

Odhaľovanie vzťahov v obsahu výučbového kurzu

Bc. Marián Šimko

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



Terminológia

- **Objekt výučby (OV)**
 - Výučbový text
- **Koncept**, pseudokoncept
 - Vedomostný element
- **Relácia**
 - Vzťah medzi konceptmi a výučbovými objektmi

Motivácia

Skvalitnenie výučby

Adekvátne prispôsobovanie

Sofistikované mechanizmy

Zložitosť systému

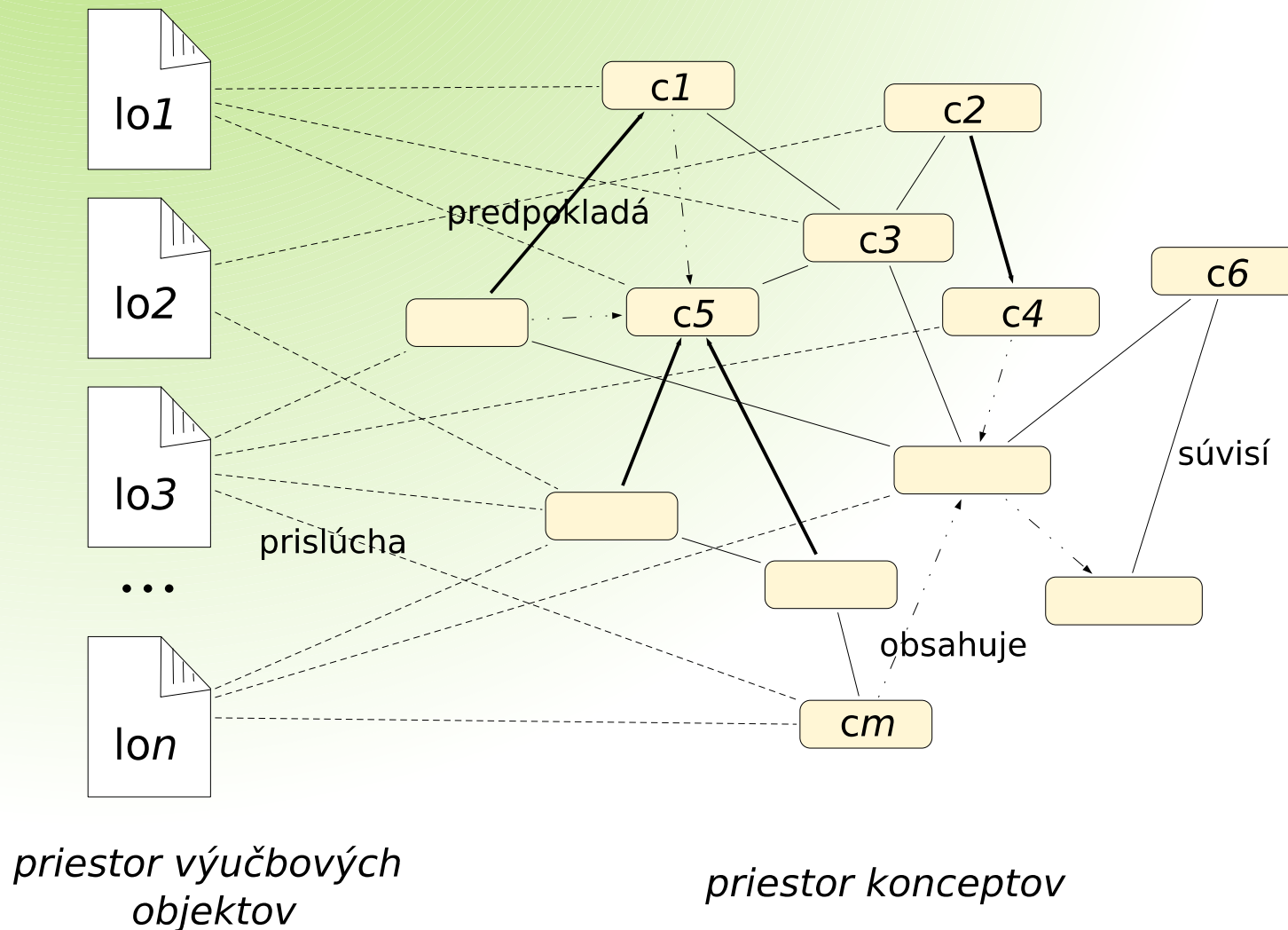
Zložitosť tvorby adaptívneho kurzu

Potreba podpory procesu tvorby

Cieľ

Automatizácia tvorby doménového modelu

Doménový model



Súčasný stav

- Tvorba kurzu – „úzke hrdlo“ adaptívnych systémov pre vzdelávanie
- „Sila inteligentného obsahu je ukrytá v znalostiach o každom jeho fragmente“*
- Absencia riešení využívajúcich potenciál odboru *objavovania znalostí*

* BRUSILOVSKY, P., KNAPP, J., GAMPER, J. Supporting teachers as content authors in intelligent educational systems. In *International Journal of Knowledge and Learning*, 2006, vol. 2, no. 3/4, pp. 191–215.

Moja práca

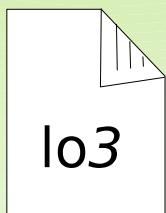
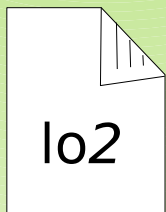
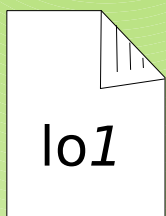


Obsah

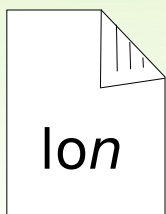
- Metóda automatizovaného získavania metadát o kurze
- Výpočet vzájomnej podobnosti pseudokonceptov
- Overenie riešenia v doméne výučby programovania
- Nástroj CourseDesigner

Metóda automatizovaného získavania metadát o kurze



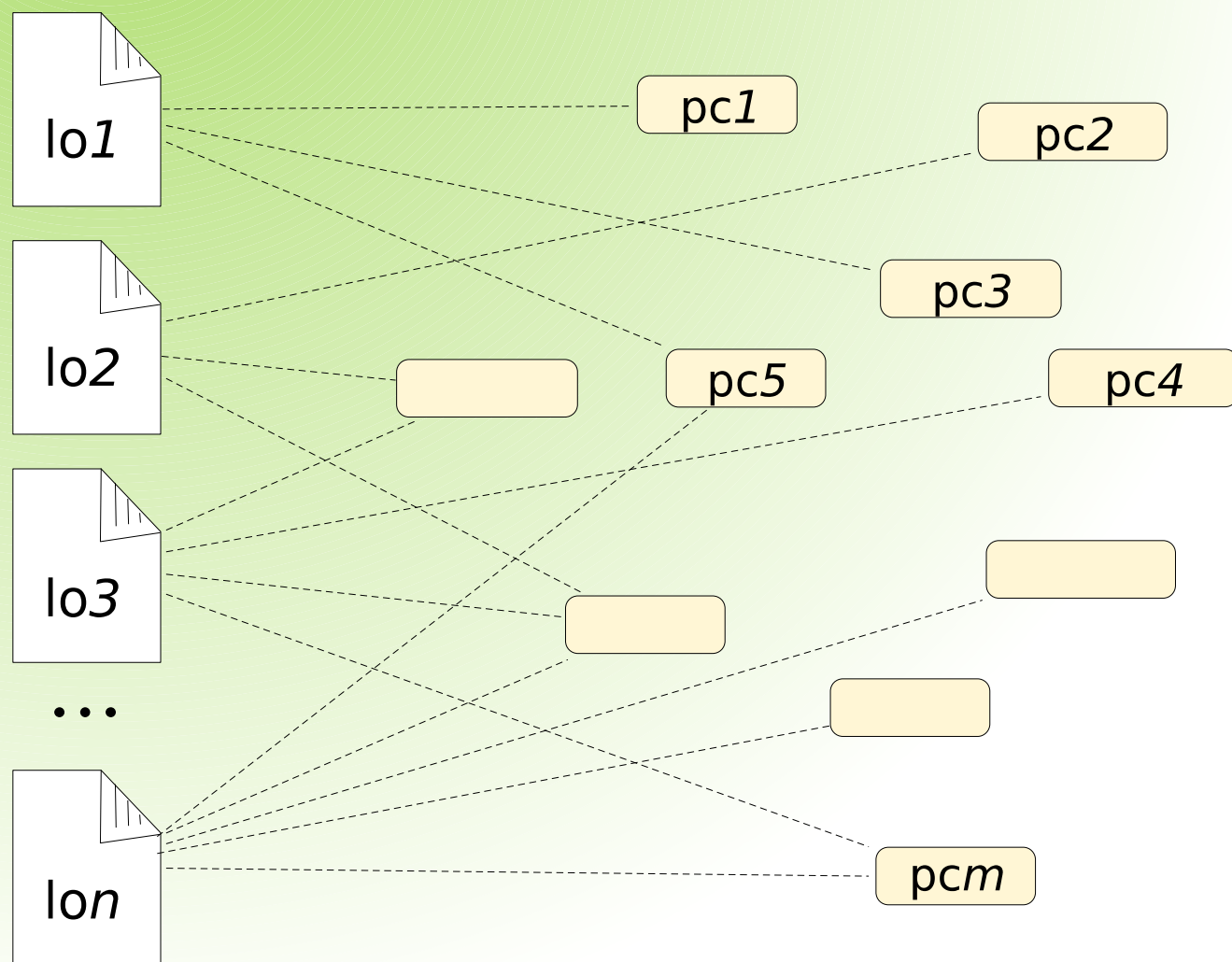


...



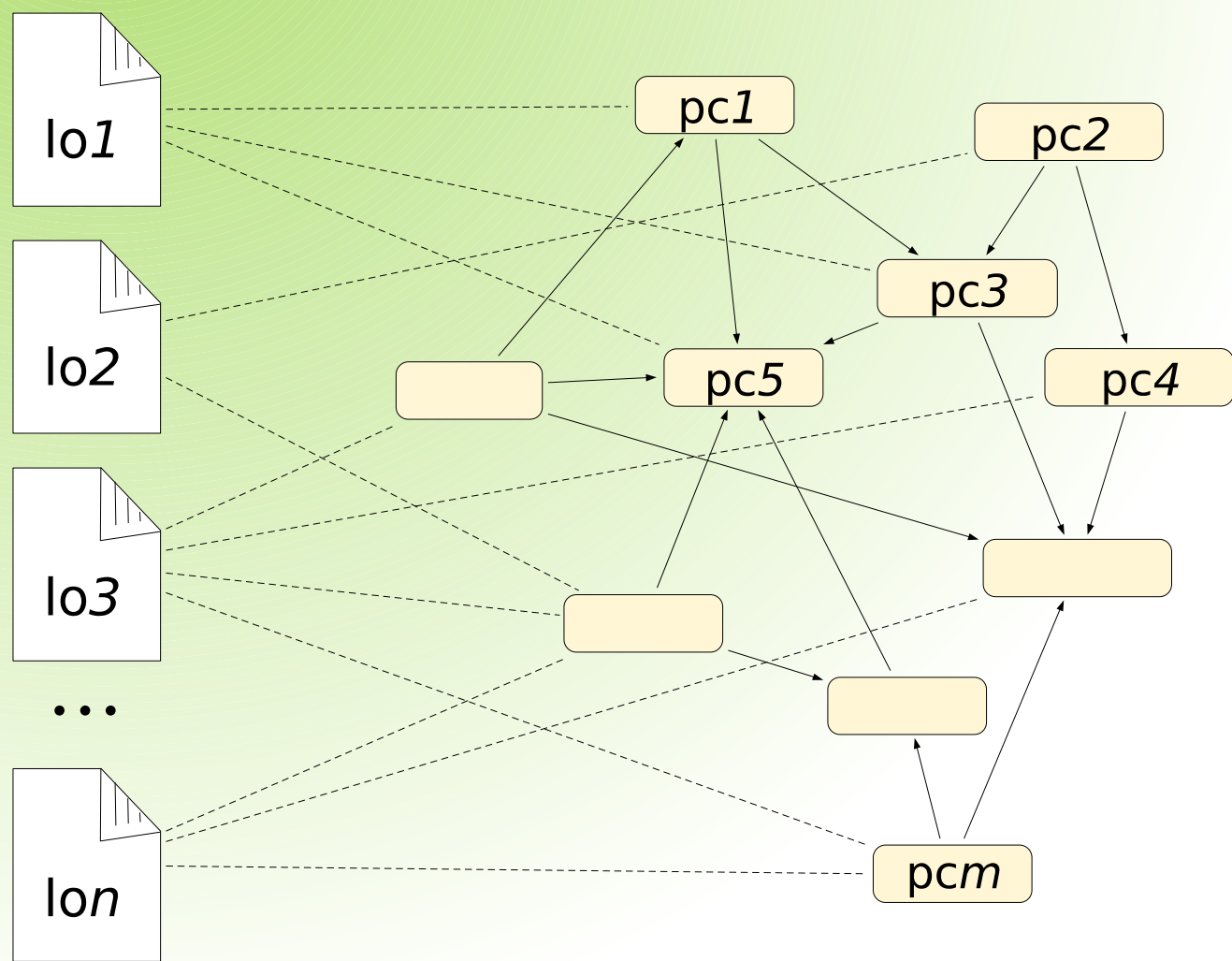
*priestor výučbových
objektov*

priestor konceptov



*priestor výučbových
objektov*

priestor konceptov



*priestor výučbových
objektov*

priestor konceptov

I.
predspracovanie
objektov
výučby

analýza textu

zostavenie vektorovej
reprezentácie

úpravy vektorov

II.
extrakcia pseudo-
konceptov

výber relevantných kľúčových slov

výpočet miery príslušnosti
kľúčových slov k objektom výučby

vytvorenie pseudokonceptov
a ich vzťahov s objektmi výučby

III.
generovanie
vzťahov

výpočet vzájomnej podobnosti
pseudokonceptov

výber najvhodnejších susedov

vytvorenie relácií

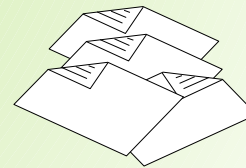
IV.
konštrukcia
doménového
modelu

asistovaná transformácia
pseudokonceptov na koncepty

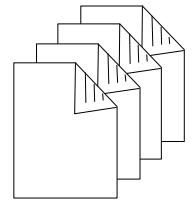
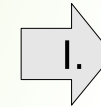
kompozícia doménového
modelu

Predspracovanie OV

- Analýza textu
 - Lexikálna analýza
 - Odstránenie stop slov, lematizácia
- Zostavenie vektorovej reprezentácie
- Úprava vektorov
 - Spracovanie registra pojmov
 - Spracovanie značkovania



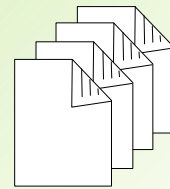
Objekty výučby



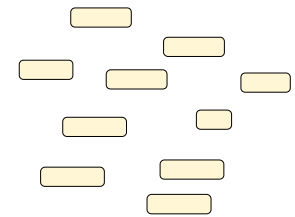
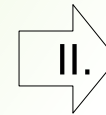
*Predspracované
OV*

Extrakcia pseudokonceptov

- Výber relevantných kľúčových slov



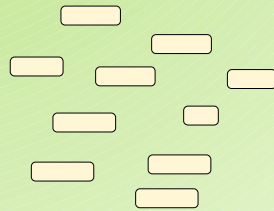
*Predspracované
OV*



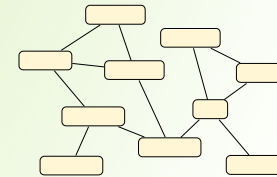
Pseudokoncepty

- Výpočet miery príslušnosti k objektom výučby
- Vytvorenie pseudokonceptov

Generovanie vzťahov



Pseudokoncepty



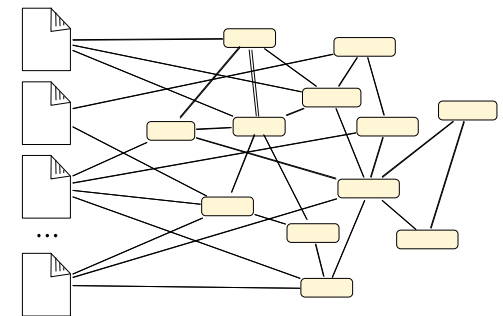
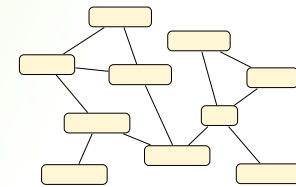
Prepojené pseudokoncepty

- **Výpočet vzájomnej podobnosti pseudokonceptov**
- Výber najvhodnejších susedov
- Vytvorenie relácií

Konštrukcia doménového modelu

- Asistovaná transformácia pseudokonceptov na koncepty
- Kompozícia doménového modelu

Prepojené pseudokoncepty



Doménový model

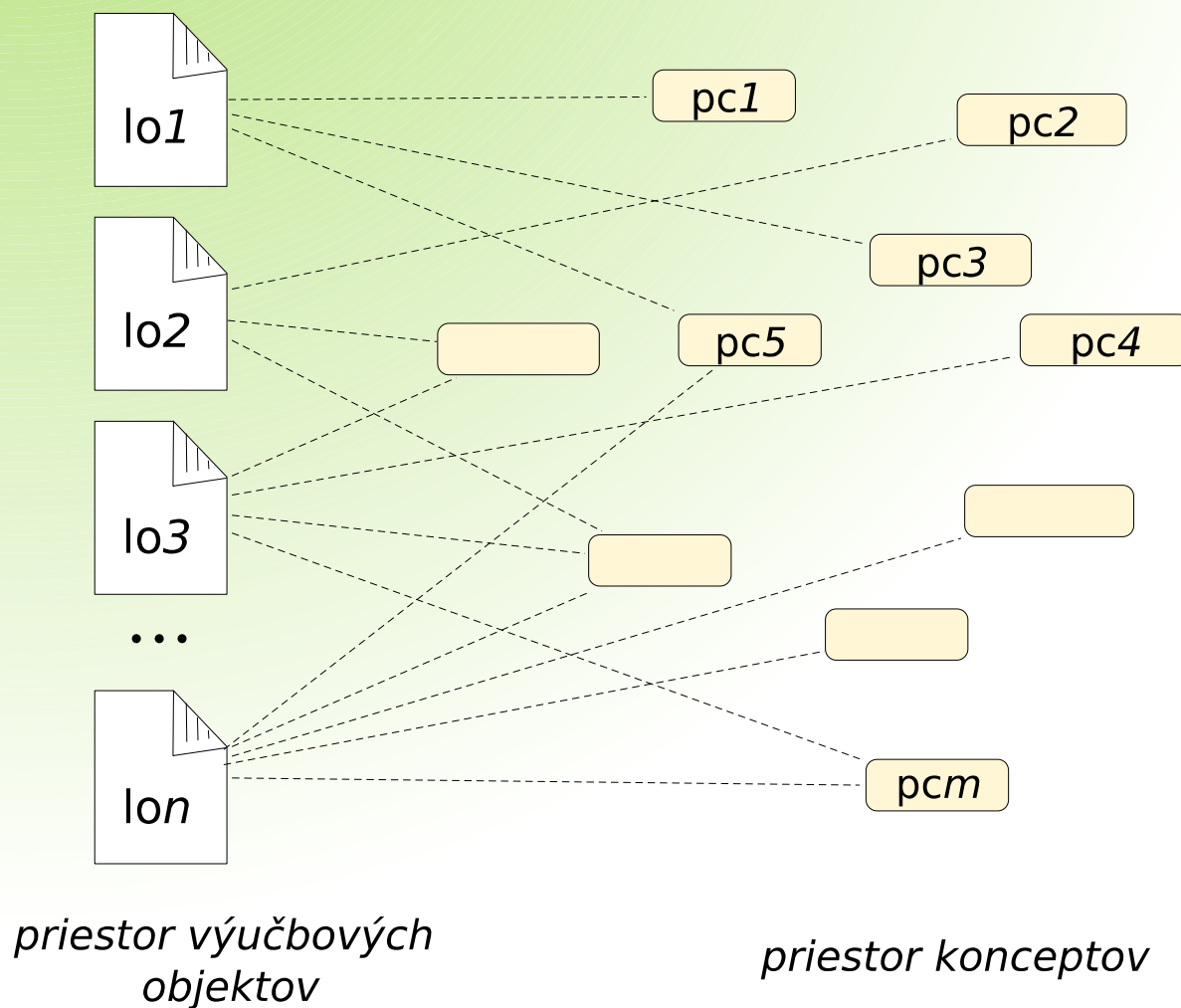
Výpočet vzájomnej podobnosti pseudokonceptov



Varianty

1. Vektorový prístup
 2. Šírenie aktivácie
 3. Analýza založená na algoritme PageRank
- + Kombinácie

Východisková pozícia



1. Vektorový prístup

- Vektorová reprezentácia pseudokonceptu

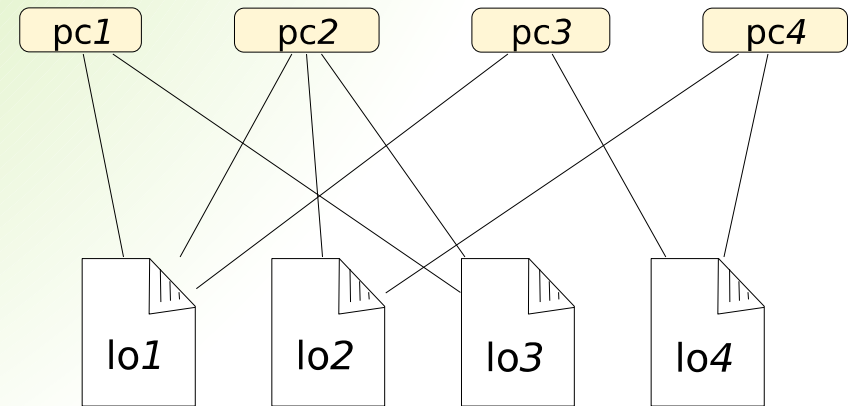
$$\vec{pc}_i = \sum_j w'_{i,j} \vec{lo}_j$$

- Podobnosť pseudokonceptov – kosínusová vzdialenosť

$$coSim(\vec{pc}_x, \vec{pc}_y) = \frac{\sum_i pc_x(i) pc_y(i)}{\sqrt{\sum_i pc_x^2(i)} \sqrt{\sum_i pc_y^2(i)}}$$

2. Šírenie aktivácie

- Neúplný doménový model ako kontextová sieť
- Metóda šírenia aktivácie
- Podobnosť pseudokonceptov:



$$saSim(pc_i, pc_j) = \frac{E_{pc_j}}{\sum_k E_{pc_k}} \log(d_{pc_i, pc_j})$$

3. Analýza PageRank

- Konštrukcia dočasného grafu na základe prepojení pseudokonceptov a objektov výučby
- Ohodnotenie pseudokonceptov *relatívne* vzhľadom na koreňový uzol
- Podobnosť konceptov implicitne ako prestíž uzla

Kombinácie variantov

- Hľadanie lepšej „východiskovej“ pozície pre aplikáciu algoritmu PageRank
- Dočasný graf je výstupom variantu
 1. vektorového prístupu
 2. šírenia aktivácie

Overenie riešenia v doméne výučby programovania



Testovacie dáta

- Kurz funkcionálneho programovania
 - 70 objektov výučby
- Manuálne vytvorený referenčný kurz
 - Jeden typ relácie medzi konceptmi
 - Tri úrovne ohodnotenia relácií:
 - Určite súvisiace
 - Možno súvisiace
 - Určite nesúvisiace

Sledované charakteristiky

- Úplnosť generovania relácií

$$recall = \frac{retrieved \cap relevant}{relevant}$$

$$recall' = \frac{retrieved \cap (relevantA \cup relevantB)}{relevantA \cup (relevantB \cap retrieved)}$$

- Presnosť generovania relácií

$$precision = \frac{retrieved \cap relevant}{retrieved}$$

- F-metrika

Výsledky

Variant	R	P	F	R'	F'
Vektorový prístup	0.594	0.509	0.549	0.793	0.620
Šírenie aktivácie	0.544	0.443	0.488	0.784	0.566
PageRank' analýza	0.500	0.569	0.532	0.741	0.652
Vektorový prístupu + PageRank' analýza	0.567	0.528	0.547	0.803	0.638
Šírenia aktivácie + PageRank' analýza	0.572	0.453	0.506	0.824	0.585

Autorský nástroj CourseDesigner



CourseDesigner

- Integrácia tvorby kurzov do portálu pre podporu vzdelávania (projekt PeWePro)
- Dva ciele:
 - Automatizácia tvorby doménového modelu
 - Experimentovanie s metódou
- Java

Zhodnotenie ...

- Analýza problémovej oblasti → *zložitost'* procesu tvorby adaptívneho kurzu
- Identifikácia potreby využitia metód a techník *objavovania znalostí*
- Návrh a overenie metódy *automatizovaného získavania metadát*
- Implementácia autorského nástroja *CourseDesigner*

... zhodnotenie ...

- Experimenty s viacerými variantmi
- Úspešnosť tvorby doménového modelu
 - 55 – 65 % (pri úspešnosti lematizácie 80,3 %)
 - spätá s reprezentáciou objektov výučby
- Potenciál pre vylepšenie

... ale nie záver

- Ambícia publikovať na medzinárodnej pôde
- Možnosti rozšírenia metódy
 - Úpravy reprezentácie objektov výučby
 - Otestovanie ďalších variantov generovania vzťahov
 - Využitie zhlučovacích algoritmov