is hosted as a separated application, deploying AWNS involves these steps:

1. [optional] A database named AWNSdb with corresponding tables in the delivered AWNSdb.sql script are created on the machine hosting the SQL Server
2. The navigation service is copied to the server machine
3. The client application is copied to the client machine

## C.2.3 Configuration

AWNS navigation service must be configured via the supplied app.config files corresponding to the application the service is hosted in. To configure the navigation to use the SQL Server as a global cache, the following section of the app.config file should be located

```
<setting name="UseSQLCache" serializeAs="String">
  <value>True</value>
</setting>
```

And the value should be set to True or False depending on whether the SQL Server is to be used as a global cache.

In case that the SQL Server has been configured and database has been deployed, the connection string used to connect to the SQL Server cache should be specified. The following section should be located

```
<connectionStrings>
  <add name="AWNS.Properties.Settings.AWNSdbConnectionString"
connectionString="Data Source=SERVERNAME\SQLEXPRESS;Initial
Catalog=AWNSdb;Integrated Security=True"
  providerName="System.Data.SqlClient" />
</connectionStrings>
```

Where SERVERNAME should be replaced with the actual path to the SQL Server.

In case service is deployed in separate application, the address of the server should be specified in both the client and server app.config file.

The following section of the server app.config file should be located

```
<service name="AWNS.ServiceLibrary.WebNavigation"
behaviorConfiguration="AWNS.ServiceLibrary.Service1Behavior">
  <host>
    <baseAddresses>
      <add
baseAddress="http://SERVERNAME/AWNS.ServiceLibrary/Navigation/" />
    </baseAddresses>
  </host>
```

and the public address and port of the server should be entered instead of SERVERNAME.

The following section in client app.config file specifies server address:

```
<client>
  <endpoint
address="http://SERVERNAME/AWNS.ServiceLibrary/Navigation/"
```

The WCF infrastructure can be configured via the app.config files as described in the WCF Configuration Schema (see External links section). Additional bindings, behaviors and endpoints can be specified.

The client app.config file which is part of the distribution is already configured with all the needed endpoints and in case the service is to be hosted in client application, the file can be left intact (except for optional SQL connection string mentioned above).

Note: In case that the navigation service is hosted in the client application, the configuration of both client and service endpoints is done in the same app.config file.

## C.2.4 Integration Guide

*AWNS* can be used as a navigational aid for browsing the Web. The user interface is shown in Fig. 1.

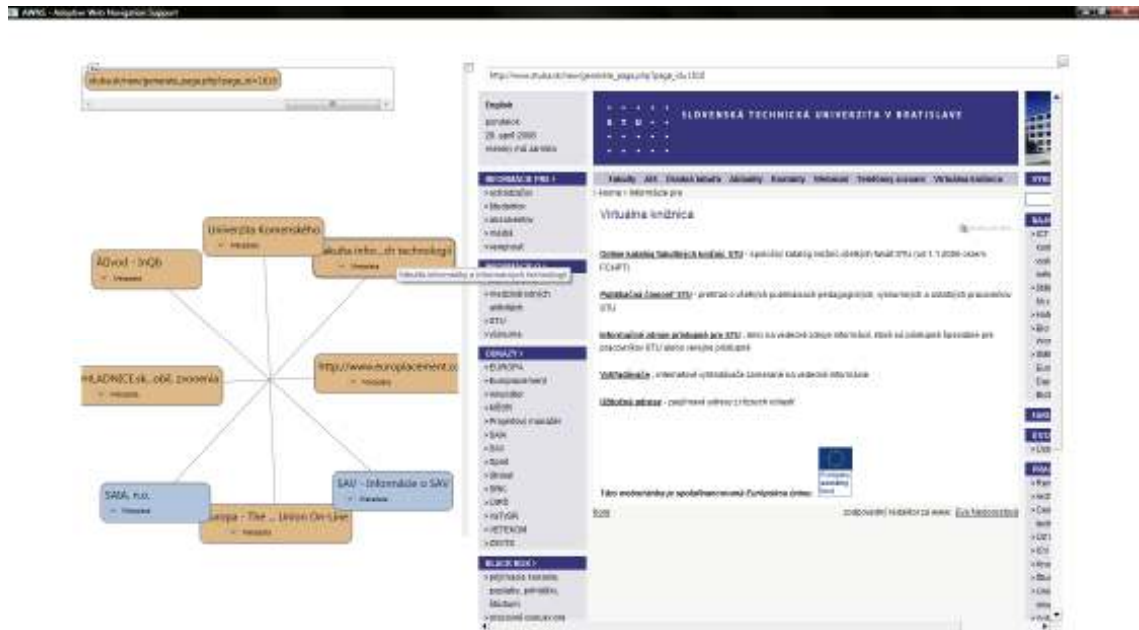


Fig 1: Graphical user interface.

Current web page is rendered on the left side. There is an address bar in the upper left. The navigation support in form of graph visualization is shown on the left side and recency-ordered session history is shown in the upper left area of the interface.

User can perform navigational steps either by following links on the rendered page, by clicking a node in the graph on the left side or by choosing a history entry. After new content is rendered in the browser, the navigation service is called and the neighborhood of the currently rendered page is visualized in the graph area on the left side. Links are added and swapped incrementally, as their evaluation becomes available. Due to implementation restrictions, the evaluation of a single URL takes at least one second. In parallel with link evaluation, node metadata is constructed and web page URLs are changed for their respective titles. For long titles, the tool tries to shorten them by removing prefixes and suffixes containing generic data (web page address, common among all subsites of the particular site) e.g. “SME.SK | A very very very very long title” becomes “A very ... long title”.

### Error handling

Possible errors can originate from bad configuration, external service unavailability network connectivity issues or database consistency violations. Most errors are handled in their respective exception handlers and selected errors are reported to end user by a messagebox. While the application can recover from most like web site being unavailable or del.icio.us functionality being denied by the provider of the del.icio.us service, some may require restarting the client application, or event changing the IP address of the machine that is making requests to del.icio.us service (the machine where the service is hosted, either the server or client machine, depending on configuration). Service application reports error information to console.

There are known timeout errors related to WCF infrastructure. Inactivity timeout for the reliable session held by WCF for the client machine is set 9 hours by the app.config file. After this time, a `TimeoutException` is thrown, making further calls from client invalid. This is by design and requires restart of the client application.

## C.3 Development Manual

### C.3.1 Tool Structure

*AWNS* consists of the following packages:

- `AWNS.ServiceLibrary` – the core assembly, base of server side implementation
- `AWNS.ServiceLibrary.Host.Core`
- – library containing reusable components that can be used for service hosting and fault handling
- `AWNS.ServiceLibrary.Host.Web` – a console application configured to host the web navigation service as a stand-alone application
- `AWNS.ServiceLibrary.Host.Publications` – a console application configured to host the publication navigation service as a stand-alone application
- `AWNS.Web` – the client application implemented as a XAML Browser Application
- `AWNS.Web.App` – the client application implemented as a WPF stand-alone application
- `AWNS.Web.Controls` – library containing client graphical user interface components
- `Baseline` – a web browser implementation with basic navigational functionality based on Back/Forward button and link following with logging support
- `Del.icio.us.NET` – a 3<sup>rd</sup> party library providing access to del.icio.us social bookmarking service, source code modified for greater flexibility and additional functionality

### C.3.2 Method Implementation

Fig 2 depicts the processing of a single navigational step using the adaptive navigation support method. User enters the URL of the starting page and the user interface passes this event to the navigation web service. The step is stored in the corresponding user model, and `WebCrawler` is used to gather list of edges connecting adjacent vertices. Evaluation is computed for each edge by using the `RelevanceEvaluator` class, which iterates through all the registered evaluation metrics and computes the weighted aggregated sum of the evaluation. The result is sorted and filtered based on the value of the evaluation and the result is returned to the client application to be displayed using the graphical user interface.



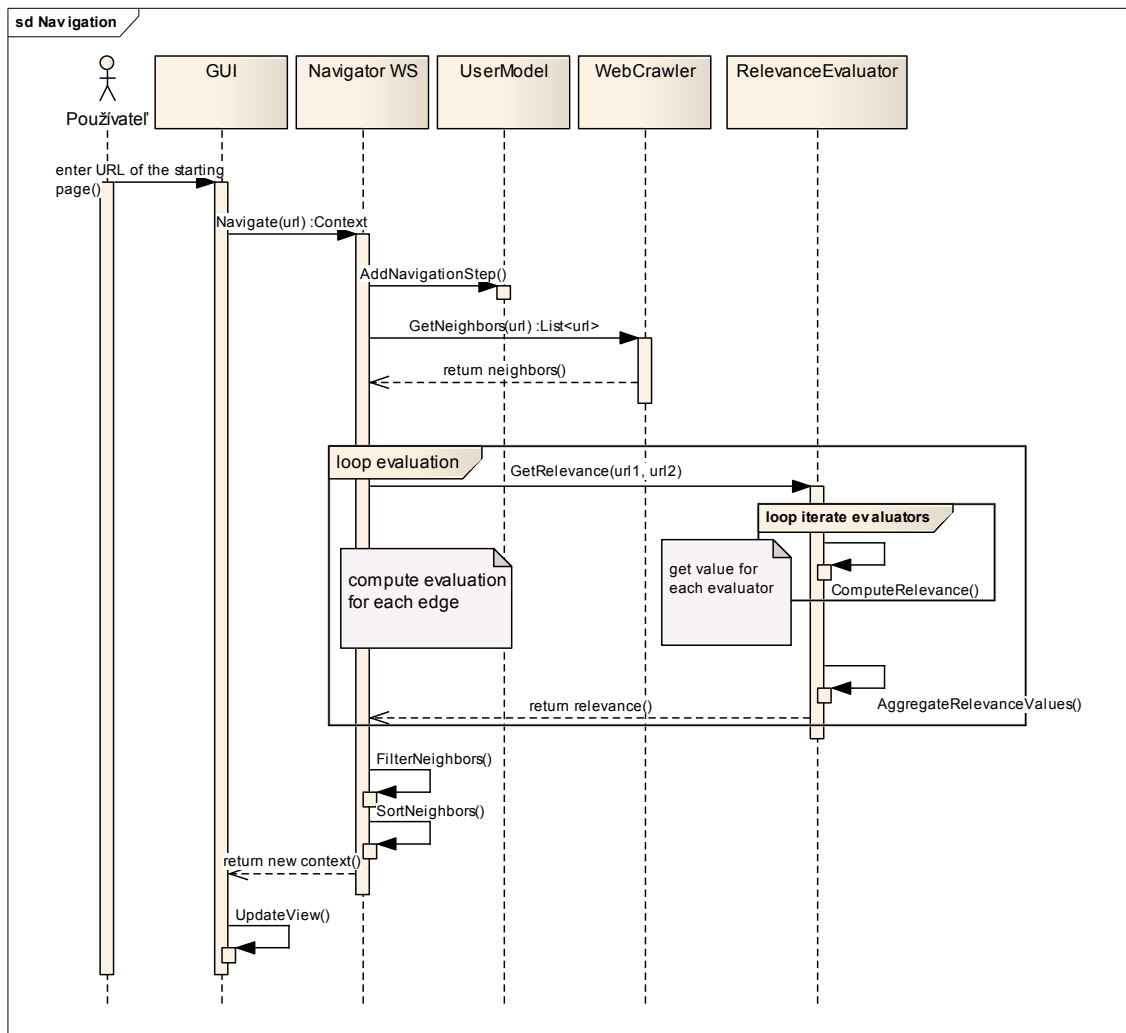


Fig 2 : Sequence diagram illustrating processing of a single navigation step using the adaptive navigation support method.

### C.3.3 Enhancements and Optimization

During the development of the AWNS tool, several performance optimizations often represented by design tradeoffs had to be made. The main bottleneck has shown to be the del.icio.us API. This allows requests no more often than 1 second from any given IP address. This fact made the implementation of evaluation metrics on the server impossible, because client requests would take progressively longer with the increasing number of connected clients. This fact had forced us to host the navigation service in the client application, and centralize the service state handling in a relational database where the client-hosted service implementations connect.

Further in the development, multiple cache mechanisms were implemented to cache web page titles, del.icio.us tags associated with urls and user model represented by tags he is using, as these aspects of the implementation have proven to be a serious bottleneck in performance.

Client application is already implemented to communicate with the service in an asynchronous manner, so the service calls do not delay processing on the client side.

Graph is constructed and filtered incrementally, which causes the state of the graph to update in approx. 1 second intervals (due to del.icio.us request interval).

## **C.4 Manual for use in Other Application Domains**

The navigation method used in the *AWNS* tool is universal and thus domain independent. In general, it can be used in any domain for which a graph representation is available, and where the neighborhood and evaluation function can be supplied.

In order to adapt the implementation of navigation service to other domain, following steps must be made:

- The neighborhood function must be specified
- The evaluation function must be supplied
- Optionally, a data representation of the vertices and edges of the graph and vertex metadata can be supplied (otherwise, supplied generic `Vertex`, `Edge` and `VertexMetadata` classes can be used)
- An interface consisting of methods specific to the navigation in the particular domain should be declared (this is due to implementation restrictions imposed by the WCF library where generic types are causing problems)
- A class inheriting from `NavigationBase<TEdge, TVertex>` and implementing the declared interface should be declared

The adapted navigation service can then be instantiated and hosted.

Depending on the hosting environment, faults should be processed accordingly. `ServiceManager` class handles most by outputting the information to the Console and it is advisable to use it to host the service in case the hosting environment is a console application.

A code sample demonstrating adaptation to a different domain is part of the distribution in the `AWNS.ServiceLibrary.Publication` assembly.

### **C.4.1 Configuration for use in Other Application Domains**

After implementation steps described in previous section, no more configuration (except as described in the installation manual) are required.

### **C.4.2 Dependencies**

Dependencies used in the described implementation of the tool are either domain dependent (the `del.icio.us.NET` library) and thus cannot be employed in implementation in other domain, or domain independent and thus do not need any additional configuration.

## **D Publikácie vzťahujúce sa k výsledkom diplomovej práce**

---

### **Hyperlinks Visualization Using Social Bookmarking**

Príspevok v spoluautorstve s prof. Máriou Bielikovou prijatý do posterovej sekcie na konferencii Hypertext 2008 v Pittsburghu.

### **Improving Navigation by Web Page Context Visualization**

Príspevok prijatý na konferenciu IIT.SRC 2008 v Bratislave.

### **Adaptívna navigácia v ontológiách**

Príspevok v spoluautorstve s prof. Máriou Bielikovou prijatý na konferenciu WIKT 2007 v Košiciach.

**Sem sa vloží príspevok z HT**

**Sem sa vloží príspevok z IIT.SRC**

**Sem sa vloží príspevok z WIKTu**

## **E Návrh článku**

---