



FACULDADE · DE · CIÊNCIAS UNIVERSIDADE · DE · LISBOA

EdV-Linguatca 2006

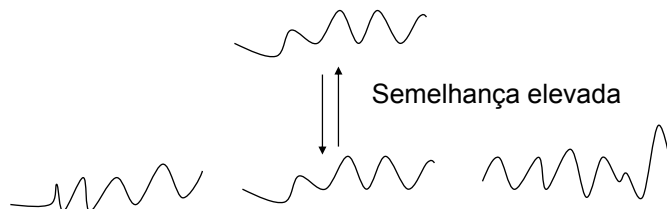
Ontologias & Terminologias: Perspectivas da engenharia

Mário J. Silva

Julho de 2006

Transferência de Características

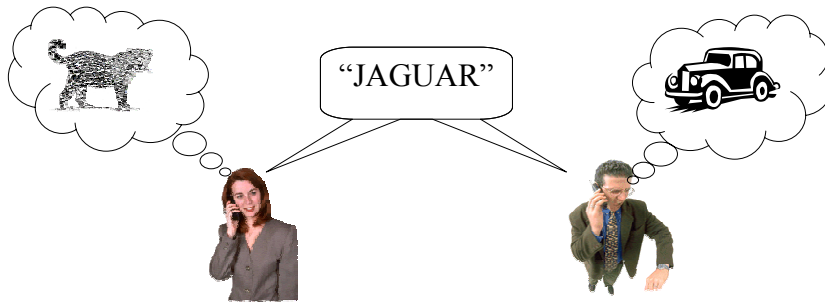
Proteína não caracterizada



Proteínas caracterizadas

- Transferimos o que sabemos
- O que sabemos é o conhecimento

Partilha de palavras e significado



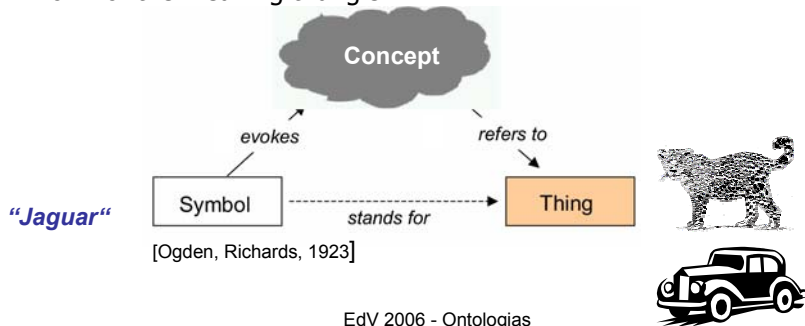
Biologia pós-genómica (e-science)

- Organismos modelo: mosca da fruta, rato, fermento, minhoca
 - Cada comunidade usa a sua terminologia.
- Como comparar genomas?
 - Torna-se necessária a sistematização do conhecimento partilhado.

<http://geneontology.org>

O Triângulo Semântico

- Humans require words (or at least symbols) to communicate efficiently. The mapping of words to things is only indirectly possible. We do it by creating concepts that refer to things.
- The relation between symbols and things has been described in the form of the meaning triangle:



EdV 2006 - Ontologias

5

Conhecimento Partilhado

- Comparações tornam-se muito mais simples:
Podemos colocar perguntas em simultâneo sobre vários recursos.
 - Uma estrutura de relacionamentos possibilita a descoberta e e formulação de abstrações.
 - Base de conhecimento partilhado permite colocar interrogações sobre bases de dados de domínios diferentes
- **Útil**
 - **Para humanos e máquinas (entre e com)**

EdV 2006 - Ontologias

6

Há 3 perspectivas (Yorick Wilks)

1. Engenharia do Conhecimento
2. Processamento de Linguagem Natural
3. Base de Dados (1. com visão ainda mais restrita)

Agenda

- Porquê Ontologias?
- **Ontologias e Engenharia do Conhecimento**
 - Ontologias na Web Semântica
- Ontologias e PLN
- Construção de Ontologias por Captura de Conhecimento
- Exercício: construção de uma ontologia
exemplo de *knowledge elicitation*

Sintaxe e Semântica

- Infix $2 + 3 = 5$
- Prefix $= + 2 3 5$
- Postfix $2 3 + 5 =$
- Binary $010 + 011 = 101$
- Roman $II + III = V$
- $7 + 3 = 42$

O que é uma Ontologia?

“uma especificação de uma conceptualização”

Conceptualização =
conjunto de conceitos usados por
pessoas para comunicar num dado
domínio
+ relacionamentos entre esses
conceitos.

**Não existe definição
universalmente aceite**

O que define uma Ontologia?

- Um vocabulário de termos comuns
- Alguma especificação do significado dos termos
- Uma percepção partilhada por pessoas e humanos.

Vocabulário Controlado

- Cada elemento tem um nome único
- A definição de cada elemento encontra-se especificada
- Deve haver um só elemento na ontologia a representar cada uma das entidades e relacionamentos.

Elementos de uma Ontologia (aceites por quase todos)

- **classes** = conjuntos de coisas
- **instances** = membros de classes
- **relationships**
- **axioms** = proposições lógicas adicionais

Hierarquia (taxonomia)

- **subclass** = classe descendente
- **direct subclass** = classe filha
- **superclass** = classe ascendente
- **direct superclass** = classe mãe

Herança

- **Hierarquia simples** ou **múltipla**
- Relação **is-a**
- Uma classe **herda** as propriedades que tiverem sido definidas para a(s) sua(s) superclasse(s)

Conhecimento Ontológico

Nem tudo pode ser definido ou sequer descrito

- Art = “I know what it is when I see it”
- ...

Definições de Ontologia em Informática

Há muitas definições:

- An ontology is an explicit specification of a conceptualization [Gruber93]
- An ontology is a shared understanding of some domain of interest. [Uschold, Gruninger96]

Uma especificação formal

EXECUTÁVEL

conceptualização de um domínio por uma

COMUNIDADE

relativa a uma parte do mundo de interesse

APLICAÇÃO

Ontologia vs. Terminologia

- Define Conceitos (termos ontológicos)
- Especificação formal, Executável
- Grão grosso: terminologias técnicas
- Explica termos lexicográficos num contexto
- Especificação Informal, não executável
- Grão-fino: polisemias

Fronteira ténue!

A semantic continuum

(Mike Uschold, Boeing Corp)

Pump: "a device for moving a gas or liquid from one place or container to another"

(pump has (superclasses (...))

Shared human consensus

Text descriptions

Semantics hardwired; used at runtime

Semantics processed and used at runtime

Implicit

Informal (explicit)

Formal (for humans)

Formal (for machines)

→ Further to the right →

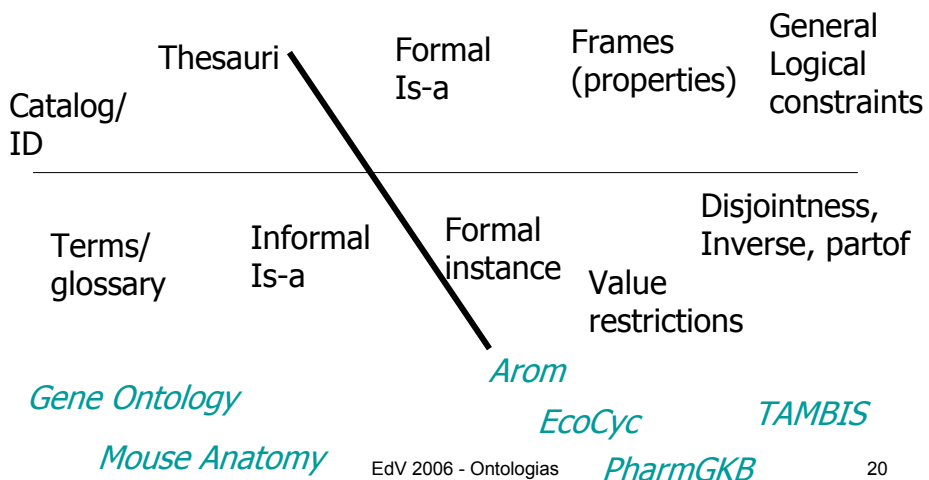
- Less ambiguity
- Better inter-operation
- More robust – less hardwiring
- More difficult

EdV 2006 - Ontologias

19

Categorização de Ontologias

(Deborah McGuinness, Stanford)



EdV 2006 - Ontologias

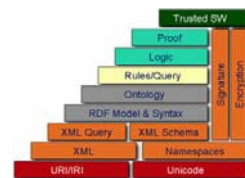
20

Porquê desenvolver ontologias?

- **Explicitar os pressupostos do domínio**
 - Para mais facilmente os poder alterar
 - Para facilitar a compreensão e actualização de dados legados.
- **Separar o conhecimento do domínio do conhecimento operacional**
 - Para poderem ser reutilizados separadamente
- **Constituir referência de comunidade que desenvolve (um conjunto de) aplicações**
 - Para partilhar uma percepção consistente do significado de um conjunto de informação.

Agenda

- Porquê Ontologias?
- Ontologias e Engenharia do Conhecimento
 - **Ontologias na Web Semântica**
- Ontologias e PLN
- Construção de Ontologias por Captura de Conhecimento
- Exercício: construção de uma ontologia exemplo de knowledge elicitation



Problemas (re)correntes em Gestão do Conhecimento

- Busca é muito limitada
 - baseada em palavras-chave
- Extração de informação é difícil.
- Inconsistências nas terminologias
 - falhas na remoção de informações desactualizadas.
- Prospecção de dados distribuída (descoberta de conhecimento implícito) é difícil.
- Views (na acepção das Bases de Dados) são difíceis de concretizar.

Gestão do Conhecimento: **a Proposta da SW**

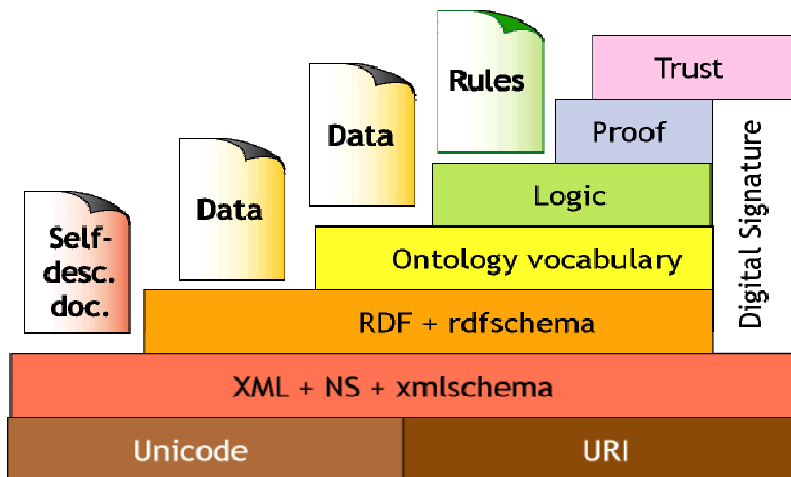
- Conhecimento particionado em domínios
- Ferramentas automáticas
 - detectam inconsistências
 - extraem conhecimento
- Plataforma para resposta a perguntas (a partir de vários documentos)
 - Alternativa à busca baseada em palavras-chave
- Possibilidade de restrição de quem pode ver fragmentos da informação.

Tecnologias da SW

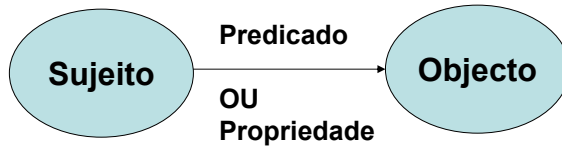
- XML e RDF
- Ontologias
- Lógica
- Utilizadores:
 - Agentes de recolha de informação
 - isto não é IA (clássica)

Cuidado com os cenários idílicos!

O bolo de camadas “layer cake”



Grafo RDF / Triplos



- Um grafo RDF representa a conjunção (AND) das proposições que correspondem a todos os triplos nele contidos.
- RDF usa URIs para identificar recursos e propriedades.

<http://www.w3.org/TR/rdf-primer/>

Representação Gráfica RDF



N3 – Notação 3

- Em RDF a informação é apenas uma colecção de proposições, cada uma delas com sujeito, predicado, complemento – e apenas isso.
- N3: um script de triplos RDF com um ponto:
<#belinda> <#knows> <#jj> .

<http://www.w3.org/2000/10/swap/Primer>

RDF em Sintaxe XML



```
<rdf:RDF>
  <rdf:Description rdf:about="#00000"> resource
  ... some xml...
</rdf:Description>
  <rdf:Description rdf:ID="#00001"> URI
    <some-pred-tag rdf:resource="00000" /> reference
    <rdf:type rdf:resource="&ns:type" /> instance-of
  </rdf:Description>
</rdf:RDF>
```

E **containers** como:
<rdf:Bag>, <rdf:Seq> (Sequence), <rdf:Alt> (alternatives)

RDF: Discussão

- RDF é independente do domínio
 - Apenas colecção de triplos, onde cada uma dos elementos (e tipos dos seus valores) é apenas um URI!
- E o vocabulário?
RDFS, RDF Schema.

RDF Schema

- Linguagem de descrição de vocabulário (linguagem “primitiva” de representação de ontologias)
 - Semelhante ao sistema de tipos das linguagens de programação OO (como Java)
 - Em vez de definir classes em termos das propriedades das instâncias, define propriedades em termos das classes a que se aplicam.

RDF Schema: *property-centric approach*



OO Clássica:

classe eg: **Book**
atributo eg: **author**
tipo eg: **Person**.

RDFS:

property eg: **author**
domain
eg: **Document**
range of eg: **Person**

- Permite estender a descrição dos recursos existentes, respeitando um dos princípios arquiteturais da Web
- **RDFS** possibilita que outros definam propriedades adicionais com **domain** eg: **Document** e **range** eg: **Person** sem redefinir descrição das classes.

Ontologias: *trade-off* fundamental

- Quanto mais rica for a linguagem mais ineficiente se tornará o suporte para raciocínio.
- Por vezes, o raciocínio não é computável.

OWL

Web Ontology Language – W3C

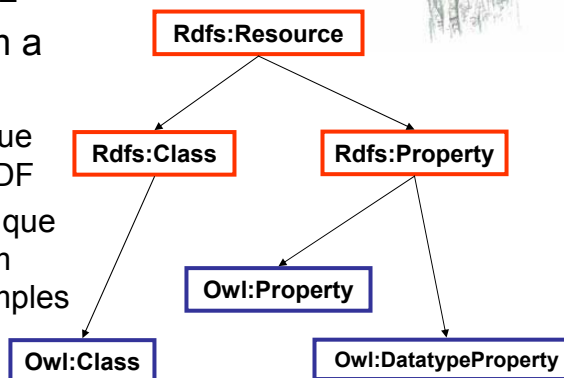


- **OWL Full**
 - compatível com RDF
 - **Undecidable**
- **OWL DL**
 - constructors de OWL não podem ser aplicados sobre constructors de OWL
 - **Raciocínio eficiente.**
- **OWL Lite**
 - Conjunto restrito de construtores.
 - **Não há** classes enumeradas, proposições disjuntivas, cardinalidade arbitrária.

OWL



- Sintaxe gráfica semelhante a UML
- RDF/XML definem a sintaxe.
 - XML que não segue as convenções RDF
 - Sintaxe abstracta que não é baseada em XML (e é mais simples de ler)



Restrições



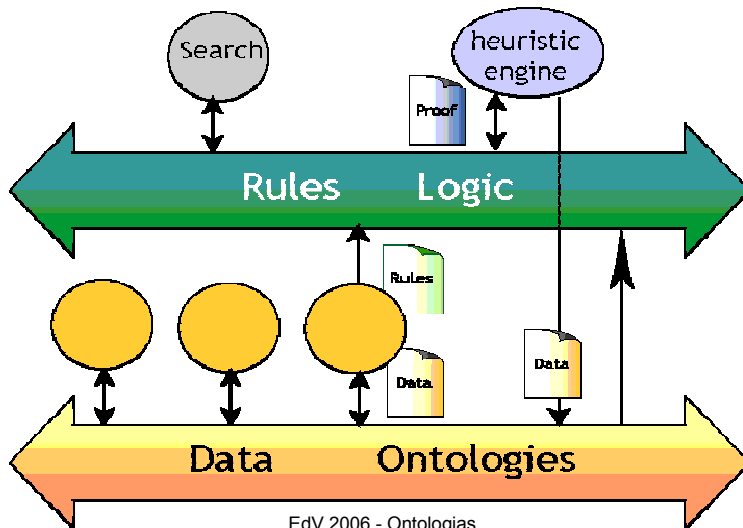
- OWL Full – sem restrições, compatível com RDF
- OWL DL
 - Particionamento do vocabulário: recursos (conceitos) podem ser apenas de um tipo: classe, atributo, ...
 - Tipos declarados explicitamente: declarar classe mesmo que se tenha já dito que é subclasse de qq coisa.
 - Separação de propriedades: no inverseOf, symmetric Property, ...
 - No transitive cardinality restrictions
 - ...
- OWL Lite
 - Constructors como owl:disjointWith não permitidos
 - Cardinalidade restrita a 0 ou 1
 - owl:equivalentClass não permitido

Compatibilidade em OWL



- Legalidade Construções:
 - Legal Lite => Legal DL => Legal Full
- Conclusões:
 - Valid Lite => Valid DL => Valid Full
- Todas as representações usam Sintaxe RDF
- Construções OWL são especializações de construções RDF

A 2ª figura...



EdV 2006 - Ontologias

39

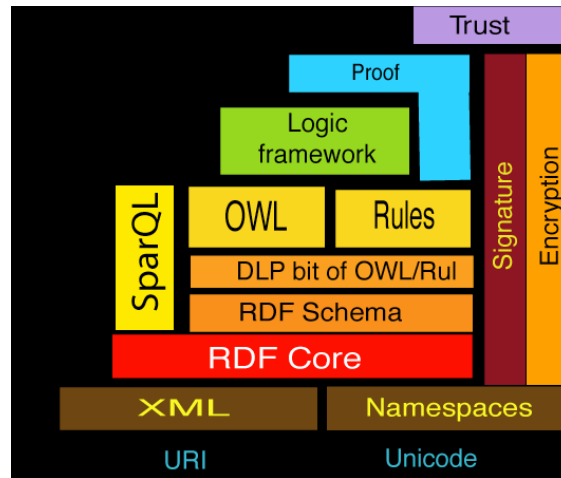
Regras e Ontologias

- A história não termina com OWL
 - Semantic Web permite várias ontologias.
- Regras como alternativas às ontologias
 - em vez de construídas sobre elas (o que iria requerer combinação)
- RIF – Rule Interchange Format
 - processo iniciado recentemente (2005)

EdV 2006 - Ontologias

40

O bolo mais fresco...@iswc2005

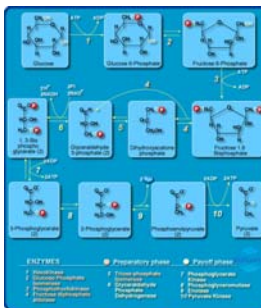


EdV 2006 - Ontologias

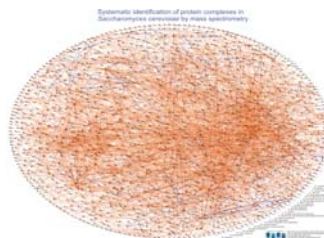
41

The domain: Biological pathways

Main categories:



Metabolic Pathways



Molecular Interaction Networks



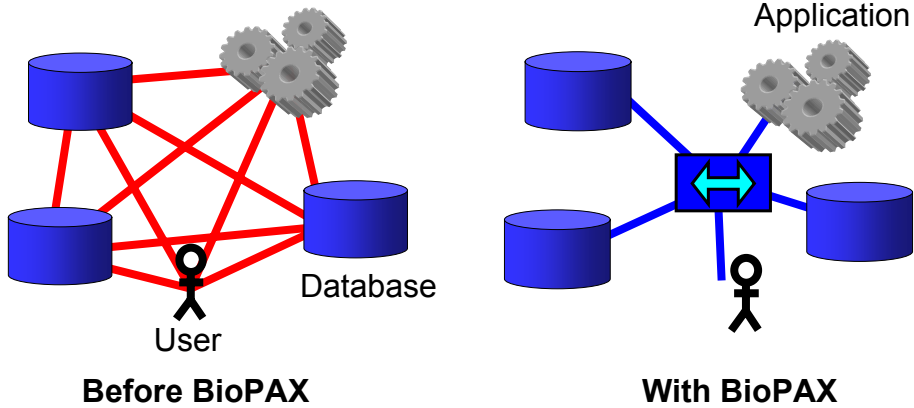
Signaling Pathways

EdV 2006 - Ontologias

42

BioPAX Motivation

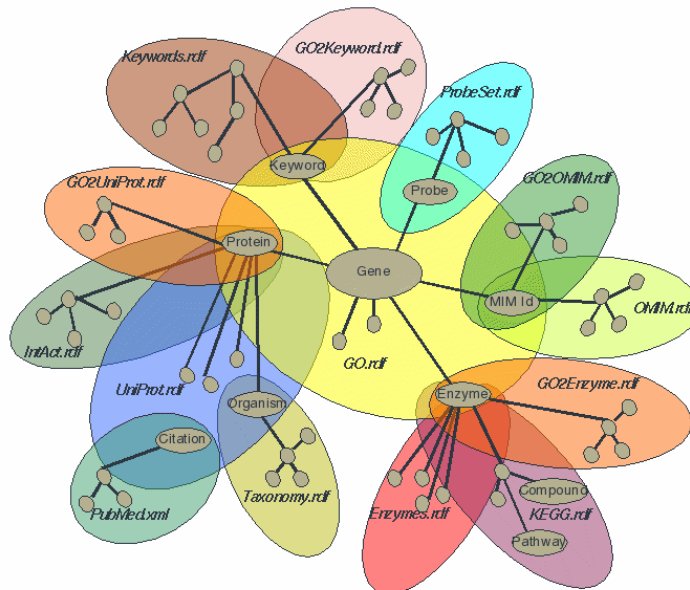
>180 DBs and tools



Usa todo o arsenal da SW de hoje: reusa ontologias, “constructs” para referência deregistos em BDs, sinónimos, Xrefs, relacionamentos, raciocínio OWL, ...

EdV 2006 - Ontologias

43



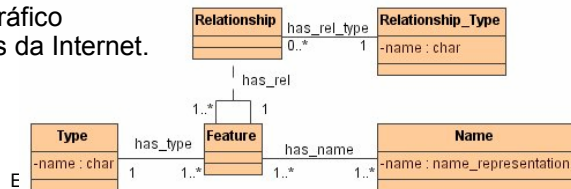
Geo-Net-PT-01



<http://www.di.fc.ul.pt/tech-reports/05-12.pdf>

- Ontologia geográfica de Portugal.
 - Criada pelo Grupo XLDB Universidade de Lisboa
 - Projecto GREASE.
 - Disponibilizada via Pólo XLDB da Linguateca.
- A Geo-Net-PT01 contém
 - 418.065 dados geográficos administrativos de Portugal
 - informação administrativa de BD legadas
 - inclui o âmbito geográfico atribuído a 686 sítios da Internet.

• OWL



Agenda

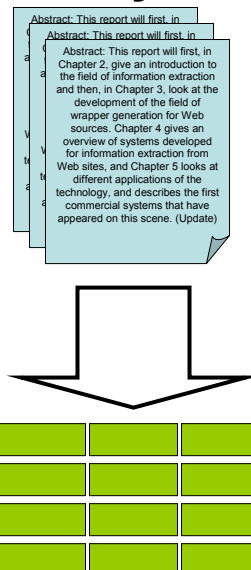
- Porquê Ontologias?
- Ontologias e Engenharia do Conhecimento
 - Ontologias na Web Semântica
- **Ontologias e PLN**
- Construção de Ontologias por Captura de Conhecimento
- Exercício: construção de uma ontologia exemplo de knowledge elicitation

Ontologias e PLN

- Terminologias a evoluir para a formalização sob a forma de ontologias
- Uso de métodos de extracção de informação para construção de ontologias.

Extracção de Informação

- Preenchimento de modelos (*templates*)
- Oportunidade de povoar ontologias com indivíduos (*instances*) identificados com recurso a métodos estabelecidos (REM)
- Extracção de sub-classes



Ontologias e Redes Sociais

- Web2.0, tags e tag clouds
- Folksonomies
 - Vocabulário in/des-controlado ☺
- + Dinâmico!
- + Partilhado!



Agenda

- Porquê Ontologias?
- Ontologias e Engenharia do Conhecimento
 - Ontologias na Web Semântica
- Ontologias e PLN
- **Construção de Ontologias por Captura de Conhecimento**
- Exercício: construção de uma ontologia exemplo de knowledge elicitation

Etapas do Desenvolvimento de uma Ontologia

1. Estabelecer o Propósito

- Sem propósito não há definição de âmbito, requisitos, avaliação

2. Aquisição Conhecimento (knowledge elicitation) Informal/Semiformal

- Coleccionar termos
- Organizá-los informalmente
- Clarificálos e produzir definições informais
- Diagramas informais

Eliciting Knowledge

Most knowledge is in the heads of experts

- Experts have vast amounts of knowledge
- Experts have a lot of tacit knowledge
- They don't know all that they know and use
- Tacit knowledge is hard (impossible) to describe
- Experts are very busy and valuable people
- Each expert doesn't know everything



...there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know.

EdV 2006 - Ontologias

53

Etapas do Desenvolvimento de uma Ontologia

3. Refinamento de Requisitos e testes

4. Construção

- Desenvolver protótipo
 - Registrar as intenções em paráfrases
- Escalar
 - validar desempenho
- Povoar
 - Possivelmente, com recurso a “text mining” e outras técnicas de PLN

EdV 2006 - Ontologias

54

Construção



Desenvolver uma ontologia baseada em lógica = programar!

1. Escolha da linguagem de representação da ontologia
2. Obtenção ferramenta(s) desenvolvimento
3. Aquisição de conhecimento do domínio
4. Reutilização das ontologias existentes

Etapas do Desenvolvimento de uma Ontologia

5. Avaliação & Quality Assurance

- Face aos objectivos
- Criar testes para gestão da mudança
- Conceber testes de regressão e “sondas”

6. Monitorar uso e evolução

A construção de ontologias é um processo, não é um produto!

Engineering - Practical Advice

(Jeremy Rogers @ Manchester)

- Ontologies are conceptualisation of domains for use on computers **TO DO SOME TASK(S)**
 - Always ask “What’s it for?”
- Perfection and completeness will seduce you
 - Only build what’s useful
- Iterative process more important than static product
- In theory, classifiers make large ontologies possible
- In practice, field is full of confusions in language, notation, assumptions & goals
 - And your users won’t want to know about any of these

Ontologias na SW: Perspectivas

OWL Lite – Less is more

- Lógica de primeira ordem simples é suficiente para a grande maioria das aplicações.
- A grande maioria do conhecimento ontológico é SIMPLES.
- Linguagens simples de representação de ontologias permitem raciocínio eficiente e são mais fáceis de usar e suportar.
- **As ontologias mais simples são também as usadas de forma mais alargada.**

Problemas que subsistem

- De onde virão as ontologias?
 - Legações: Wordnet, esquemas de bases de dados, thesauri...
 - Geradas por aplicações de aprendizagem/PLN
- De onde virá o markup semântico?
- De onde virão as ferramentas?
- Como lidar com tantas ontologias? (ontology mapping problem)
 - Métodos de gestão de conhecimento
 - Projectos em E-science (ciências integrativas)