

Slovenská technická univerzita v Bratislave  
FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ  
Ilkovičova 3, 812 19 Bratislava



# Tímový projekt

## Ponuka

Grafická podpora vyhľadávania znalostí v dokumentoch

Tím 12:

Bc. Marian Beňo, Bc. Miloš Blaško, Bc. Ľubomír El'ko, Bc. Ján Kmeťko, Bc. Lukáš Lazarcík, Bc. Tomáš Mičko

**Kontakt:**

[sproty@googlegroups.com](mailto:sproty@googlegroups.com)

## Obsah

---

1	Zloženie tímu .....	1
2	Motivácia .....	2
3	Koncepcia riešenia .....	3
	Príloha A - Zoradenie všetkých tém podľa priority .....	4
	Príloha B - Aktuálny rozvrh všetkých členov tímu.....	5

# 1 Zloženie tímu

---

Tím pozostáva zo 6 študentov inžinierskeho študijného programu Softvérové inžinierstvo, pričom všetci členovia úspešne absolvovali bakalársky stupeň štúdia na FIIT v programe Informatika.

**Bc. Marián Beňo** - Od začiatku roka pracuje ako technický konzultant pre firmu Metalogix software, ktorá sa zaoberá archiváciou e-mailov a dát. Táto práca mu priniesla poznatky z oblasti Exchange servera, Windows serverov 2003/2008, SQL serverov 2005/2008 a s tým spojenými databázami. V bakalárskom projekte vytvoril regionálny informačný systém, kde využil technológie HTML, CSS, PHP a MySQL. Neskôr by sa rád venoval tvorbe web stránok a web aplikácií a rád by rozšíril svoje vedomosti v oblasti databáz. Z tohto dôvodu začal študovať technológiu JSP/Wicket a Hibernate.

**Bc. Miloš Blaško** - Počas štúdia si osvojil najmä programovacie jazyky C a Java. Väčšinu projektov vrátane bakalárskej práce (s témou „Rozpoznávanie hlasových povelov“) vypracoval práve v Jave, a tak získal rôznorodé skúsenosti napríklad s IDE Eclipse alebo 3<sup>rd</sup> party knižnicami. Má bohaté aj menej bohaté skúsenosti s jazykmi SQL (MS Access, MySQL), HTML, CSS, PHP, JavaScript, ActionScript 3 a Flash. V práci si rozšíril svoje vedomosti z Javy o technológie Hibernate, Spring, Struts, JSP a zoznámil sa s SVN a metodikou SCRUM. Od projektu očakáva nadobudnutie nových skúseností s vývojom softvéru v rôznorodom tíme.

**Bc. Ľubomír Epko** – Ovláda viacero programovacích jazykov (Pascal, C, Delphi, Java, Prolog, Lisp, Assembler) a prístupov k programovaniu (procedurálne, OO, funkcionálne či logické). Má skúsenosti s webovými technológiami HTML a CSS, relačnými databázami a modelovaním v jazyku UML. Rád sa učí nové technológie čomu nasvedčuje aj úspešne vypracovaná bakalárska práca v prostredí Linux (QEMU ako plugin pre webový prehliadač), za ktorú dostal pochvalu dekana a zúčastnil sa s ňou na konferencii IIT.SRC 2009.

**Bc. Ján Kmet'ko** - Dva roky pracuje vo firme PosAm ako Java programátor. Pred tým sa dva roky venoval programovaniu back-endu web stránok v PHP. Bohaté skúsenosti s programovaním J2EE aplikácií a prácou v tíme (JSE, Spring, Hibernate, Wicket, JSP, Ant, Maven, SVN, SQL). Na niekoľkých projektoch pracoval aj v roli analytika. Víťaz PosAm Java akadémie 2007. V bakalárskom projekte navrhol a vytvoril systém pre FIIT na hodnotenie výučby študentmi (TeaEval), ktorý v týchto dňoch ide do ostrej prevádzky. V budúcnosti by chcel pôsobiť v roli analytika, pre jeho dobré abstraktné myslenie.

**Bc. Lukáš Lazarčík** - Počas štúdia získal skúsenosti s viacerými programovacími jazykmi, najmä s platformou Java SE, tieto vedomosti ďalej prehľbuje štúdiom JEE technológií ako JSP, Hibernate, Spring MVC. Okrem Java technológií prišiel do styku s rôznymi webovými technológiami ako PHP, Javascript, jazykom SQL a databázovým serverom MySQL. V rámci bakalárskej práce riešil projekt Analýza logu webového servera, v ktorom sa zaoberal správaním užívateľov na webovom portáli. Od tímového projektu očakáva, že už nadobudnuté znalosti prehľbí, získa ďalšie a využije ich aby výsledný produkt bol na vysokej úrovni.

**Bc. Tomáš Mičko** – Počas štúdia dosahoval nadpriemerné výsledky, čo svedčí o jeho chuti učiť sa nové veci a o zmysle pre zodpovednosť a spoľahlivosť. Na úrovni školských projektov získal praktické skúsenosti s programovacími jazykmi C, Java SE, SQL, XML. V bakalárskej práci sa zaoberal využitím modelov v softvérových projektoch, kde získal prehľad o nových metodológiách modelovania softvéru.

## 2 Motivácia

---

S pribúdajúcim časom sa hromadí množstvo dokumentov, z ktorých môžeme čerpať dôležité informácie a preto je potrebné, aby sme sa zamerali na ich efektívne vyhľadávanie a spracovanie. Veľmi zaujímavým prístupom je práve vizualizácia výsledkov hľadania, ktorá prehľadne znázorňuje vzájomné prepojenia dokumentov. Práve tieto informácie totiž autori potrebujú a využívajú pri tvorbe nových dokumentov.

Po oboznámení sa s podobnými projektmi (napr. [www.foaf.sk](http://www.foaf.sk)) nás táto téma natoľko zaujala, že by sme ju radi spracovali ako tímový projekt. Na základe vlastných skúseností pri vyhľadávaní dokumentov vieme aké ťažké je filtrovať získané výsledky. Veríme, že naše nápady a postrehy by mohli byť prínosom v spomínanej oblasti.

Všetci členovia nášho tímu majú niekoľkoročné skúsenosti s programovacím jazykom Java. Niektorí z nás v tomto jazyku úspešne vypracovali svoje bakalárske práce, čo nás istou mierou predurčuje k úspešnému zvládnutiu tohto zadania. Vypracovaním projektu by sme si s určitosťou prehĺbili a doplnili zručnosti z každej etapy tvorby softvérového produktu a oboznámili sa s novými technológiami. Zároveň máme množstvo nápadov a vízií ako efektívne, pútavo a organizovane prezentovať výsledky hľadania koncovým používateľom.

Väčšina z nás má zapísaný predmet Objektovo - orientovaná analýza a návrh softvéru. Práve poznatky ako vzory, refactoring alebo prefactoring by sme mohli zužitkovať priamo na tomto tímovom projekte, a tak by sme získali veľmi cenné praktické skúsenosti.

### 3 Konceptia riešenia

Riešenie tohto projektu vidíme v rozdelení na dve hlavné časti:

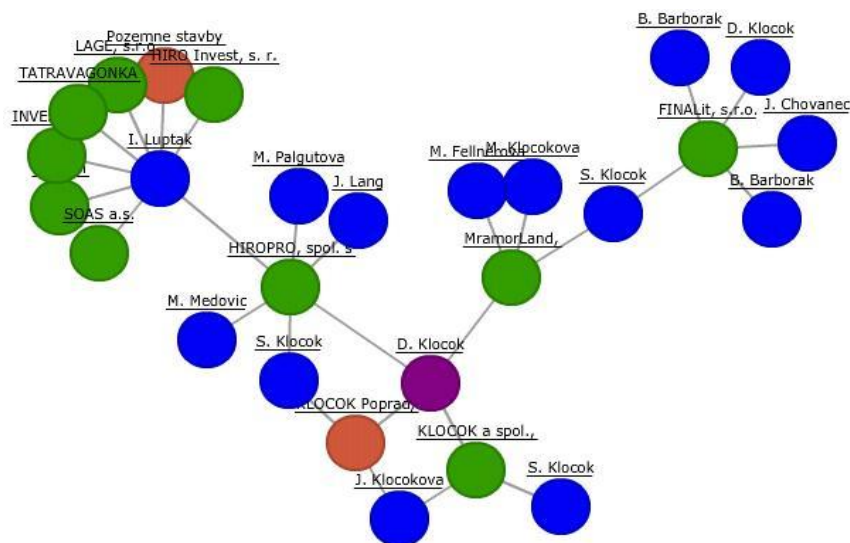
- Vyhľadávanie, indexovanie a prepájanie dokumentov vzhľadom na ich vzájomnú relevanciu
- Vizualizácia prepojených vyhľadaných dokumentov

Takéto rozdelenie sa prejaví pri klient - server architektúre, keďže prvá časť bude sústredená na strane servera a druhá na strane klienta.

Dokumenty budú uložené v databáze spolu s informáciami o nich (ako napr. kľúčové slová). Do databázy sa budú vkladať ručne, prípadne môže server prehľadávať internet a jednotlivé dokumenty sťahovať a zároveň ich kategorizovať do skupín podľa ich obsahu. Server bude postupne jednotlivé dokumenty analyzovať a na základe ich vzájomnej relevancie ich prepájať. Vyhľadávanie a prepájanie sa bude realizovať viacerými algoritmi, medzi ktorými si bude môcť používateľ vybrať alebo ich kombinovať.

Na druhej strane bude mať klient na starosti vizualizáciu dokumentov. Dokumenty a ich prepojenia budeme zobrazovať pomocou grafu. Uzly budú reprezentovať dokumenty a čiary prepojenia. Tu existuje veľa možností ako zobrazovať prepojenia tak, aby bolo prehľadne prezentované, ktoré prepojenia majú podobný alebo rozličný charakter. Napr. prepojenie „rovnaký autor“ môže byť prezentované ako spojovacie čiary rovnakej farby iného odtieňa, resp. čím väčšia vzájomná relevancia tým hrubšia čiara. Jednotlivé uzly a prepojenia v grafe bude môcť používateľ presúvať, mazať, pridávať alebo definovať nové. Naším cieľom je poskytnúť užívateľsky prívetivú prezentáciu hierarchie dokumentov, pričom používateľ si bude môcť prispôbiť grafickú reprezentáciu jednotlivých prepojení (farbu, veľkosť, tvar, štýl a pod.).

Na nasledujúcom obrázku č. 1 vidíme vizualizáciu použitú na stránke [www.foaf.sk](http://www.foaf.sk), ktorou by sme sa chceli do istej miery inšpirovať.



Obr. 1. Vizualizácia prepojení z [www.foaf.sk](http://www.foaf.sk)

## **Príloha A - Zoradenie všetkých tém podľa priority**

---

1. Grafická podpora vyhľadávania znalostí v dokumentoch (Dokumenty)
2. Webové stránky pre cestovnú kanceláriu (Cestovka)
3. Informačný systém stredných škôl (SS IS)
4. Textový editor obohatený o grafické prvky (Editor)
5. Portál pre časopis (Časopis)
6. Digitálne mapy (Digmapy)
7. Evidencia publikačnej činnosti (EPCA) (EPCA)
8. Automatizovaná podpora predmetu z oblasti programovania (DSAPodpora)
9. Využitie sociálnych sietí pri vytváraní pracovných tímov (Sociálne siete)
10. Mobilný cestovný poriadok pre iPhone (Mobilný Poriadok)
11. RoboCup tretí rozmer (RoboCup 3D)
12. Hierarchická wiki s právami (Wiki)
13. Elastické komunikačné centrum (EKCentrum)
14. Webový portál pre zdravotne postihnutých občanov (ZŤP Portál)
15. Vizualizácia softvérových artefaktov v 3D priestore (3DVizual)
16. Virtuálna FIIT (VFIIT)
17. Dizajn s použitím obohatenej reality (ARDizajn)
18. Web 2.0 v knižniciach alebo od OPACu k portálu (DLPortál)
19. Knižnica (Knižnica)
20. Podpora kontroly plagiarizmu (Plagiarizmus)
21. Imagine Cup 2010: Game Design (IC Game Design)

## Príloha B - Aktuálny rozvrh všetkých členov tímu

	7:00 - 7:50	8:00 - 8:50	9:00 - 9:50	10:00 - 10:50	11:00 - 11:50	12:00 - 12:50	13:00 - 13:50	14:00 - 14:50	15:00 - 15:50	16:00 - 16:50	17:00 - 17:50	18:00 - 18:50	19:00 - 19:50	20:00 - 20:50	
PO						Pokročilé databázové technológie		Objektovo orientovaná analýza a návrh softvéru		Tvorba softvérového systému v tíme		Výskum softvérových systémov			
						Účtovníctvo									
UT	Kódovanie					Účtovníctvo		Manažment projektov softvérových a informačných systémov		1 Manažment projektov softvérových a informačných systémov		2 Manažment projektov softvérových a informačných systémov			
ST					Aspektovo - orientovaný vývoj softvéru								Aspektovo - orientovaný vývoj softvéru		
ŠT	Kódovanie		Návrh prekladačov		Návrh prekladačov		Architektúra softvérových systémov			Objektovo orientovaná analýza a návrh softvéru		Objektovo orientovaná analýza a návrh softvéru			
PI			Pokročilé databázové technológie												

1-2 členovia tímu	Preferovaný čas
3-4 členovia tímu	Jeden z termínov voľný
5-6 členovia tímu	Zamestnanie