

Zápisnica zo stretnutia č. 10

Dátum:
11.3.2010

Miesto:
IBM laboratórium

Čas:
13:00 – 14:00

Zúčastnení členovia:
Bc. Marián Beňo
Bc. Miloš Blaško
Bc. Ľubomír El'ko
Bc. Ján Kmeťko
Bc. Lukáš Lazarčík
Bc. Tomáš Mičko

Vedúci pedagóg: Ing. Ivan Polášek, PhD.

Stretnutie viedol: Bc. Ľubomír El'ko

Zapisovateľ: Bc. Miloš Blaško

Program stretnutia:

1. Úvodná diskusia
 - diskusia k priebehu riešenia zadefinovaných úloh
2. Rozobratie aktuálnych problémov
 - zadané/nezadané metadáta dokumentu
 - zadávanie údajov pri pridávaní dokumentov
 - hromadné pridávanie dokumentov
3. Zadefinovanie nových úloh
 - pridanie mouse listenerov k uzlom a hranám grafu
 - zabezpečenie nezobrazovania väzieb zrušených používateľom
 - prepracovať zobrazenie výsledkov hľadania v dokumentoch
 - pridanie filtra na zobrazovanie iba zvolených väzieb
 - upratanie build systému
 - analýza možností práce s metadátami uložených dokumentov
 - spracovanie sťahovania súborov zo servera ku klientovi
 - získanie/vytvorenie väčšej testovacej bázy znalostí

Priebeh stretnutia

1. *Úvodná diskusia*

V tomto týždni došlo k nedorozumeniu a tímu sa nepodarilo stretnúť s vedúcim. Čas strávený v IBM laboratóriu sme ale využili na diskusiu o stave riešenia zadaných úloh.

Do tohto stretnutia sa nám podarilo spraviť nasledujúce:

Miloš Blaško

- rozbehanie bugzilly

- inštalácia javy a nastavenie práv webservera na labss2

Lubomír Elko

- migrácia z nástroja Lucene na Compass

- úprava závislých zdrojových kódov

Tomáš Mičko

- editácia dokumentov

- skúšobné sťahovanie dokumentov, upload novej verzie dokumentu

- vytváranie používateľských väzieb v grafe

Ján Kmeťko

- presun atribútov z triedy File do triedy Document

- analýza pridania autentifikácie používateľov

- analýza nastavenia applet policy

Lukáš Lazarčík

- TF-IDF vektor

- návrh algoritmu na zistenie podobnosti dokumentov na základe ich obsahu

- analýza named entity recognition algoritmov a knižníc na vyhľadávanie osôb a názvov dokumentov (referencií) v neštruktúrovanom texte

Marián Beňo

- aktualizácia plánu a zistenie termínov a formy odovzdávania výstupov

2. *Rozobratie aktuálnych problémov*

Uvedomili sme si, že sme ešte stále nevyriešili problém získavania autorov a názvov ukladaných dokumentov. Tieto údaje sa dajú získať z metadát, ale nemáme istotu, že budú zadané alebo či budú správne. Takisto je možné tieto údaje zadať v dialógovom okne pri pridávaní nového dokumentu do bázy znalostí, čo zas môže byť pracné. Najjednoduchšie riešenie by bolo povinné uvádzanie autora aj názvu dokumentu pri jeho pridaní do bázy znalostí. Doba trvania pridávania nových dokumentov by sa ale predĺžila a je možné, že by tieto údaje pri niektorých dokumentoch ani nebolo možné zo strany používateľa zistiť.

Ďalšiu vec, ktorú sme ešte nevyriešili je možnosť editovania názvu a autora už pridaného dokumentu. V takomto prípade by nám pomohlo, ak by sme mohli programovo meniť metadáta dokumentov, inak by mohlo dôjsť k nekonzistentnosti medzi reálnymi údajmi a údajmi v našom systéme. Momentálne je to vyriešené tak, že používateľ si dokument stiahne k sebe, upraví metadáta a na serveri prepíše starý dokument novým.

Poslednou diskutovanou vecou bolo pridávanie väčšieho množstva dokumentov naraz bez ručného zadávania údajov. Takéto hromadné pridávanie má opodstatnenie najmä

pri zavádzaní nášho nástroja na už existujúcu bázu znalostí (niekoľko sto dokumentov). Problém je rovnaký ako u vyššie spomenutých záležitostiach – neuvedené/nesprávne metadáta.

3. Zadefinovanie nových úloh

Nové úlohy, ktoré vyplynuli zo stretnutia:

- pridanie mouse listenerov k uzlom a hranám grafu – napr. dvojklik na zobrazenie detailu, pravý klik na zobrazenie pop-up menu s ďalšími možnosťami (vymazať, stiahnuť...) a pod.
- zabezpečenie nezobrazovania väzieb zrušených používateľom – ide len o malú úpravu v kóde - bug
- prepracovať zobrazenie výsledkov hľadania v dokumentoch – krajšie a prehľadnejšie zobrazenie výsledkov hľadania s možnosťou výberu centrálného dokumentu
- pridanie filtra na zobrazovanie iba zvolených väzieb
- upratanie build systému – treba pridať závislosti na Compass a nové knižnice
- analýza možnosti práce s metadátami uložených dokumentov – extrakcia je už otestovaná, hodila by sa nám aj možnosť zapisovania/zmeny metadát dokumentov
- spracovanie sťahovania súborov zo servera ku klientovi – sťahovanie dokumentov ku klientovi funguje, je potrebné pridať indikátor stavu sťahovania (progress bar) plus otestovať správanie pri veľkých súboroch
- získanie/vytvorenie väčšej testovacej bázy znalostí – doteraz sme nástroj testovali len s malou množinou dokumentov (okolo 10), takže by bolo vhodné získať alebo nejako vytvoriť početnejšiu množinu spracovávaných dokumentov

4. Stav úloh z prechádzajúceho stretnutia

Číslo úlohy	Úloha	Člen tímu	Výstup	Termín zadania	Termín ukončenia	Stav
9.1	Rozbehanie Bugzilly	Miloš Blaško	Bugzilla	25.02.10	11.03.10	Splnené
9.2	Kontaktovať Lacka ohľadom Yetty	Miloš Blaško	Yetty	25.02.10	11.03.10	Splnené
9.3	Dopracovanie front-end časti (filtre, nastavenia a pod.)	Miloš Blaško, Tomáš Mičko	Zdrojový kód	25.02.10	11.03.10	Pokračuje
9.4	Doplnenie GUI o vytváranie mauálnych väzieb	Miloš Blaško, Tomáš Mičko	Zdrojový kód	25.02.10	11.03.10	Splnené
9.5	Migrácia na Compass	Ľubomír Eľko, Ján Kmeťko	Zdrojový kód	25.02.10	11.03.10	Splnené
9.6	Indexovanie PDF a DOC	Ľubomír Eľko	Zdrojový kód	25.02.10	11.03.10	Pokračuje
9.7	Získavanie metadát z dokumentov	Ľubomír Eľko	Zdrojový kód	25.02.10	11.03.10	Splnené
9.8	Administrátorské rozhranie	Ján Kmeťko	Zdrojový kód	25.02.10	11.03.10	Pokračuje
9.9	Zavedenie používateľov	Ján Kmeťko	Zdrojový kód	25.02.10	11.03.10	Pokračuje
9.10	Applet Policy – upload súborov	Ján Kmeťko	Zdrojový kód	25.02.10	11.03.10	Pokračuje
9.11	Support service vrstvy, DB, back-up	Ján Kmeťko	-	25.02.10	13.05.10	Pokračuje
9.12	Presun properties z File do Document	Ján Kmeťko	Zdrojový kód	25.02.10	11.03.10	Splnené
9.13	Implementácia algoritmov pre vytváranie väzieb	Lukáš Lazarčík, Marián Beňo	Zdrojový kód	25.02.10	11.03.10	Pokračuje
9.14	Aktualizácia dokumentácie	Lukáš Lazarčík	Dokument	25.02.10	13.05.10	Pokračuje
9.15	Aktualizácia a support build systému	Lukáš Lazarčík	-	25.02.10	13.05.10	Pokračuje
9.16	Aktualizácia plánu a zistenie termínov a formy odovzdávania výstupov	Marián Beňo	Dokument	25.02.10	11.03.10	Splnené

5. Zadelenie nových úloh na vypracovanie

Číslo úlohy	Úloha	Člen tímu	Výstup	Termín zadania	Termín ukončenia	Stav
9.3	Dopracovanie front-end časti (filtre, nastavenia a pod.)	Miloš Blaško, Tomáš Mičko	Zdrojový kód	25.02.10	18.03.10	-
9.6	Indexovanie PDF a DOC	Ľubomír El'ko	Zdrojový kód	25.02.10	18.03.10	-
9.8	Administrátorské rozhranie	Ján Kmeťko	Zdrojový kód	25.02.10	18.03.10	-
9.9	Zavedenie používateľov	Ján Kmeťko	Zdrojový kód	25.02.10	18.03.10	-
9.10	Applet Policy – upload súborov	Ján Kmeťko	Zdrojový kód	25.02.10	18.03.10	-
9.11	Support svc, DB, back-up	Ján Kmeťko	-	25.02.10	13.05.10	-
9.13	Implementácia algoritmov pre vytváranie väzieb	Lukáš Lazarčík, Marián Beňo	Zdrojový kód	25.02.10	18.03.10	-
9.14	Aktualizácia dokumentácie	Lukáš Lazarčík	Dokument	25.02.10	13.05.10	-
9.15	Aktualizácia a support build systému	Lukáš Lazarčík	-	25.02.10	13.05.10	-
10.1	Pridanie mouse listenerov do grafu	Tomáš Mičko	Zdrojový kód	11.03.10	18.03.10	-
10.2	Nezobrazovanie zrušených väzieb	Ján Kmeťko	Zdrojový kód	11.03.10	18.03.10	-
10.3	Prepracovať výsledky hľadania	Miloš Blaško	Zdrojový kód	11.03.10	18.03.10	-
10.4	Filter na zobrazenie väzieb	Tomáš Mičko, Miloš Blaško	Zdrojový kód	11.03.10	18.03.10	-
10.5	Upratanie build systému	Ján Kmeťko, Lukáč Lazarčík	Zdrojový kód	11.03.10	18.03.10	-
10.6	Analýza práce s metadátami	Ľubomír El'ko	Dokument	11.03.10	18.03.10	-
10.7	Spracovanie sťahovania súborov	Tomáš Mičko, Miloš Blaško	Zdrojový kód	11.03.10	18.03.10	-
10.8	Väčšia testovacia база	Všetci	-	11.03.10	18.03.10	-