

ROZŠÍRENIE MODELU CERIF PRE LEPŠIU ORIENTÁCIU V OBLASTI VEDY A VÝSKUMU

Bc. Metod Majchrák

Ing. Nadežda Andrejčíková PhD.

O čom budeme hovoriť...

- Aktuálne problémy v oblasti vedy a výskumu
- Pozícia začínajúceho výskumníka - use case
- Model CERIF
- Niektoré existujúce riešenia
- Návrh nášho riešenia
- Návrh modelu
- Získavanie a spracovanie dát
- Vyhľadávanie v kontexte – prezentácia dát
- Prínosy

Hlavné problémy súčasného stavu

- **Inštitucionálna, národná aj medzinárodná úroveň**
 - Pracoviská, organizácie, výskumníci, výsledky, granty, konferencie a iné aktivity
 - Izolovanosť a jedno-účelové zameranie informačných zdrojov
 - Heterogenita prostredia, rôzne dátové štruktúry
 - Absencia jednoznačných identifikátorov
 - Minimálne vyjadrenie sémantiky, chýbajúce vzťahy

Pohľad mladého výskumníka

- Každý výskum ja založený na existujúcich znalostiach
- Ako získať prehľad v problémovej oblasti?

Ktoré **publikácie** sa zaoberajú danou témou?

Ktoré **konferencie** sa venovali danej téme?

Aké **projekty** boli riešené na danú tému?

Ktorí **výskumníci** spolupracujú v danej oblasti

V akých **organizáciách** pracujú?

Aké sú ďalšie **projekty** týchto organizácií?



Ktoré **organizácie** usporadúvajú tieto **konferencie**?

Ktorí **výskumníci** sa týchto konferencií zúčastnili?

Ktoré **publikácie** sa stretli s najväčším **ohlasom**?

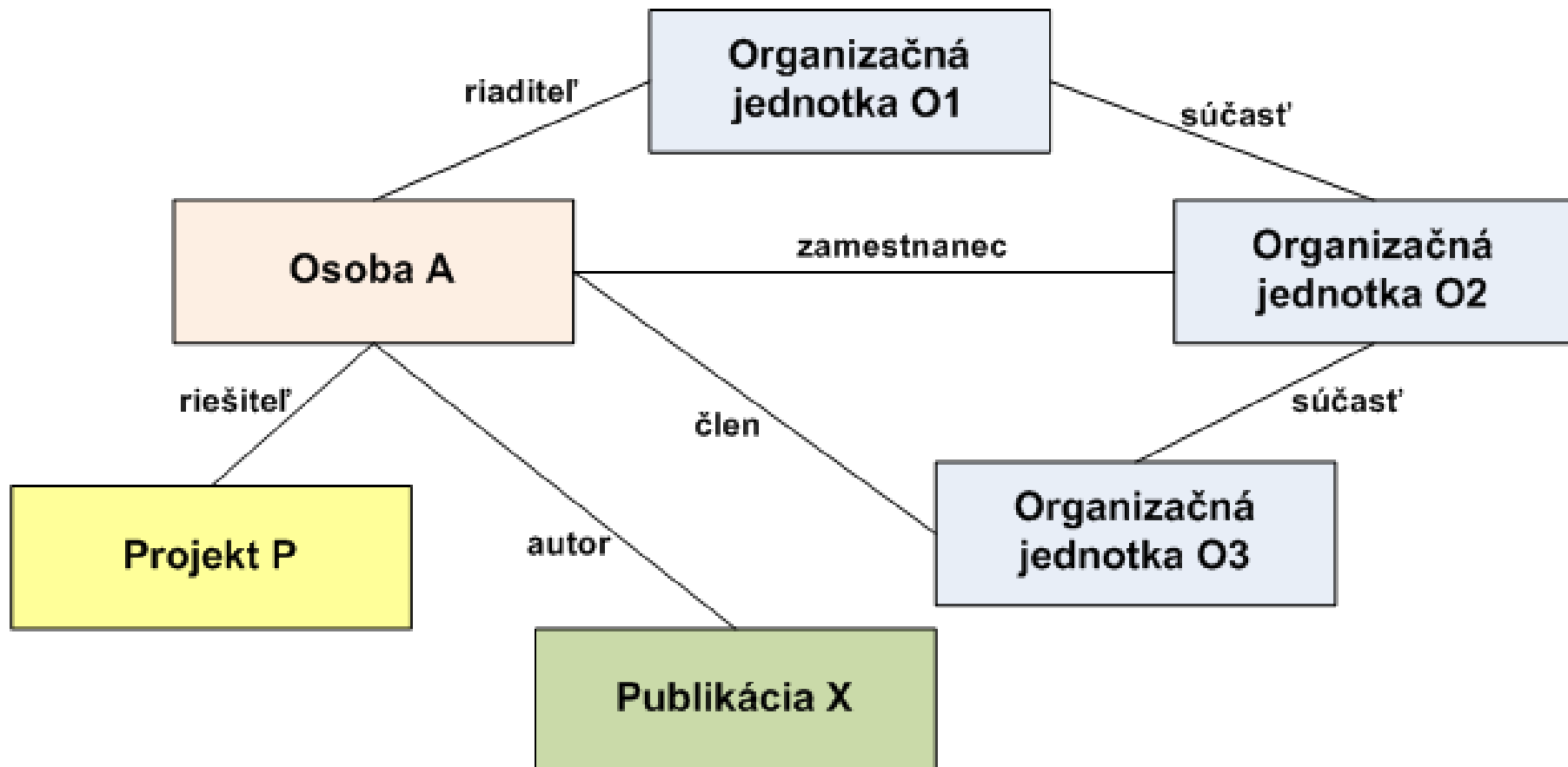
Kedy sa **téma** prvý krát spomína a v ktorom roku sa o nej písalo najviac?

CERIF

- euroCRIS 1991
- Štandard, zaoberajúci sa entitami v oblasti vedy a výskumu a ich vzájomnými vzťahmi
- Štyri skupiny vzájomne prepojených entít:
 - Základné entity (core entities) – projekt, osoba, organizácia
 - Výsledkové entity (result entities) – publikácia, patent, produkt
 - Entity druhej úrovne (2nd level entities)
 - Väzobné entity (Link entities)

CERIF – vzťahy a ich sémantika

- V modeli sú prepojené všetky hlavné entity a vzťahy medzi nimi môžu byť ohodnotené.



CERIF

- euroCRIS 1991
- Štandard, zaoberajúci sa entitami v oblasti vedy a výskumu a ich vzájomnými vzťahmi
- Štyri skupiny vzájomne prepojených entít:
 - Základné entity (core entities) – projekt, osoba, organizácia
 - Výsledkové entity (result entities) – publikácia, patent, produkt
 - Entity druhej úrovne (2nd level entities)
 - Väzobné entity (Link entities)
- **logický aj fyzický dátový model**
- **rozšíriteľnosť**
- **vlastná xsd schéma**

Niektoré existujúce riešenia

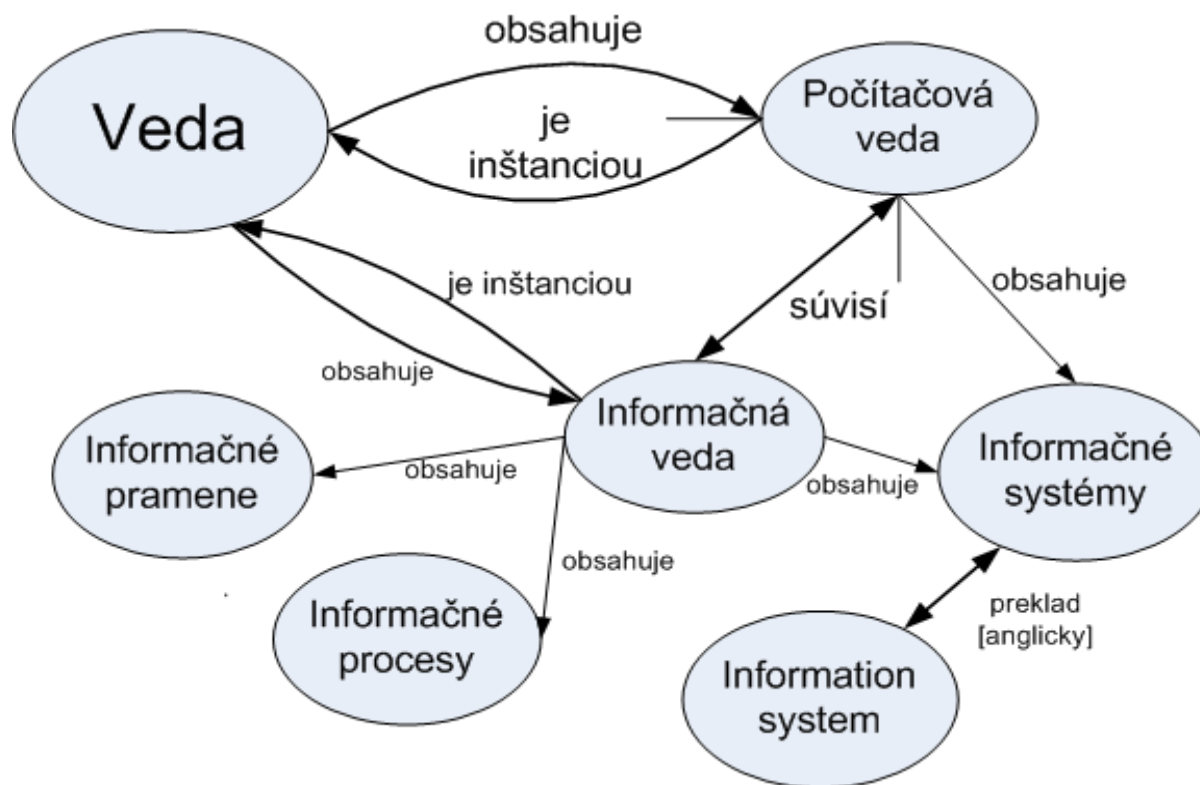
- Národná úroveň
 - Nórsko, Holandsko, Rakúsko, Slovinsko, Maďarsko a ďalšie
- Inštitucionálna úroveň
 - Univerzity vo Anglicku, Škótsku, Dánsku a iných krajinách
- Viaceré problémy:
 - Identifikácia jedno-jednoznačných termínov
 - Duplicita dát
 - Prepojenie dát podľa významu

Náš projekt

- Návrh riešenia pre FIIT STU
- Východiskový stav – zdroje dát
 - Výsledky výskumnej činnosti – CREPČ
 - OLIB
 - Projekty AIS
 - Ohlasy – ručné spracovanie z databáz WoS a Scopus
 - Akcie, konferencie, www stránky pracovníkov
- Návrh dátového modelu(rozšírenie modelu CERIF)
- Metódy pre získavanie a spracovanie dát
- Zobrazenie dát

Návrh modelu

- CERIF + tematická mapa pojmov
- Tematické pojmy môžu byť prepojené medzi sebou rôznymi vzťahmi:



Opis základného procesu

Vyhľadanie zdrojov

Získanie obsahu a jeho analýza

Spracovanie obsahu a mapovanie na dátový model

Zobrazenie entít a vytvorených vzťahov

Získavanie dát

- Automatizované získavanie dát pomocou technológií
 - Architektúra orientovaná na služby
 - Automatizované získavanie ohlasov z renomovaných citačných databáz
- Automatizované získavanie dát z rôznych formátov
 - MARC 21
 - UNIMARC
 - CREPČ XML prípadne iná schéma
- Spracovanie dát nezávislé od ich zdroja
 - Poskytované ako služba

Spracovanie dát

- Identifikácia entít, ich vlastností a sémantiky
- Ošetrovanie homoným, viacvýznamovosti
 - Zavedenie jednoznačných identifikátorov pre inštancie entít
 - Rozlišovanie entít podľa kontextu a vlastností
- Automatizované spracovanie textu
 - Regulárne výrazy, N-gramy, Dice's koeficient, Jaro–Winklerova vzdialenosť , Levensteinova vzdialenosť
- Ošetrovanie preklepov
- Ošetrovanie duplícít
- Odstraňovanie autocitácií

Zobrazenie dát – mapa pojmov

Výsledky pre vyhľadávaný pojem: Semantic

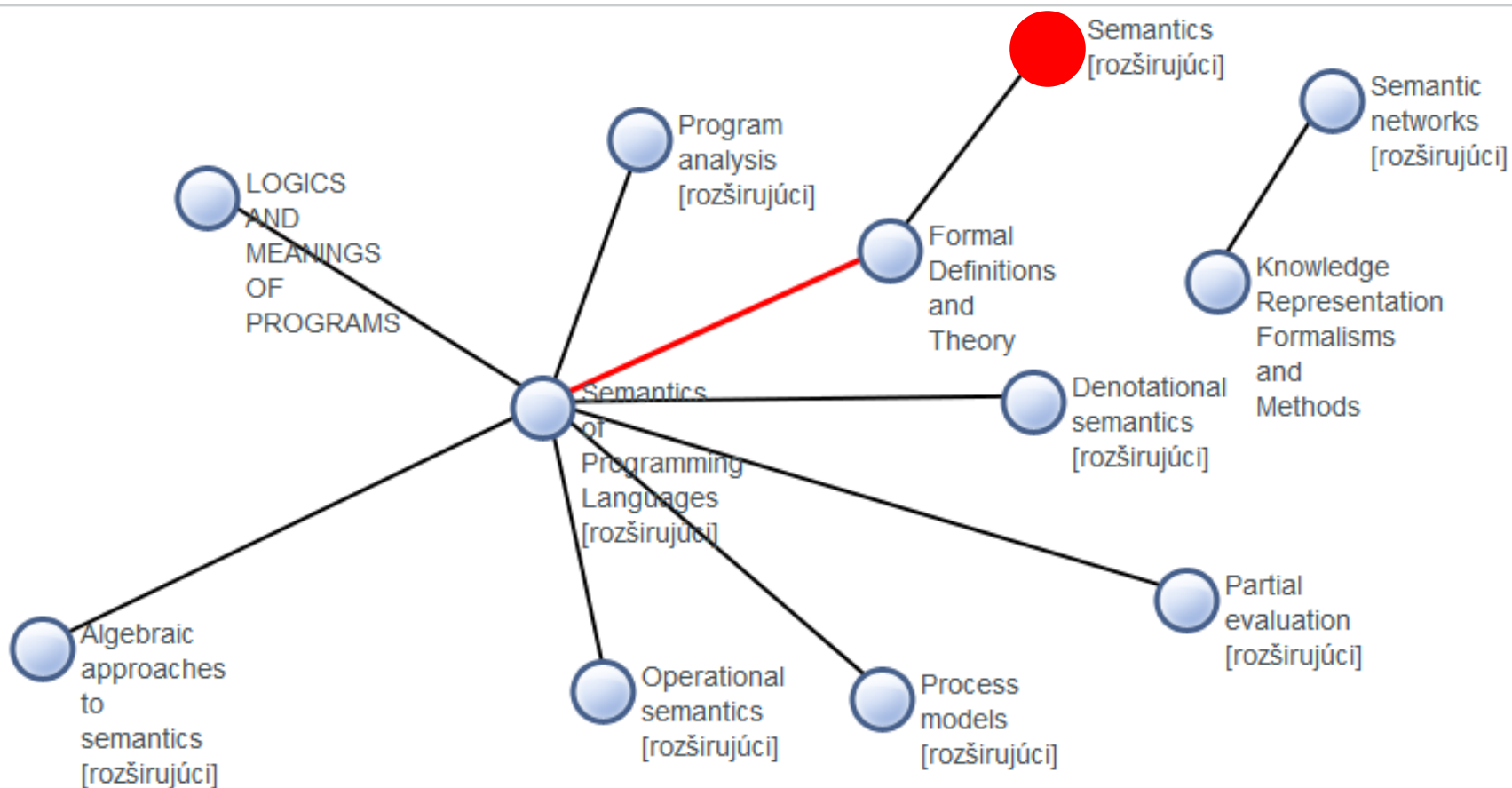
Mapa pojmov

Publikácie [0]

Výskumníci [0]

Projekty [0]

Červená farba čiary vyjadruje vzťah súvisiacich pojmov a čierna vyjadruje vzťah rozšírenia medzi pojmi.



Zobrazenie dát

Úvod Vyhľadávanie Návod

Vyhľadávanie v databáze:

Vyhľadaj

Zobrazenie vyhľadávaných informácií

Výsledky pre vyhľadávaný pojem: semantic

Mapa pojmov				Publikácie [81]				Výskumníci [46]				Projekty [5]				Konferencie [3]			
Názov				Ohlasy				Rok				Projekt							
An Approach to Named Entity Disambiguation Based on Explicit Sem.				0				2011				neviaže sa na projekt							
SPEM Ontology as the Sematic Notation for Method and Process Def				0				2011				Adaptívny sociálny web a jeho služby p							
Semantic media adaptation and personalization				0				2011				neviaže sa na projekt							
Exploratory Search in the Adaptive Social Semantic Web				0				2011				neviaže sa na projekt							
Acquisition of Semantic Metadata via Interactive Games				0				2011				neviaže sa na projekt							

Zhodnotenie

- Univerzálna použiteľnosť metód
- Rozšírenie modelu CERIF
- Automatizácia získavania dát
- Spracovanie získaných dát
- Spracovanie dát z ľubovoľného zdroja
 - Import v rôznych formátoch
 - Spracovanie dát ako služba
- Pohľad na vedu a výskum v kontexte a súvislostiach

Ďakujem za pozornosť.

xmajchrakm2@fiit.stuba.sk

andrejcikova@fiit.stuba.sk