

UNIVERSIDADE FEDERAL DE SANTA CATARINA

## Construindo Aplicações na Web Semântica Problemas Fundamentais (Desafios) da Web Semântica

Renato Fileto  
fileto@inf.ufsc.br

Programa de Pós-graduação em Ciência da Computação – PPGCC  
Departamento de Informática e Estatística – INE  
Centro Tecnológico – CTC

## Tópicos

- **Construção de ontologias**
  - (Semi-)automatizada e interativa
- **Geração de anotações semânticas de recursos**
  - Muitas aplicações requerem automatização
- **Evolução de ontologias**
  - Checar se alterações não introduzem inconsistências e gerenciar versões
- **Integração de ontologias (*ontology matching*)**
  - A obtenção de consenso para a construção e adoção de uma única ontologia é inviável em muitos domínios
- **Dedução/raciocínio**
  - Teorias computáveis e não computáveis; ontologias inconsistentes, ...

## Edição de Ontologias no Protégé

- Editor de ontologias e arcabouço para o desenvolvimento de bases de conhecimento livre e de código aberto baseado em Java.
- Suporta os modelos de *frames* (RDF) e OWL.
- Exporta as ontologias e bases de conhecimento para uma variedade de formatos, incluindo RDF(S), OWL e XML Schema.
- Extensível via *plugins*.

## Construção de ontologias

- **Processo para geração automatizada de ontologias a partir de esboços, dados e documentos** (nova geração de ferramentas)
  - Entendimento do domínio
  - Entendimento dos dados
  - Definição das tarefas
  - **Geração semi-automática**
  - Avaliação de qualidade
  - Refinamento com interação humana

## Definições de “geração de ontologia” Grobenik and Mladenic, 2006

- **Indução de conceitos** (a partir das instâncias)
- **Indução de relações** (a partir de conceitos e instâncias a eles associadas)
- **Popular ontologia** (dada uma ontologia e instâncias de seus conceitos)
- **Geração de uma ontologia completa** (a partir das instâncias e algum conhecimento/informação)
- **Atualização/extensão de ontologias** (dada uma ontologia e informação como novas instâncias ou padrões de uso da ontologia)

## Document Clustering - Vector Model

### Vector of word occurrences in the documents:

$$d_{ik} = TF(W_k, d_i) * IDF(W_k)$$

### Where

- $TF(W_k, d_i)$  = number of times  $W_k$  occurs in document  $d_i$
- $IDF(W_k) = \log ( D / DF(W_k) )$
- $D$  = number of documents
- $DF(W_k)$  = number of documents in which word  $W$  occurs

### Document distance metric:

$$\cos(d_i, d_j) = \frac{\sum_w d_{iw} d_{jw}}{\sqrt{\sum_w d_{iw}^2} \sqrt{\sum_w d_{jw}^2}}$$





## Annotation example (cont.)

The morphological analyser gives the roots of all the words.

[There]: root—there

[are]: root—be

[deliverables]: root—deliverable

There are — — —.

triggers the creation of one or more new classes in the ontology.

## Annotation example (cont.)

There are projects.

There are workpackages, tasks and deliverables.

SEKT is a project.

MUSING, 'Knowledge Web', and 'Presto Space' are projects.

Projects have workpackages.

Workpackages can have tasks.

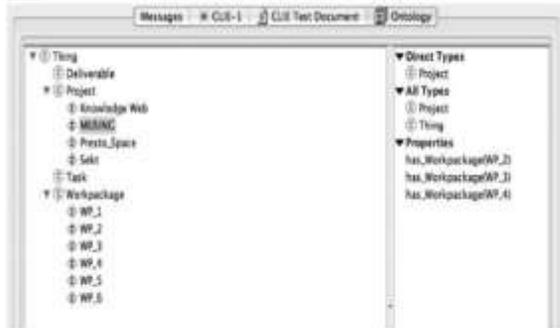
WP1, WP2, WP3, WP4, WP5 and WP6 are workpackages.

SEKT has WP1.

MUSING has WP2, WP3 and WP4.

'Knowledge Web' has WP5 and WP6.

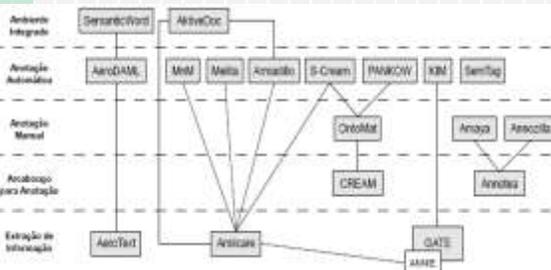
## Ontologia resultante



## Quadro Comparativo (Nunes e Fileto, 2007)

Nome da Ferramenta	Tipo da Ferramenta	Análises Geradas	Processo Anotação	Evolução Ontologia	Formato Documentos
OntoMap: Anotador	Série-automática	CWL	Interativo	Sim	HTML
Sensato: Word	Série-automática	DAML+OIL	IE (AeroDAME)	-	Word (DOC)
Annotica	Série-automática	RDF	Interativo	-	HTML
Share	Série-automática	CWL	OEIE	Sim	HTML, text, email, images
KIM	Autentadora	RDF+OWL	OEIE (GATE)	Sim	HTML
Midas	Série-automática	RDF+DAML+OIL	IE (Annotica)	Sim	HTML, text
MAM	Híbrida	CWL	OEIE (Annotica)	Sim	HTML, text
Gerador de Anotação Semântica para o processo de autoria	Série-automática	CWL	Interativo	Não	Writer (SCWF)

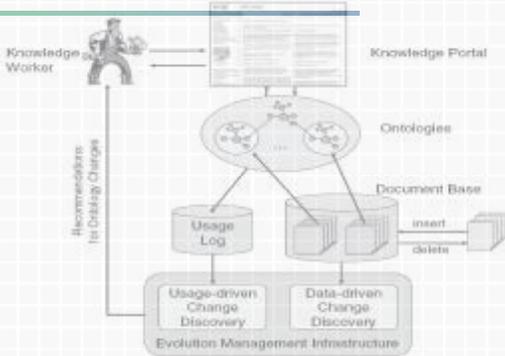
## Dependências entre plataformas (Rui Tramontin Junior, 2007)



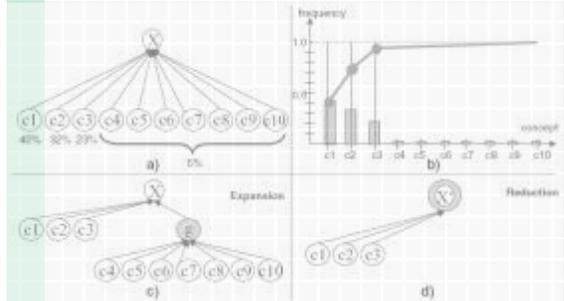
## Comparação de desempenho (Reeve e Han 2005)

Plataforma	Categoria	Precisão	Cobertura	Medida F
Armadillo	Descoberta de padrões	91.0	74.0	87.0
KIM	Regras	86.0	82.0	84.0
MnM	Indução	95.0	90.0	n/d
MUSE	Regras	93.5	92.3	92.9
PANKOW	Descoberta de padrões	65.0	28.2	24.9
SemTag	Regras	82.0	n/d	n/d

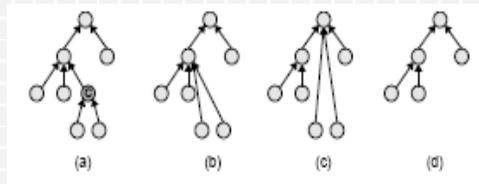
## Evolução de ontologias (Bloehdorn et al. 2006)



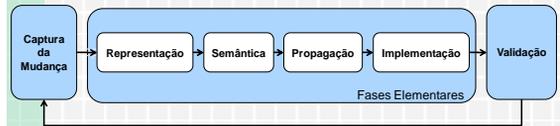
## Exemplo: Uso não uniforme de conceitos



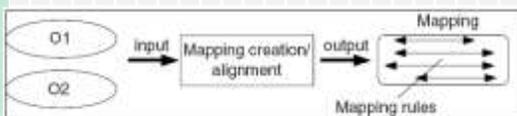
## Alternativas para remoção do conceito C que tem sub-conceitos



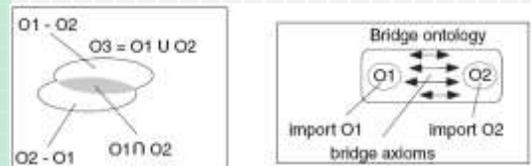
## Processo de evolução de ontologias



## Mapeamentos entre ontologias (Ontology Matching)

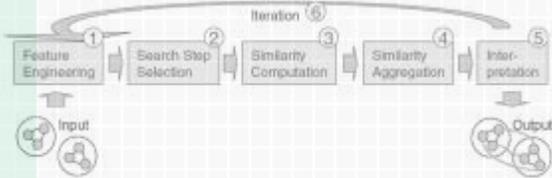


## Fusão de ontologias



## Processo semi-automático de alinhamento

(Huang et al., 2006)



## Referência

Jerome Euzenat, INRIA Rhone-Alpes, France,  
Pavel Shvaiko, University of Trento, Italy.

### Ontology Matching

URL: <http://book.ontologymatching.org>  
Springer-Verlag, Heidelberg (DE), 2007  
343 p., 67 illus., Hardcover  
ISBN: 978-3-540-49611-3

## Referências em Web semântica

- T. Berners-Lee, J. Hendler e O. Lassila. **The Semantic Web**. Scientific American, May 2001.
- Thomas B. **Explore's Guide to the Semantic Web**, Manning Publications, Greenwich, CT, 2005.
- Davies, J., Studer, R., Warren, P. (Eds.) **Semantic Web Technologies: trends and research in ontology-based Systems**, John Wiley & Sons, 2006.
- Breitman, K.K., Casanova, M.A., Truszkowski, W. **Semantic Web: Concepts, Technologies and Applications**. Series: NASA Monographs in Systems and Software Engineering, Springer 2007.
- Grigoris Antoniou and Frank van Harmelen **A Semantic Web Primer**, 2nd edition, The MIT Press, Cambridge, MA, USA, 2008.
- Kashyap, V., Bussler, C., Moran, M. **The Semantic Web - Semantics for Data and Services on the Web**, Series: Data-Centric Systems and Applications. Springer, 2008.
- Hitzler, P., Krotzsch, M., Rudolph, S. **Foundations of Semantic Web Technologies**. Chapman & Hall/CRC 2009.

## Ferramentas de desenvolvimento

- <http://protege.stanford.edu/>
- <http://jena.sourceforge.net/>
- <http://www.w3.org/2001/Annotea/>
- <http://www.ontotext.com/kim/>
- <http://esw.w3.org/topic/SemanticWebTools>
- <http://www.daml.org/2003/05/swmu-tools-tutorial/Overview.html>

## Perguntas?

