



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



Používateľská príručka Tímový projekt

Grafická podpora vyhľadávania znalostí v dokumentoch

Tím 12: Šproty

*Bc. Marian Beňo, Bc. Miloš Blaško, Bc. Lubomír Elko, Bc. Ján Kmeťko, Bc. Lukáš
Lazarčík, Bc. Tomáš Mičko*

Vedúci: Ing. Ivan Polášek PhD.

Kontakt:

sproty@googlegroups.com

1. Používateľská príučka

V tejto časti dokumentácie opíšeme návod na používanie klientskej časti aplikácie, ktorá je realizovaná v podobe appletu.

1.1 Systémové požiadavky

- Nainštalovane JRE (Java Runtime Environment) – minimálne vo verzii 1.6
- Pripojenie k internetu a internetový prehliadač (Internet Explorer 7 a vyšší, Mozilla Firefox 3.0)

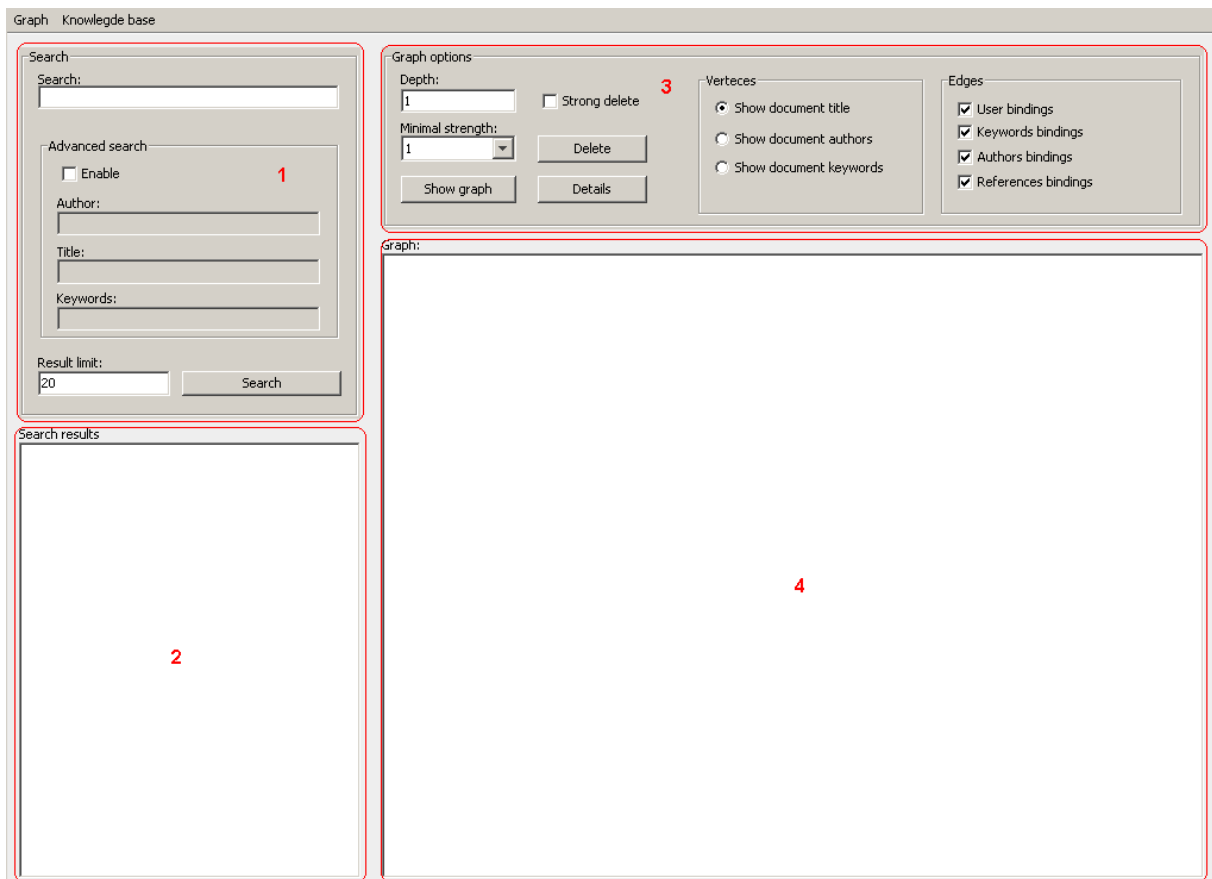
1.2 Spustenie aplikácie

Pre spustenie aplikácie je potrebné otvoriť internetový prehliadač (IE alebo Firefox) a zadať adresu <http://labss2.fiiit.stuba.sk/TeamProject/2009/team12is-si/knowledgeBase/>

1.3 Opis okna appletu

Po načítaní stránky sa zobrazí základné okno appletu (obr. 1), ktoré sa dá logicky rozdeliť na niekoľko častí:

1. Táto časť slúži na vyhľadávanie v báze znalostí.
2. V tejto časti sa zobrazujú výsledky vyhľadávania
3. V tejto časti je možné nastavovať rôzne možnosti v grafe.
4. Táto časť appletu slúži na zobrazovanie samotného grafu dokumentov z bázy znalostí.



Obr. 1. Základné okno appletu

Okrem týchto častí sa ešte v hornej časti appletu nachádza menu.

1.4 Vyhľadávanie v báze znalostí

Na vyhľadávanie v báze znalostí slúži prvá časť okna appletu. Je možné vykonať dva typy vyhľadávania:

- Základné – spočíva vo vyplnení položky „*Search*“ a kliknutí na tlačidlo „*Search*“.
- Pokročilé (Advanced) – povoľuje sa zaškrtnutím políčka „*Enable*“, kedy sa odblokujú položky „*Author*“, „*Title*“ a „*Keywords*“. Vyhľadávanie sa opäť spustí kliknutím na tlačidlo „*Search*“.

Rozdiel medzi týmito dvoma vyhľadávaniami je jednoduchý. Pri základnom vyhľadávaní sa zadaná fráza hľadá v celom obsahu súborov, ktoré sú v báze znalostí. Pri pokročilom vyhľadávaní sa vyhľadá iba v metadátoch jednotlivých dokumentov. Je pritom možné definovať, ktoré metadáta sa majú prehľadávať (autori, názvy a/alebo kľúčové slová) podľa toho, ktoré položky pri vyhľadávaní sa vyplnili.

V tejto prvej časti appletu je ešte jedna položka – „*Result limit*“, ktorá určuje maximálny počet zobrazených výsledkov vyhľadávania.

Po vyplnení požadovaných položiek a stlačení tlačidla „*Search*“, alebo stlačení tlačidla „*Enter*“ na klávesnici, sa spustí vyhľadávanie. Výsledky sa zobrazia v druhej časti appletu (obr. 2). Výsledky sú zoradené podľa relevantnosti, pričom viac relevantné dokumenty sú navrchu.



Obr. 2. Zobrazenie výsledkov vyhľadávania

Výsledky sú vďaka farebnosti a rôznej veľkosti písmen prehľadné. Modrou farbou je navrchu každého výsledku zobrazený názov dokumentu. Pod ním je zelenou farbou uvedený autor

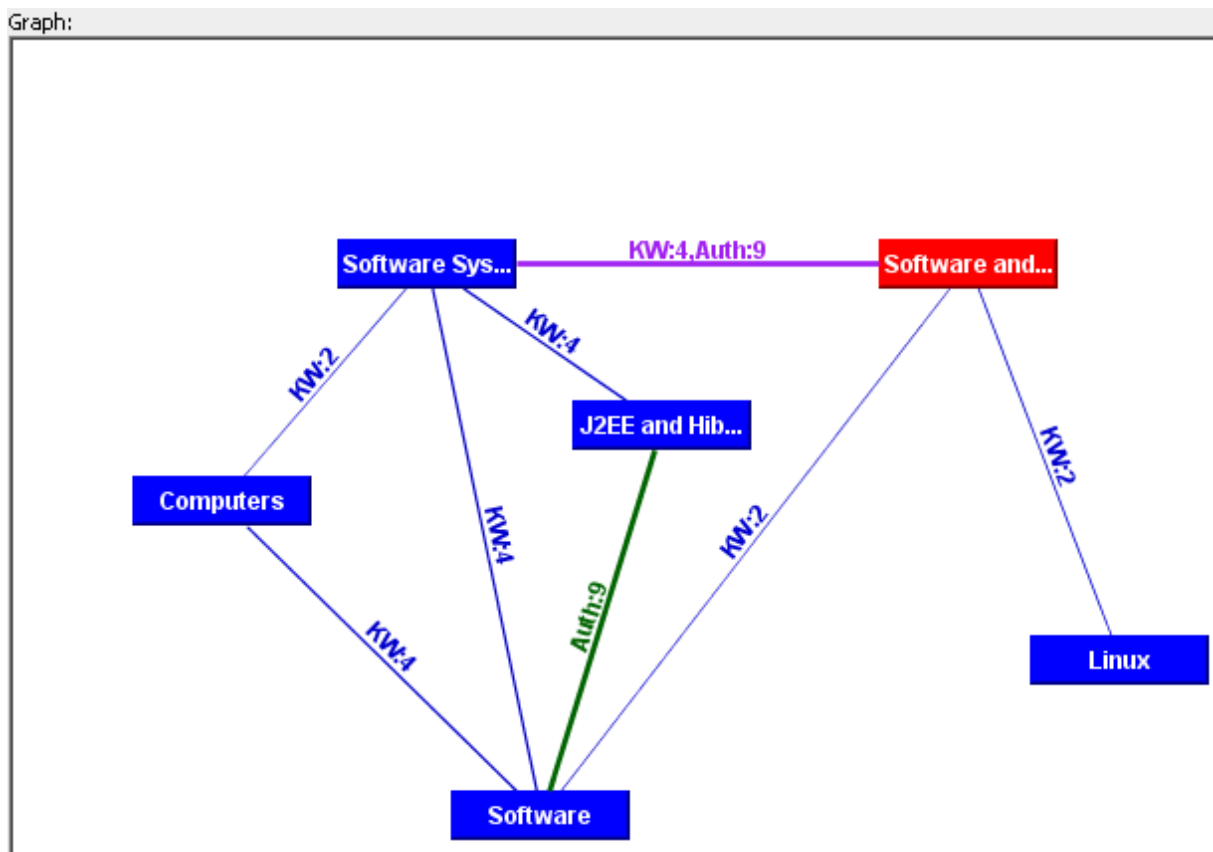
dokumentu. Nakoniec sú čiernou farbou vypísané kľúčové slová daného dokumentu, ktoré sú oddelené čiarkou.

1.5 Zobrazenie grafu dokumentov

Po tom, čo sa zobrazili výsledky vyhľadávania je možné vytvoriť okolo vyhľadaného dokumentu graf reprezentujúci grafické znázornenie bázy znalostí vrátane prepojení medzi jednotlivými dokumentmi. Zobraziť graf je možné dvoma spôsobmi:

1. Dvojklikom na jeden z dokumentov zobrazených vo výsledkoch vyhľadávania.
2. Označením dokumentu zobrazeného vo výsledkoch vyhľadávania (kliknutím na neho) a následným kliknutím na tlačidlo „Show graph“.

Následne sa vo štvrtej časti okna appletu zobrazí graf reprezentujúci časť bázy znalostí (obr. 3), ktorá prislúcha k okoliu dokumentu, ktorý sa zobrazil ako centrálny – je to dokument v grafe označený červenou farbou, dokument, okolo ktorého sa zobrazil graf.



Obr. 3. Graf reprezentujúci bázu znalostí

1.5.1 Opis zobrazeného grafu

Graf je zložený z vrcholov reprezentujúcich dokumenty a z hrán, ktoré reprezentujú prepojenia medzi dokumentmi. V grafe sa môže nachádzať viacero typov vrcholov a hrán, preto ich význam bližšie opíšeme.

V grafe sa môžu nachádzať tri typy vrcholov. Sú oddelené iba farbou zobrazenia:

- Červenou farbou je znázornený centrálny dokument. Je to ten, ktorý sa označil vo výsledkoch vyhľadávania a zobrazili sme graf okolo neho.

- Zelenou farbou je zobrazený dokument, ktorý sme rozbalili (expand) – v grafe na obr. 3 sa takýto vrchol nenachádza, preto bude mechanizmus rozbalenia opísaný neskôr.
- Modrou farbou sú zobrazené všetky ostatné dokumenty v grafe.

Každý dokument je v grafe reprezentovaný jedným vrcholom. Názov vrcholu reprezentuje názov reprezentovaného dokumentu. To, aký nápis sa zobrazuje vo vrchole sa dá meniť nastaveniami, ktoré opíšeme neskôr. Všetky vrcholy v grafe majú rovnakú veľkosť, preto sa môže stať, že vo vrchole sa nebude zobrazovať celý nápis, ale iba jeho časť. Zobrazenie stručných informácií o dokumente v grafe sa dá vyvolať krátkym podržaním kurzora myši nad vybraným vrcholom. Po krátkom čase sa zobrazí popisok (tooltip) s informáciami o dokumente.

Medzi dokumentmi v grafe sa zobrazujú hrany, ktoré reprezentujú prepojenia medzi jednotlivými dokumentmi. Tieto prepojenia sú buď definované používateľom aplikácie, alebo sú vytvorené automaticky bázou znalostí na serveri pomocou špeciálnych heuristických algoritmov na základe obsahu súborov, alebo na základe metadát dokumentov. Podľa typu prepojenia, ktoré reprezentujú, majú hrany rôznu farbu a popis:

- Prepojenie dokumentov na základe rovnakých kľúčových slov je reprezentované hranou modrej farby. Popis hrany sa skladá z dvoch častí. Prvú časť tvoria písmená KW (z angl. KeyWords), a druhú časť tvorí číslo, ktoré je v rozmedzí 1 až 9. Toto číslo vyjadruje silu prepojenia – prepojenie o sile 1 je najslabšie.
- Prepojenie dokumentov na základe rovnakých autorov je reprezentované hranou zelenej farby a popisom, ktorý má v prvej časti písmená Auth (z angl. Author).
- Prepojenie vytvorené na základe referencie – jeden dokument sa odkazuje na iný, je tvorené hranou červenej farby a popisom, ktorý má v prvej časti písmená Ref (z angl. Reference).
- Graf môže obsahovať aj hranu čiernej farby, ktorá reprezentuje prepojenie definované používateľom. Takáto hrana má v popise frázu „User“.

V grafe sa môžu nachádzať aj hrany fialovej farby, ktoré majú trochu zložitejší popis. Tieto hrany reprezentujú zložené prepojenie – tvorené viacerými jednoduchými prepojeniami. V popise takýchto hrán sa nachádza informácia, aké typy prepojení sú na daných hranách. Napríklad v grafe na obrázku 3 je fialová hrana s popisom „KW:4,Auth:9“, čo znamená, že hrana reprezentuje dve prepojenia. Jedno vytvorené na základe rovnakých kľúčových slov o sile 4 a druhé vytvorené na základe rovnakých autorov dokumentov o sile 9.

Sila prepojenia je v grafe znázornená taktiež hrúbkou hrany. Na obrázku 3 je vidieť, že prepojenia o sile 2 majú užšiu hrúbku hrany ako prepojenia o sile 9.

1.6 Manipulácia grafu

Applet umožňuje manipuláciu zobrazeného grafu. Ľubovoľný vrchol je možné v grafe presúvať – meniť jeho pozíciu. Je možné označiť viacero vrcholov a presúvať ich hromadne. Na manipuláciu grafu slúži časť tri v okne appletu. Postupne opíšeme význam jednotlivých položiek.

1.6.1 Položka „Depth“

Nastavením čísla v tejto položke sa určí do akej maximálnej hĺbky sa má graf zobrazit'. Hĺbka v našom grafe znamená vzdialenosť od centrálného vrcholu. V grafe na obrázku 3 je vrchol s názvom „Linux“ v hĺbke jeden od centrálného vrcholu, zatiaľ čo vrchol s názvom „Computers“ je v hĺbke dva. Tento graf bol zobrazený s nastavením položky „Depth“ rovnou dvom. Graf už do väčšej hĺbky nešiel.

1.6.2 Položka „Minimal strength“

Nastavením tejto položky sa určí minimálna sila hrán, ktoré sa v grafe majú zobrazit'. Ak sa táto položka nastaví napr. na hodnotu 3, tak v grafe sa budú zobrazovať iba prepojenia, ktoré majú silu tri a väčšiu.

1.6.3 Tlačidlo „Show graph“

Stlačením tohto tlačidla sa vytvorí graf, ktorý bude mať centrálny dokument taký, ktorý je v danej chvíli označený vo výsledkoch vyhľadávania. Graf sa vytvorí podľa pravidiel (nastavení), ktoré sú aktuálne nastavené v časti tri okna appletu.

1.6.4 Tlačidlo „Delete“ a položka „Strong delete“

Slúži na vymazanie vrcholu (dokumentu) alebo hrany (prepojenia medzi dokumentmi) z grafu. Vymaže sa práve označený vrchol alebo hrana. Vrcholy a hrany sa v grafe označujú tak, že sa na ne klikne ľavým tlačidlom myši. Po stlačení tlačidla sa aplikácia používateľa opýta, či chce skutočne vymazať danú položku z grafu. Ak sa označí fialová hrana (reprezentujúca viacero prepojení), aplikácia si vyžiada označiť zoznam prepojení, ktoré si používateľ žiada odstrániť. Ak sa vymazáva vrchol, zmažú sa aj všetky hrany, ktoré sú napojené na tento vrchol. Tlačidlo „Delete“ má klávesovú skratku – klávesu „Delete“.

S tlačidlom „Delete“ je spojená aj položka „Strong delete“. Ak je táto položka zaškrtnutá, znamená to, že pri mazaní sa daný dokument, alebo prepojenie zmaže aj z bázy znalostí (teda zo servera) a pri ďalšom zobrazení grafu sa už dokument alebo prepojenie nezobrazí. Ak táto položka nie je zaškrtnutá, mazanie sa uskutoční iba lokálne, tzn. dokument alebo prepojenie sa zmaže iba z grafu a pri opätovnom zobrazení grafu bude už opäť viditeľný(é).

UPOZORNENIE!!!: Je dôležité zaškrtnávať položku „Strong delete“ iba v prípade ak skutočne používateľ chce zmazať dokument alebo prepojenie aj z bázy znalostí a nechce aby sa táto položka zobrazovala ďalším používateľom aplikácie.

1.6.5 Tlačidlo „Details“

Ak je označený vrchol, alebo hrana v grafe, po stlačení tohto tlačidla sa zobrazia detaily danej položky (obr. 4 zobrazuje detaily dokumentu reprezentovaného vrcholom v grafe, obr. 5 zobrazuje detaily prepojení reprezentovaných hranou v grafe).

Medzi zobrazovanými informáciami o dokumente sú v zobrazenom poradí:

- ID dokumentu v báze znalostí, používateľ ho nemôže meniť
- Názov dokumentu

- Autor dokumentu
- Kľúčové slova dokumentu
- Typ dokumentu
- Dátum pridania do bázy znalostí
- Používateľ, ktorý daný dokument do bázy vložil (používatelia nie sú vo verzii 1.0 implementovaný)
- Názov súboru dokumentu
- Veľkosť súboru

The screenshot shows a window titled "Document details" with the following fields and values:

Document ID:	11
Title:	Linux
Author:	Lubomir Elko
Keywords:	linux, programming, digital,
Document type:	Technical documentation
Add date:	2010-04-13 15:04:35.921
Added by:	1
File name:	3.doc
File size:	32 KB

Buttons at the bottom: Upload new file, Download, Save changes, Edit, Close.

Obr. 4. Detailné informácie o dokumente.

V tomto okne sa nachádzajú aj tri zaujímavé tlačidlá – „Upload new file“, „Edit“ a „Download“. Prvé tlačidlo slúži na nahratie novej verzie súboru pre daný dokument. Po stlačení tlačidla sa vyvolá okno, v ktorom je potrebné vybrať súbor na nahratie.

Tlačidlo „Edit“ slúži na zmenu atribútov dokumentu. Po jeho stlačení sa povolia položky, ktoré je možné meniť. Po ich zmene je potrebné kliknúť na tlačidlo „Save changes“, čo spôsobí uloženie zmien na server.

Stlačením tlačidla „Download“ sa vyvolá uloženie súboru zo servera na lokálny počítač.

Medzi zobrazovanými informáciami o prepojení sú v zobrazenom poradí:

- Prepojenie, ktoré je zobrazované (slúži na rozlíšenie viacerých prepojení, ktoré sú na fialovej hrane)
- ID prepojenia v báze znalostí, používateľ ho nemôže meniť
- Typ prepojenia

- Sila prepojenia
- Dátum pridania do bázy znalostí
- Dokumenty, ktoré prepojenie spája (rozlišuje sa medzi odkazovaným a odkazujúcim dokumentom)
- Označenie, či je prepojenie obojsmerné
- Označenie, či je prepojenie validné – prepojenie nie je validné, ak sa vymaže z bázy znalostí
- Meno používateľa, ktorý prepojenie vymazal (nie je implementované vo verzii 1.0)
- Dátum, kedy bolo prepojenie vymazané (nie je implementované vo verzii 1.0)
- Meno používateľa, ktorý väzbu vytvoril (týka sa iba používateľských prepojení (user bindings))
- Meno algoritmu, ktorý prepojenie vytvoril
- Verzia algoritmu, ktorý prepojenie vytvoril
- Prídavné textuálne informácie (komentáre) o prepojení

Binding:	Keywords: 4
Binding ID:	20
Binding type:	Similar content
Strength:	4
Date added:	13.4.2010 15:04
Referencing document:	Software Systems
Referenced document:	Software and HW
Bidirectional:	True
Valid:	True
Invalidated by:	N/A
Invalidated date:	N/A
Created by:	
Algorithm name:	ik.fiit.sproty.knowledgebase.algorithms.KeywordAlgorithm
Algorithm version:	algVersion
Comments:	comment

Buttons: Save changes, Edit, Close

Obr. 5. Detailné informácie o prepojení.

Zobrazenie detailov je možné aj vybraním položky „*Show details*“ z kontextového menu, ktoré sa zobrazí po kliknutí s pravým tlačidlom na danú položku v grafu (vrchol alebo hranu). Podobné kontextové menu s rovnakou položkou „*Show details*“ sa zobrazí aj pri kliknutí s pravým tlačidlom myši na dokument, ktorý je zobrazený vo výsledkoch vyhľadávania v časti dva okna appletu.

1.6.6 Položka „*Verteces*“

V tejto položke sa nachádzajú tri „radio buttony“, pomocou ktorých sa dá meniť typ nápisu vo vrchoch grafu. Vo vrchole grafu sa môžu nachádzať tri typy nápisov:

- Názov dokumentu, ktorý vrchol reprezentuje

- Autora dokumentu, ktorý vrchol reprezentuje
- Kľúčové slová dokumentu, ktorý vrchol reprezentuje

1.6.7 Položka „Edges“

V tejto položke sa nachádzajú štyri „checkboxy“, ktorými sa dá upravovať typ prepojení, ktoré sa majú v grafe zobraziť. Zobrazia sa iba prepojenia, ktoré budú zaškrtnuté. Pre aplikovanie zmien týchto nastavení je nutné znova vytvoriť graf.

1.6.8 Kontextové menu grafu

Po stlačení pravého tlačidla myši nad vrcholom grafu sa zobrazí kontextové menu s povolenými položkami „*Show details*“, „*Expand*“ a „*Collapse*“. Položka „*Show details*“ už bola vysvetlená vyššie prejdeme preto na ozrejenie zvyšných dvoch položiek.

Po kliknutí na položku „*Expand*“ nad daným vrcholom sa vybraný vrchol rozbalí, t.j. do grafu sa pridajú ďalšie dokumenty a prepojenia podľa platných nastavení zobrazenia grafu, ktoré sú nastavené v časti tri okna appletu. Rozbalený vrchol sa označí zelenou farbou. Ak sa po stlačení položky „*Expand*“ graf nezmení, znamená to, že nie sú v báze znalostí k dispozícii žiadne ďalšie dokumenty alebo prepojenia, ktoré by vyhovovali nastaveniam grafu.

Položka „*Collapse*“ má opačný efekt ako „*Expand*“. Po kliknutí na ňu zmiznú z grafu všetky vrcholy, ktoré sú spojené s centrálnym vrcholom iba cez daný „*collapsováný*“ vrchol. Vrcholy, ktoré majú k centrálnemu dokumentu prepojenie aj cez iné vrcholy v grafe ostanú.

Po vyvolaní kontextového menu nad prázdny miestom v grafe sa zobrazí menu s povolenou iba jednou položkou – položkou „*Create virtual document*“. Slúži na vytvorenie virtuálneho dokumentu. Je to dokument, ktorý nemá priradený žiadny fyzický súbor. Virtuálny dokument môže slúžiť na vytvorenie grafu s dokumentmi, ktorým sa vytvorí prepojenie na daný virtuálny dokument. Alebo na združenie dokumentov (napríklad všetky dokumenty z daného projektu) - pomocou používateľských väzieb - okolo tohto virtuálneho dokumentu. Po vyvolaní tejto položky kontextového menu sa zobrazí dialógové okno, v ktorom je potrebné vyplniť niekoľko položiek, aby sa vytvoril želaný virtuálny dokument (obr. 6). Ak nechceme, aby systém našiel prepojenie s ostatnými dokumentmi je nutné vyplniť iba položku „*Title*“, ktorá označuje názov dokumentu. V opačnom prípade sa môžu vyplniť aj položky „*Author*“ – autor dokumentu a „*Keywords*“ – kľúčové slová dokumentu. Položka „*Document type*“ označuje typ dokumentu. V položke „*Description*“ je možné uviesť bližší opis dokumentu – dôvod vytvorenia atď.

Create virtual document [X]

Author:

Title:

Keywords:
comma separated keywords e.g. 'kw1, long kw2, kw3' (without quotes)

Document type:

Description:

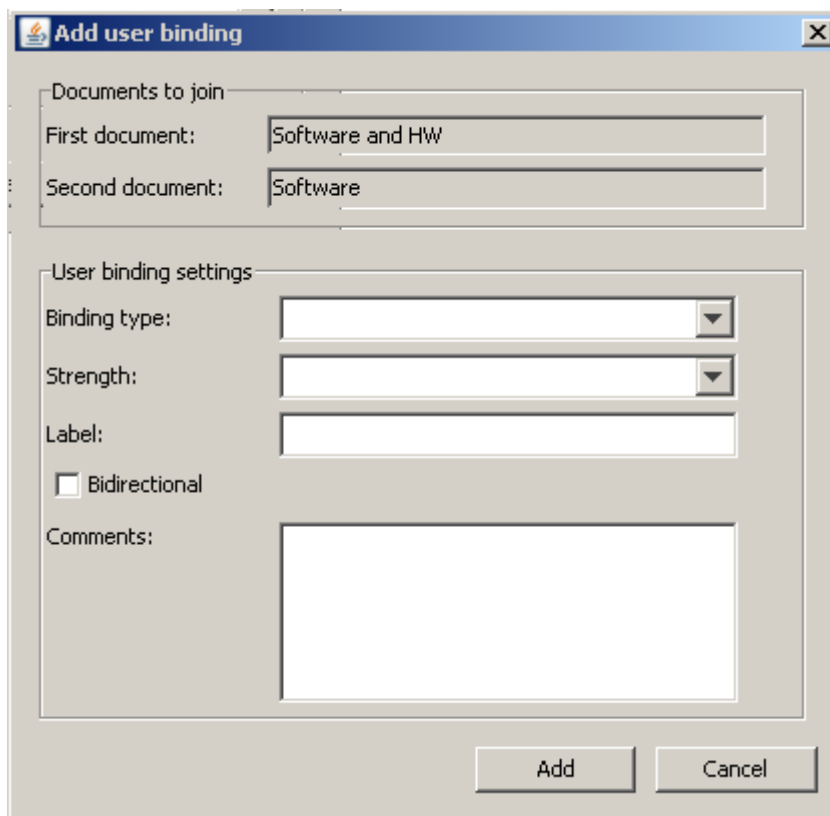
[Create] [Cancel]

Obr. 6. Vytvorenie virtuálneho dokumentu.

1.7 Menu

Menu appletu tvoria dve časti – „*Graph*“ a „*Knowledge base*“. V časti „*Graph*“ sa nachádzajú dve položky – „*Reload layout*“ a „*Clear graph*“. Prvou položkou sa vyvolá preusporiadanie vrcholov grafe – vhodné pri zmenení počtu vrcholov v grafe (napr. pri rozbalení vrcholu). Aplikácia sa snaží vytvoriť čo najlepšie usporiadanie vrcholov v grafe, ktoré spočíva v usporiadaní vrcholov tak, aby sa hrany medzi nimi čo najmenej križovali. Druhou položkou sa vyprázdni aktuálny graf, t.j. zmažú sa všetky vrcholy a hrany v grafe a na obrazovke ostane iba prázdna plocha grafu.

V časti „*Knowledge base*“ sa nachádzajú tri položky – „*Add user binding*“, „*Create virtual document*“ a „*Add document*“. Prvá položka je určená na vytvorenie používateľského prepojenia – prepojenie medzi dokumentmi vytvorené používateľom. Pre správne fungovanie tejto funkcionality je potrebné pred stlačením danej položky označiť v grafe presne dve dokumenty, ináč nebude funkcionality fungovať. Označenie dvoch dokumentov je možné takto: označí sa prvý dokument, podrží sa tlačidlo „ctrl“ na klávesnici a označí sa druhý dokument. Prvý označený dokument sa pri vytváraní prepojenia zoberie ako odkazujúci a druhý dokument ako odkazovaný. Po stlačení položky sa otvorí dialógové okno určené na vyplnenie atribútov prepojenia (obr. 7).



Obr. 7. Vytvorenie používateľského prepojenia.

Položka „*Create virtual document*“ slúži na vytvorenie virtuálneho dokumentu. Je to rovnaká funkcia aká sa vyvolá pri stlačení položky „*Create virtual document*“ v kontextovom menu grafu. Táto funkcia bola vysvetlená v dokumente vyššie.

Položka „*Add document*“ slúži na pridanie súboru (dokumentu) do bázy znalostí na serveri. Po kliknutí na túto položku sa otvorí dialógové okno slúžiace na vyplnenie atribútov vkladaného dokumentu (obr. 8). V tomto okne je potrebné kliknutím na tlačidlo „*Browse*“ vybrať zo súborového systému počítača súbor, ktorý chceme vložiť do bázy znalostí. Ešte je potrebné vyplniť metadáta dokumentu – autora, názov, kľúčové slová a typ dokumentu. Je možné nechať aplikáciu, nech vyberie tieto metadáta z metadát súboru. Túto funkciu je vhodné použiť iba v prípade, keď sú metadáta v súbore vyplnené. Po kliknutí na tlačidlo „*Add document*“ sa dokument vloží do bázy znalostí.

The image shows a dialog box titled "Add document". At the top, there is a "File:" label followed by an empty text input field and a "Browse" button. Below this is a "Metadata" section enclosed in a rounded rectangle. It contains four fields: "Author:" with a text input field and a hint "comma separated authors, e.g. Author1, Author2"; "Title:" with a text input field; "Keywords:" with a text input field and a hint "comma separated keywords, e.g. kw1, long kw2, kw3"; and "Type:" with a dropdown menu. At the bottom of the metadata section is a checkbox labeled "Load metadata from file (metadata must have been filled)". At the very bottom of the dialog box are two buttons: "Add document" and "Cancel".

Obr. 8. Pridanie dokumentu do bázy znalostí.

1.8 Pridaná funkcionálnosť

Do grafu sa dá pridať vrchol aj použitím funkcie „*Drag and Drop*“, a to pretiahnutím dokumentu z výsledkov vyhľadávania do priestoru grafu. V grafe sa pre novo pridaný dokument zobrazia aj prepojenia.

Zmenu centrálného dokumentu je možné vykonať nielen stlačením tlačidla „*Show graph*“, alebo dvojklikom na dokument vo výsledkoch vyhľadávania, ale aj dvojklikom na vybraný dokument (vrchol) v zobrazenom grafe. Táto akcia spôsobí vytvorenie nového grafu s novým centrálnym dokumentom. Graf sa vytvorí podľa pravidiel nastavených v časti tri okna appletu.