

Zápis zo stretnutia č. 5

Názov projektu:	Znalostný manažment na báze technológie .NET
Vedúci projektu:	Ing. Ivan Polášek, PhD.
Téma stretnutia:	Analýza myšlienok, ktoré budú použité v projekte
Miesto stretnutia:	Softvérové štúdio D07b, FIIT STU
Dátum stretnutia:	07. 11. 2007
Čas stretnutia:	15:00 – 17:30
Prítomní:	Ing. Ivan Polášek, PhD. Bc. Daniel Princzkel Bc. Lukáš Sim Bc. Michal Pažitný Bc. Zdenko Porubčan
Hostia:	Jaroslav Košťal Juraj Kojda
Neprítomní:	Bc. Vladimír Mlynarovič Bc. Peter Lauro
Vypracoval:	Bc. Michal Pažitný

Priebeh stretnutia

1. Keďže projekt Lucky Number 7 stále nie je použiteľný, rozhodli sme sa začať v tvorbe nového systému.
2. So študentmi, ktorí riešia tému Dolovania dát vo svojich Bakalárskych projektoch zrejme nebudeme spolupracovať.
3. V novom systéme bude použitý Clips aj Lucene (či už pre .NET alebo Javu).
4. Clips nepracuje s pravdepodobnosťou, zodpovedný člen tímu dostal za úlohu premyslieť implementáciu.
5. Členovia tímu zreferovali svoje kapitoly z Vybraných štúdií:

Kapitola 2 - Vlado - neprítomný

Ostatní sme prebrali:

Odkazovanie na iné dokumenty na základe podobnosti, ktoré budú medzi sebou asociované – podľa ďalších bodov (zhluky, projekty, ľudia a pod.).

Objasnené čo sú obalovače, ale ak ich použijeme, tak maximálne na deľbu dokumentu na kapitoly.

Kapitola 3 - Zdeno objasnil page ranking, existujúce vzorce na výpočet relevancie, ale je to určené pre priestor, v ktorom sú stránky prepojené cez odkazy.

Algoritmy – rankovacie a prepojovacie.

Zhlukovanie – kategorizácia, dvojrozmerné mapy.

So zhlukovaním spomínané neurónové siete.

Opísal nástroje na vytváranie ontológií.

Uvažovali sme, či použijeme neurónovú sieť, ale zatiaľ sme tento nápad ani neakceptovali ani nezamietli.

Kapitola 4 - Dano

Uchovávanie vyhľadávaných dokumentov používateľa, vytvoril by sa nejaký profil, na základe ktorého by inteligentný systém vyberal vhodné dokumenty.

Hlasovanie.

V druhej sekcii bola načrtnutá predpríprava – čo používateľ používa, to je v systéme natiahnuté, aby to bolo hneď použiteľné, ale túto funkčnosť nebudeme implementovať, pretože tento princíp nepovažujeme za hlavný.

Kapitola 5 - Simo

Sociálne siete, komunity na webe – je to podobné ako so zhlukovaním. Opäť je to určené pre web, ale my by sme mohli uvažovať nejaké dokumentácie, ľudia, tímy, projekty.

Ponuky na základe podobnosti typov projektov, dokumentácií a podobne.

Node ranking – ohodnotenie väzieb (vzťahov) medzi jednotlivými objektmi – kapitoly, dokumenty, a pod.

Kniha o *data mining-u* – Mišo

Po preštudovaní zistil, že zatiaľ nám v tvorbe projektu nepomôže, je tam vysvetlené indexovanie a vyhľadávanie na mnohých modeloch a vzorcoch. Použitím Lucene sa vyhneme potrebe implementovať vlastné indexovanie a vyhľadávanie dokumentov.

6. GnomeDataMine sme odmietli, pretože systém má podporu v oblasti data mining-u v triedach Lucene.
7. Dohoda, čo bude v analýze, špecifikácii a návrhu.
8. Stanovili sme si úlohy (tab. Nové úlohy)
9. Najbližšie stretnutie (6) je v stredu 14. 11. 2007 o 14:45 v softvérovom štúdiu D07b, FIIT.

Staré úlohy

ID	Úloha	Zodpovedný	Zadaná	Termín	Stav
1.4	Štúdium dokumentov predchádzajúceho tímu. Inštalácia a spustenie systému. Overiť súčinnosť app. servera s klientom a modulmi (Vyhľadávací, Znalostný).	všetci	12. 10. 2007	06. 11. 2007	zrušená
2.6	Referát zdrojového kódu. Rozdelenie úlohy:	všetci	19. 10. 2007	06. 11. 2007	zrušená
	\\src\BuilderAdministrationApplication	Lukáš Sim, Vladimír Mlynarovič	24. 10. 2007	06. 11. 2007	zrušená
	\\src\server_side	Michal Pažitný, Zdenko Porubčan	24. 10. 2007	06. 11. 2007	zrušená
	\\src\client_side	Daniel Princzkel	24. 10. 2007	06. 11. 2007	zrušená
3.1	Doplniť dokumentáciu o spôsobe inštalácie celého	všetci	24. 10. 2007	17. 12. 2007	zrušená

	systemu.				
3.3	Referovanie do research článkov kapitoly knihy <i>Štúdie vybraných tém softvérového inžinierstva (2)</i> a knihy <i>Data Mining</i> . Člen tímu zreferuje pridelenú tému a navrhne možnosť jej rozvoja. Rozdelenie úlohy:	všetci	24. 10. 2007	06. 11. 2007	splnená
	Kapitola 2	Vladimír Mlynarovič	24. 10. 2007	06. 11. 2007	splnená
	Kapitola 3	Zdenko Porubčan	24. 10. 2007	06. 11. 2007	splnená
	Kapitola 4	Daniel Princzkel	24. 10. 2007	06. 11. 2007	splnená
	Kapitola 5	Lukáš Sim	24. 10. 2007	06. 11. 2007	splnená
	Kniha <i>Data Mining</i>	Michal Pažitný	24. 10. 2007	06. 11. 2007	splnená
3.4	Vyhľadať vhodné zdroje na www.acm.org www.computers.org	všetci	24. 10. 2007	06. 11. 2007	zrušená
4.1	Overiť, či vie Clips pracovať s pravdepodobnosťou (ľavá strana pravá strana)	Zdenko Porubčan	31. 10. 2007	06. 11. 2007	splnená
4.2	Objasniť spôsob, akým sú vkladané relevancie zväzkov dokumentov.	všetci	31. 10. 2007	06. 11. 2007	splnená
4.3	Preskúmať použiteľnosť GnomeDataMine a pravidiel z kapitoly 3.	Zdenko Porubčan	31. 10. 2007	06. 11. 2007	splnená
4.4	Doplniť systém o pridanie sekundárnych kľúčových slov podľa typu projektu	všetci	31. 10. 2007	Letný semester	zrušená

Nové úlohy

ID	Úloha	Zodpovedný	Zadaná	Termín	Stav
5.1	Premyslenie implementácie pravdepodobnosti v Clipse (použitie funkcií <i>and</i> , <i>or</i> , <i>not</i> pri zadávaní dotazov).	Zdenko Porubčan	7. 11. 2007	14. 11. 2007	otvorená
5.2	Naštudovať si parametrizovanie indexácie dokumentov v Lucene (hlavne či je možné obmedziť počet kľúčových slov z jedného dokumentu)	Michal Pažitný	7. 11. 2007	14. 11. 2007	otvorená
5.3	Z vybraných štúdií vybrať niekoľko myšlienok (6-8) a zvoliť koncepciu pre	všetci	7. 11. 2007	14. 11. 2007	otvorená

	kvalitnejšie vyhľadávanie ako mali LN7 (nové vzorce, metriky), vytvoriť dômyselnejší algoritmus na výpočet relevancie dokumentov				
5.4	Definovanie dynamických ontológií jednotlivých dokumentov, ktoré použijeme v znalostnom systéme (názov dokumentu, autor, obsah; zdrojové kódy, technická dokumentácia, používateľská príručka a pod.).	Zdenko Porubčan	7. 11. 2007	14. 11. 2007	otvorená
5.5	Analyzovať a navrhnuť modul pre spracovanie a vyhodnocovanie relevancie dokumentov na základe <i>node ranking-u</i> .	Lukáš Sim	7. 11. 2007	14. 11. 2007	otvorená
5.6	Analyzovať a navrhnuť modul pre spracovanie a vyhodnocovanie relevancie dokumentov na základe <i>profilovania používateľov (sociálne siete)</i> .	Daniel Princzkel	7. 11. 2007	14. 11. 2007	otvorená
5.7	Analyzovať a navrhnuť modul pre indexáciu a vyhodnocovanie relevancie dokumentov použitím knižníc Lucene.	Michal Pažitný	7. 11. 2007	14. 11. 2007	otvorená
5.8	Spracovať analýzu projektu Lucky Number 7.	Vladimír Mlynarovič	7. 11. 2007	14. 11. 2007	otvorená
5.9	Vytvorenie analýzy, špecifikácie a návrhu projektu – každý člen tímu prispieva textom o jemu pridelenej časti systému. Vytvorená dokumentácia musí byť odovzdaná vedúcemu...	všetci	7. 11. 2007	14. 11. 2007	otvorená
5.10	Vytvorenie architektúry použitia Clipsu a Lucene v projekte.	Zdenko Porubčan, Michal Pažitný	7. 11. 2007	21. 11. 2007	otvorená