

# OnAIR: Improving multimedia search results by applying ontologies.

Raphael Mendes de O. Cóbe<sup>1</sup>  
raphael@ccsa.ufrn.br

<sup>1</sup>Instituto de Matemática e Estatística - IME  
Universidade de São Paulo - USP

Universidade de São Paulo



# Agenda I

- 1 Introdução
  - Web Semântica
  - Buscas Semânticas
- 2 Ontologias
  - Conceitos
  - Protégé-OWL editor
  - Reasoners
  - Jena - Java Semantic Web Framework
- 3 OnAIR

# Agenda

- 1 Introdução
  - Web Semântica
  - Buscas Semânticas
- 2 Ontologias
  - Conceitos
  - Protégé-OWL editor
  - Reasoners
  - Jena - Java Semantic Web Framework
- 3 OnAIR

# Histórico

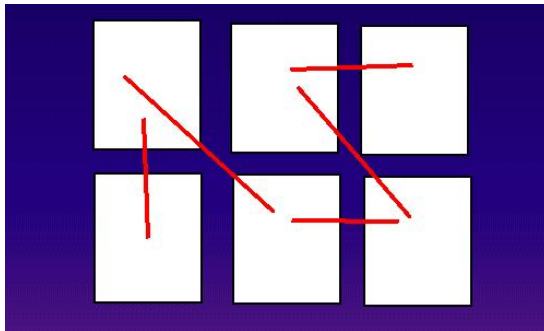
## *Evolução da Web*

- **1989** - Tim Berners-Lee (Http + Html);
- **1994** - W3C;
- **1994** - Web Semântica;
  - 1a Conferência Internacional WWW;

# Motivação

A web atualmente:

- Conjunto de Nós e Links;
- “To a computer, then, the web is a flat, boring world devoid of meaning.” (Berners-Lee).
- Muito pouco a respeito do conteúdo (não textual) é levado em conta;



Universidade de São Paulo



Figura: [Tim Berners-Lee 1994. <http://www.w3.org/Talks/WWW94Tim/>

# Motivação

Mudança no pensamento [1]

- Objetos publicados na Web descrevem coisas reais;
- Definição de relacionamentos especiais entre conceitos;
- Por Exemplo:
  - Um documento pode descrever um autor;
  - Um outro documento pode descrever uma música;
- Como relacionar dois documentos ?
  - Extensão da idéia de Link entre documentos;
  - Links como (diversos tipos de) relacionamentos;

# Motivação

Mudança no pensamento [2]

- Interpretação por seres humanos;
- Capacidade de abstração da realidade;

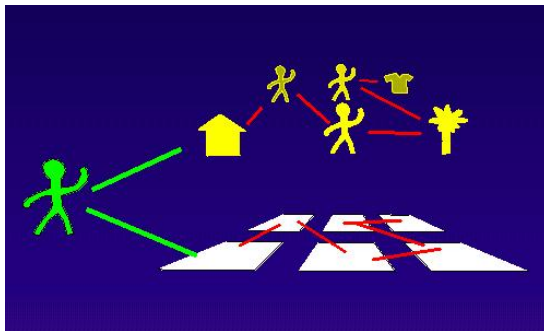
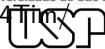


Figura: [Tim Berners-Lee 1994, <http://www.w3.org/Talks/WWW94Tim/>]

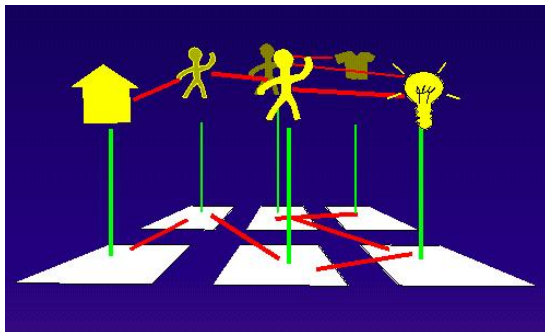
Universidade de São Paulo



# Motivação

## Uma nova Visão da Web

- Tarefas reais desempenhadas por máquinas;
- Por Exemplo:
  - Um programa pode procurar por um imóvel e negociar a troca de donos para um novo dono. [BLH01]



Universidade de São Paulo

USP

Figura: [Tim Berners-Lee 1994, <http://www.w3.org/Talks/WWW94Tim/>]



# Recuperação de Informação

- Aquisição organização, armazenamento, recuperação e distribuição de informação;
- Necessidade do Usuário expressa através de uma consulta;
- Pode não expressar da melhor forma esta necessidade;
- Resultados relevantes com base na consulta;
- Dificuldade em estabelecer relevância de documentos;
- **Polissemia e Sinonímia**;

# O Projeto OnAIR

- Quebra de Vídeos em trechos;
- Palavras chave descrevendo trechos;
- Definição de ontologias com palavras chave (ontologias de domínio);
- Classificação automática de ontologias;
- Expansão de consultas;
- Cálculo de similaridade entre termos;
- Redução da sinonímia e polissemia;

# Agenda

- 1 Introdução
  - Web Semântica
  - Buscas Semânticas
- 2 Ontologias
  - Conceitos
  - Protégé-OWL editor
  - Reasoners
  - Jena - Java Semantic Web Framework
- 3 OnAIR

# Ontologias

- Utilizadas em diversas áreas: filosofia, IA, desenvolvimento de sistemas, direito, medicina, etc...
- Vocabulário comum;
- Reuso de Conhecimento de um domínio específico;
- Descrição de Objetos, Propriedades e relacionamentos;
- Formas Lógicas de primeira ordem;
  - Lógicas de predicado;
  - Quantificadores universais e existenciais;
  - *Concepts and Roles*;

# O Uso de Ontologias

- **Taxonomias**
  - Yahoo! Directory;
- **Catálogos Online:**
  - Catálogo da Amazon.com;
- **Terminologias** de domínio específico:
  - SNOMED - Clinical Terms: Terminologia para medicina;
  - UNSPSC - Terminologia para produtos e serviços;

# Ontologias

## Implementação

- Separação em TBox e ABox;
- Formatos XML;
- RDF(S);
- OWL (Lite — DL — Full);
- Jena - Java Semantic Web Framework
  - *Programming toolkit.*
  - API para abstração da representação utilizada;
    - DAML+OIL, RDF(S), OWL (Lite, DS e Full);
- Protégé: Editor OWL;

# Ontologias

Protégé-OWL editor

- Mozilla Public License V.1.1;
- Editor para Ontologias OWL e RDF(S);
- Escrito em Java;
- Versão 4.0;
- Arquitetura baseada em eventos;
- Fácil extensão e construção de plugins;
- Plugins para visualização de ontologias - *OwlViz*;

# Ontologias

Protégé-OWL editor

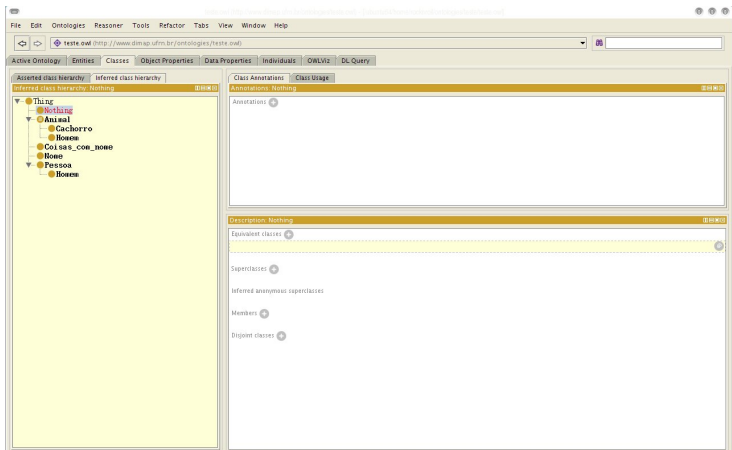


Figura: Screenshot do Protégé



# Ontologias

Protégé-OWL editor

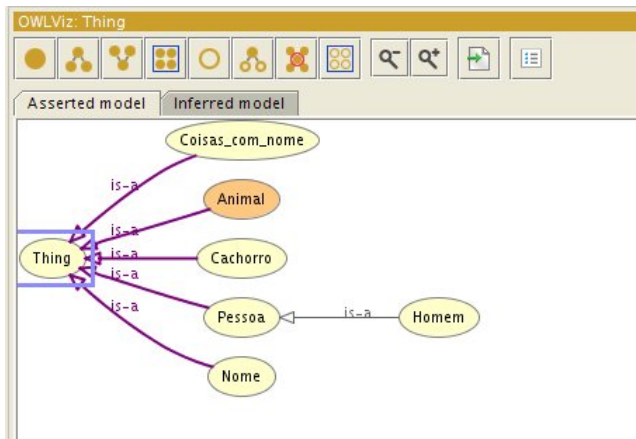


Figura: Screenshot do Protégé

# Reasoners

- Inferências entre conceitos;
- Faz uso do poder da lógica descritiva;
- Operações lógicas sobre ontologias;
- Detecção de Inconsistências;
- FaCT++
  - **GPL**;
  - Implementado em C++;
  - Alto desempenho;
  - Compatível formato DIG;
- Pellet OWL Reasoner
  - Dupla Licença: Comercial e **AGPL**;
  - Implementado em Java;
  - Desempenho não tão bom;

# Exemplo de Uso

- Exemplo de uso (Retirado do portal FaCT++):
  - A bus driver is a person that drives a bus;
  - A bus is a vehicle;
  - Everything that drives a vehicle is a driver;
  - A bus driver drives a vehicle, so must be a driver;

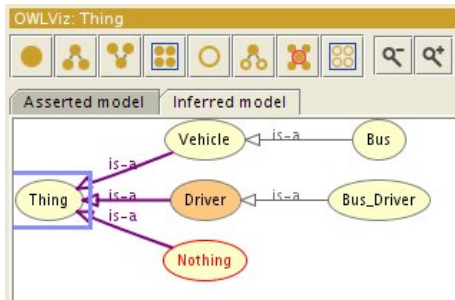


Figura: Hierarquia Inferida

# Jena - Java Semantic Web Framework

## Reasoning [1]

- Principal Motivação para a utilização de ontologias;
- Derivação de informações adicionais;
- Classificação taxonômica;
- API para inferências;
- Criar novo modelo de objetos;
- Hierarquia Inferida pode substituir Hierarquia comum;
- Compatibilidade com Reasoners DIG (*DL Implementation Group*);
  - Interface para a comunicação com reasoners (conexões HTTP);

# Jena - Java Semantic Web Framework

## Reasoning [2]

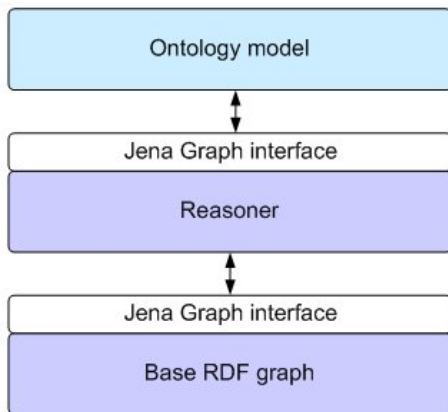
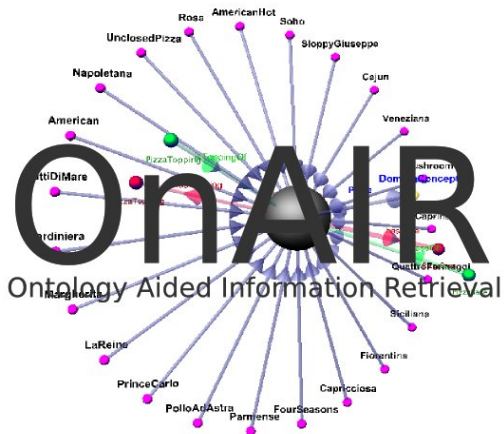


Figura: Jena Reasoning API

# Agenda

- 1 Introdução
  - Web Semântica
  - Buscas Semânticas
- 2 Ontologias
  - Conceitos
  - Protégé-OWL editor
  - Reasoners
  - Jena - Java Semantic Web Framework
- 3 OnAIR

# Ontology Aided Information Retrieval



# Descrição

- Recortar um vídeo de longa duração em pequenos trechos;
- Separação de acordo com o seu conteúdo (texto e meta-dados);
- Meta-dados providos por um especialista do domínio;
- Pergunta em Linguagem Natural;
- Recuperação de lista ordenada de documentos (vídeos) que melhor respondem a pergunta passada;
- Ontologias de domínio;
- Classificação Taxonômica;



# Descrição

## Entradas

- Coleção de Vídeos e transcrições:
  - Definição de espaços de vídeos e suas transcrições;
  - Faz uso de um arquivo de texto (Legenda do vídeo);
  - Até o momento, feito manualmente. :-)
- Palavras chave:
  - Palavras associadas a trechos de vídeo e que se encontram na Ontologia de domínio;
  - Possíveis substitutos para a transcrição;
- Ontologia de domínio:
  - Formato OWL contemplando conceitos utilizados no conjunto de vídeos;

# Descrição

## Artefatos Gerados

- Nova Ontologia:
  - Criação de nova ontologia criando Novas Classes para indivíduos;
  - Criação de propriedades Objeto para propriedades comuns;
- Índice Invertido:
  - Lista de termos e contagem de ocorrência por termos para cada documento;
  - Também é gerado um índice invertido para Palavras-chave;

# Processo de Recuperação

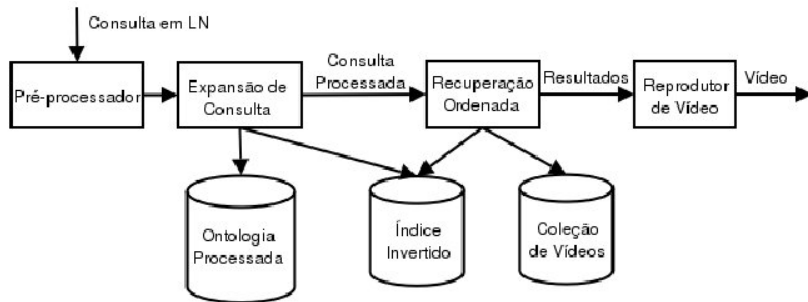


Figura: Fluxo para recuperação da informação

# Processo de Recuperação

- **Pré-processamento:**

- Correção ortográfica e sugestões;
- Definição de pesos para a consulta;
  - Termos que se encontram na ontologia recebem peso maior;
  - Baseado na contagem de termos que aparecem na coleção de vídeos;

- **Expansão da Consulta:**

- Inclusão de termos \*similares\* contidos na ontologia;
  - Termos diretamente sinônimos;
  - Propriedades que relacionam termos;

- **Busca em palavras-chave e transcrições;**

## Screenshots

The screenshot displays the OnAIR web interface. At the top, the title "OnAIR: Ontology-Aided Information Retrieval" is visible in the browser's address bar. The main header features the "OnAIR" logo on the right and navigation links for "about OnAIR" and "about Joe Yoder" on the left. Below the header, a video player is shown with the title "Joseph Yoder talking about patterns". To the left of the video player, there is a search results panel with the heading "put your question" and a search input field containing "What are Design Patterns?". Below the search input, the text "click on one result: watch all the videos sequentially" is displayed, followed by a numbered list of results:

1. An example of better communication through design ...
2. Patterns Definition
3. Yoder's wills
4. Course explanation
5. Why Design Patterns?

At the bottom of the search results panel, a progress indicator shows "00:13 / 00:48". The video player shows a man in a patterned shirt gesturing with his hands. In the bottom right corner of the video player area, the text "idade de São Paulo" and the "USP" logo are visible, along with navigation icons for a menu, back, search, and refresh.

## Screenshots

Arquivo

Personagem:

Título:

Clips				Recursos de apoio		Associacoes de recursos	
Nome	Arquivo	Duracao	Palavras Chave				
Book's general ...	WebContent/resourc...	5234	gof design_pa...				
Refactoring	WebContent/resourc...	5234	refactoring				
eXtreme Progr...	WebContent/resourc...	5234	programming ...				
Course explana...	WebContent/resourc...	5234	patterns cover				
Why Design Pat...	WebContent/resourc...	5234	principals desi...				
Good OO progr...	WebContent/resourc...	5234	methods obje...				
Parallel evolutio...	WebContent/resourc...	5234	communicatio...				
Guy Steele talk	WebContent/resourc...	5234	words language				
Design Patterns...	WebContent/resourc...	5234	extending co...				
An example of ...	WebContent/resourc...	5234	patterns com...				
Design Patterns...	WebContent/resourc...	5234	power pattern...				
Motivation to d...	WebContent/resourc...	5234	programming ...				
UNIX Parallel to...	WebContent/resourc...	5234	trade guru pat...				
Patterns Definit...	WebContent/resourc...	5234	problem soluti...				
Christopher Ale...	WebContent/resourc...	5234	architecture s...				
What are forces...	WebContent/resourc...	5234	system forces ...				
Engeneering ex...	WebContent/resourc...	5234	documenting ...				

Nome:

Arquivo:

Duracao(s):

Ponto inicio (srt):

Ponto fim (srt):


Palavras Chave:

Texto do clip:

idade de São Paulo

**ISP**

☰ 🔍 ↻

-  T. Berners-Lee and J. Hendler.  
Scientific publishing on the semantic web.  
*Nature*, 410:1023–1024, 2001.