

Ontologias (I)

Linguagens de Anotação de Documentos

Informação Estruturada

- Temos visto tecnologias para estruturar conhecimento, em contraste com informação não-estruturada como textos em linguagem natural
- No entanto mesmo dados estruturados têm as suas limitações no que toca à partilha de informação, o que se tem tornado um problema na *web*
- Além da estrutura da informação (*sintaxe*) torna-se cada vez mais importante definir também a sua *semântica* associada aos dados

Limitações: XML

- Por exemplo, esquemas XML (DTD ou o mais avançado XSD)
- Muito orientado à estrutura física do ficheiro, usada para persistir a informação textualmente
- Parte da informação do modelo conceptual (semântica) perde-se ao codificar o DTD
- Ainda pode ser interpretado por humanos que percebam o domínio de aplicação, mas não por máquinas (a não ser que se codifique a semântica)

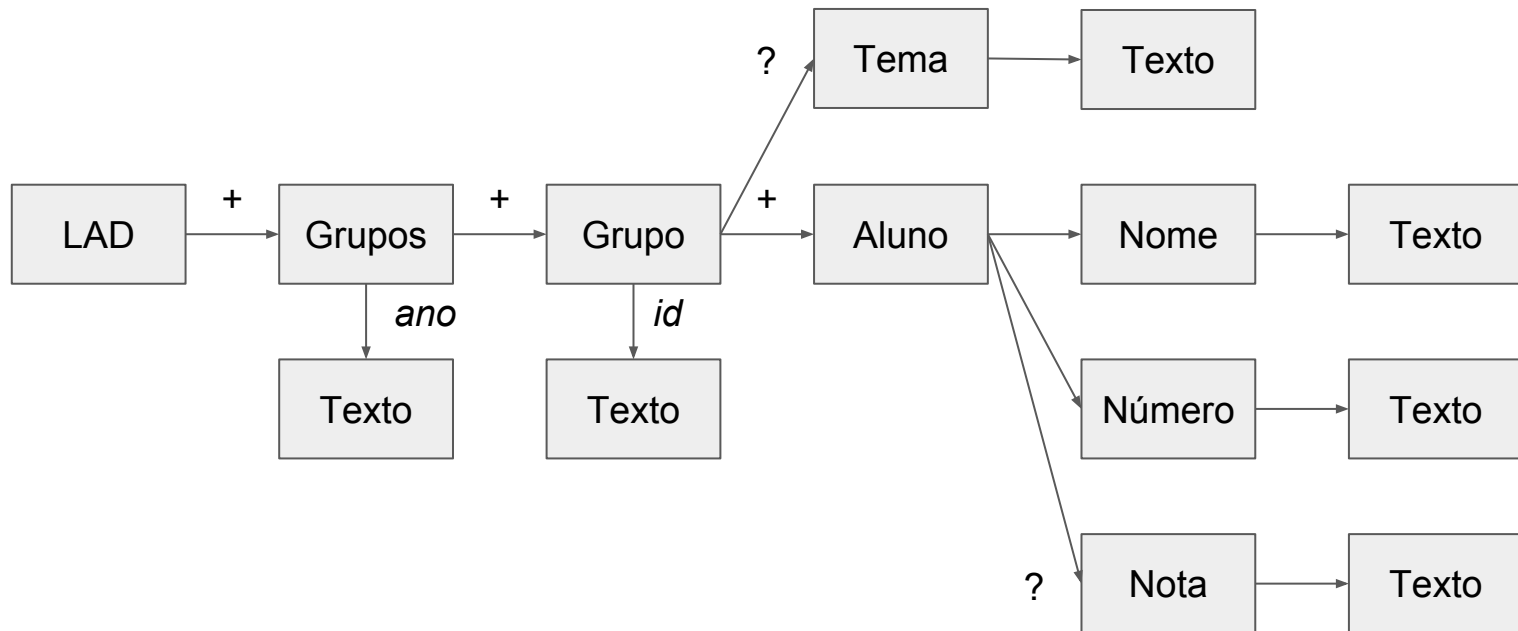
Limitações: XML

- Estrutura hierárquica em árvore é suficiente em muitos casos mas também tem limitações
- Apenas relações de “conteúdo”, do que é composto cada elemento (“*um grupo tem alunos e um tema*”)
- E se precisarmos de relações entre elementos de outra natureza?
- Podemos usar atributos com referências para outros elementos, mas não é natural numa árvore
- Além disso não têm semântica associada

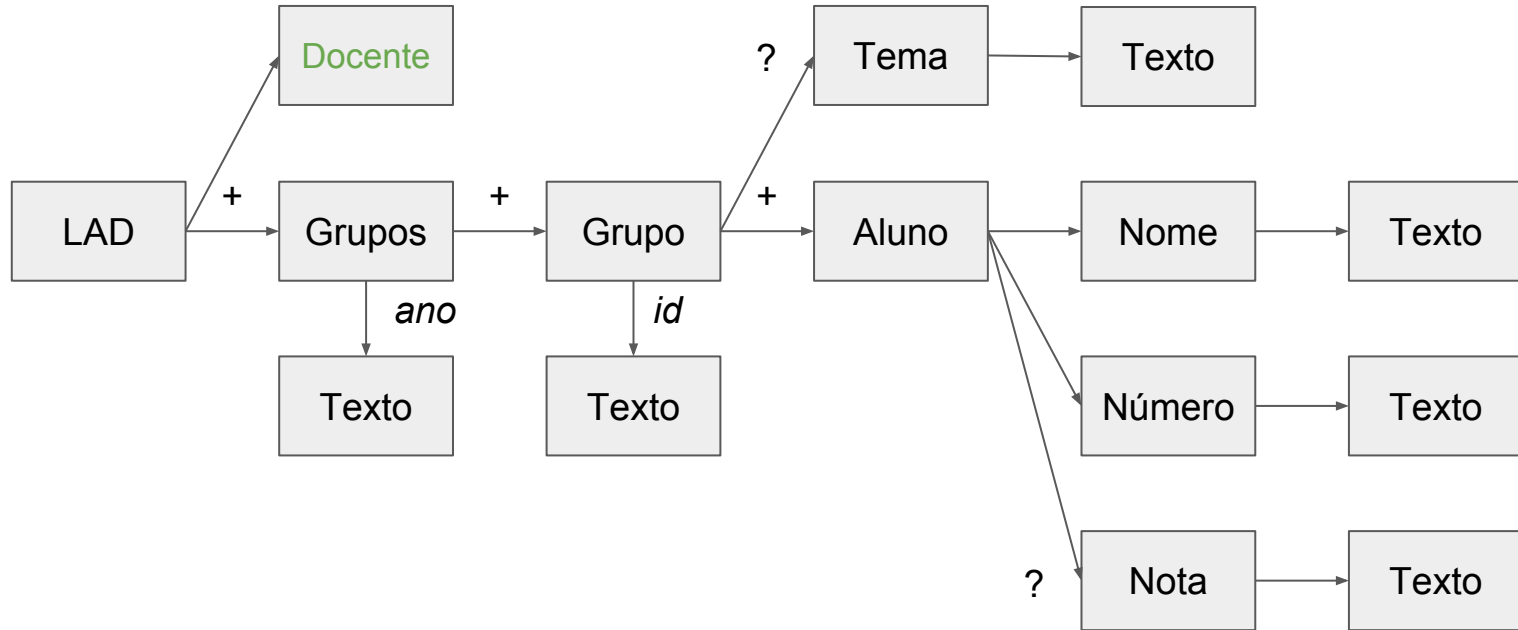
Limitações: XML

- Mais: consultas são sintáticas, tornam-se difíceis de gerir para a informação complexa e de grandes dimensões
- Não são lógicas, não podem usufruir totalmente de sistemas de derivação de conhecimento, provadores de teoremas automáticos

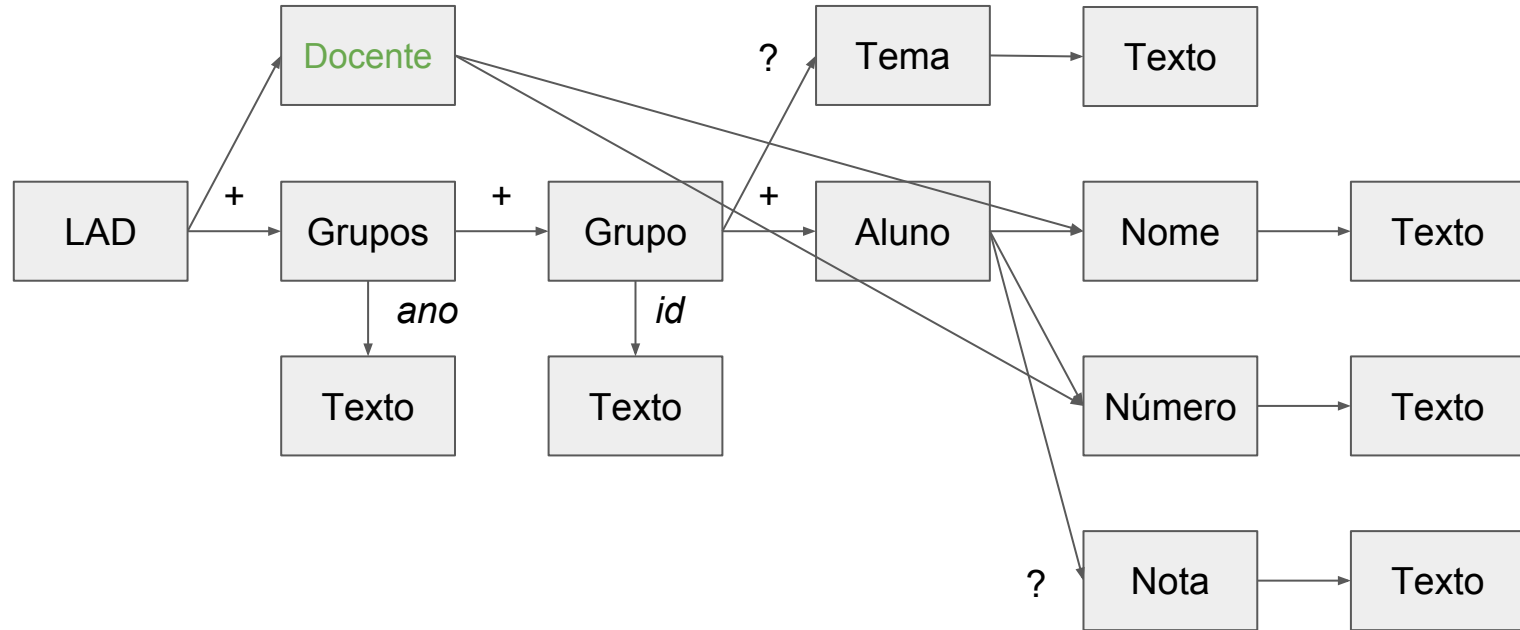
Limitações: XML



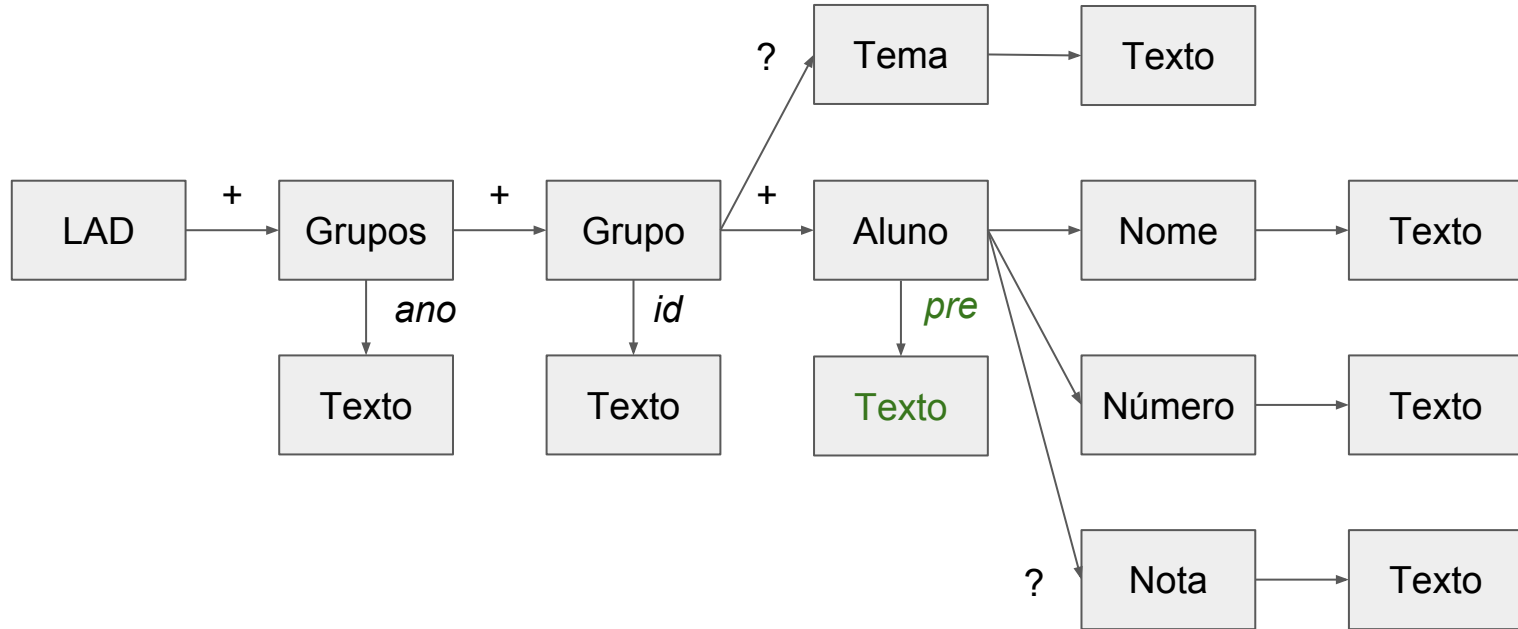
Limitações: XML



Limitações: XML



Limitações: XML



Limitações: XML

- Qual é a semelhança entre um *aluno* e um *docente*? Que propriedades são partilhadas?
- Se um aluno pode estar relacionado com vários grupos (mudanças) onde o inserir?
- Quais as restrições (cardinalidade) sobre os alunos de um grupo?

Para lá das árvores

- O conhecimento não é tão bem comportado como uma árvore XML
 - Relações de natureza diferente entre os vários elementos
 - Relações muitos-para-muitos (grafos)
 - Restrições adicionais sobre o que é considerado válido
 - ...
- Esta informação existe no nosso modelo conceptual, mas é muitas vezes descartada na concretização da estrutura de dados

Ontologias

- Um meio para estabelecer conceitos comuns para representar conhecimento num domínio de aplicação
- Definem conceitos, as suas propriedades e relações entre eles, assim como restrições adicionais
- O objetivo é a partilha de conhecimento de modo consistente
- Semântica é parte integrante, para que as ferramentas consigam interpretar os dados automaticamente
- As instâncias são tratadas ao mesmo nível que a ontologia

Ontologias

- As relações entre os elementos são arbitrárias, sem as restrições do XML
- Ênfase na semântica: o tipo de relação entre as entidades está explícita
- Conteúdo lógico, matemático, permite utilizar sistemas poderosos de derivação de conhecimento e impor noções avançadas de consistência
- Impacto na *performance* das ferramentas: e.g., processar XML é extremamente eficiente

Ontologias (Filosofia)

- Implementação do conceito filosófico de ontologia
- *“parte da filosofia que trata da natureza do Ser, ou seja, da realidade, das existências dos entes e das questões metafísicas em geral”* (wiktionary)
- Nas ciências da computação: define as entidades que existem, as suas propriedades e as suas relações
- Ao contrário dos filósofos, não estamos interessados se as entidades reflectem ou não essências fundamentais, mas em fixar conceitos do domínio

Ontologias (Ciências da Computação)

- Especificação de conceitos e a relação entre eles para um domínio de aplicação ou comunidade
- Frequentemente usadas implicitamente
 - Definição de taxonomias, sistemas de classificação, organização de informação...
- Como especificam conceitos relevantes ao domínio, permitem processar automaticamente consultas avançadas e derivar novo conhecimento
- Utilizadas formalmente pela primeira vez na inteligência artificial onde a captura de conhecimento é essencial para automatizar tarefas

Ontologias (Ciências da Computação)

- Essas formalizações eram majoritariamente usadas pelas aplicações internamente, por isso não havia necessidade de formalizar os conceitos
- Isto mudou quando com a explosão das aplicações *web*
- Tornou-se necessário conseguir exprimir ontologias formalmente, para que vários serviços possam usufruir de dados disponibilizados por outros

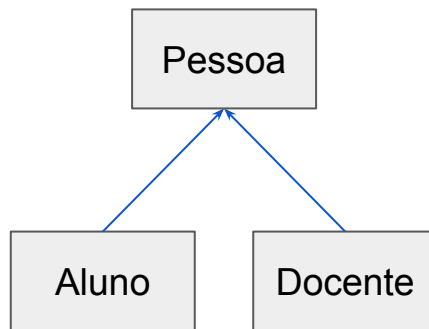
Componentes de uma Ontologia

- *Indivíduos / Objetos*: objetos base, concretos
- *Classes / conceitos*: conjuntos, tipos de objetos
- *Propriedades / Atributos*: propriedades dos objetos e das classes
- *Relações*: relacionam objetos e classes
- *Axiomas / Restrições*: regras impostas nos objetos

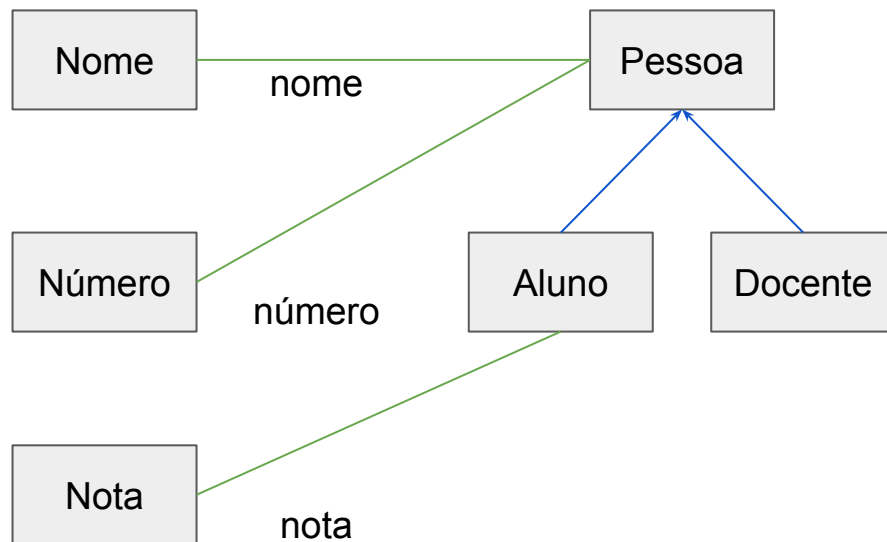
Componentes de uma Ontologia

- Na base duma ontologia estão classes ligadas entre si que definem a estrutura do conhecimento
 - De modo abstrato, uma taxonomia, vocabulário estruturado
- Passamos depois a introduzir relações adicionais e restrições à sua forma
 - Isto é aquilo que se chama da ontologia
- As instâncias obedecem então à estrutura e às restrições
 - Criamos uma base de conhecimento

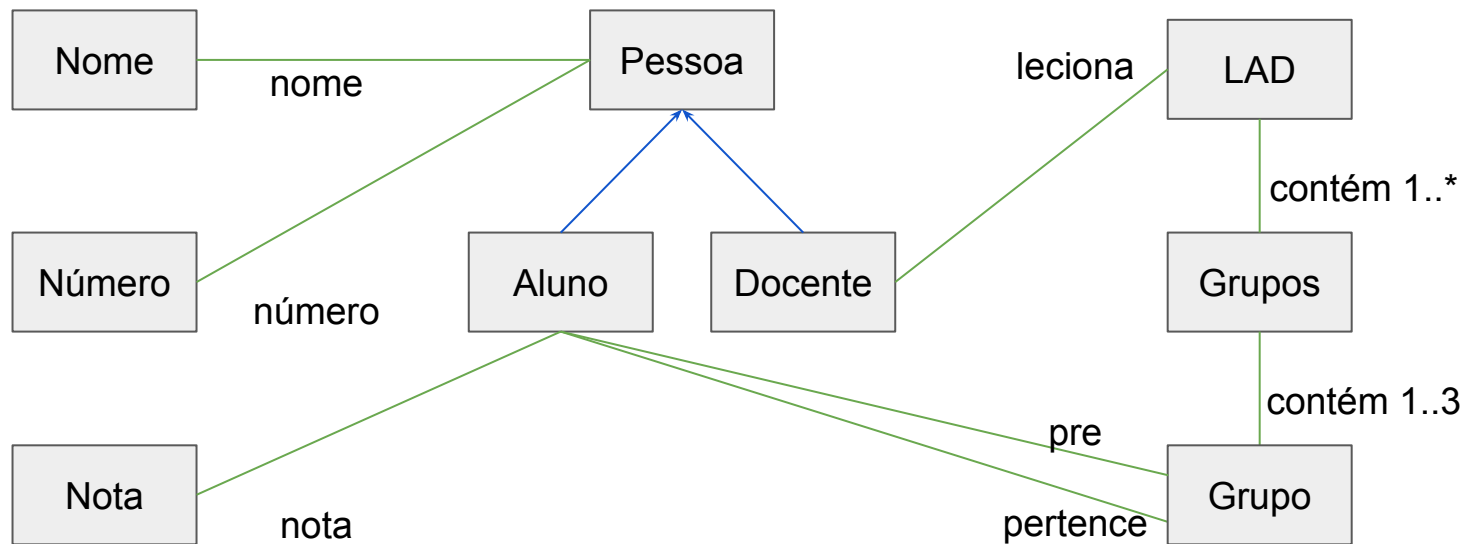
Exemplo: Grupos



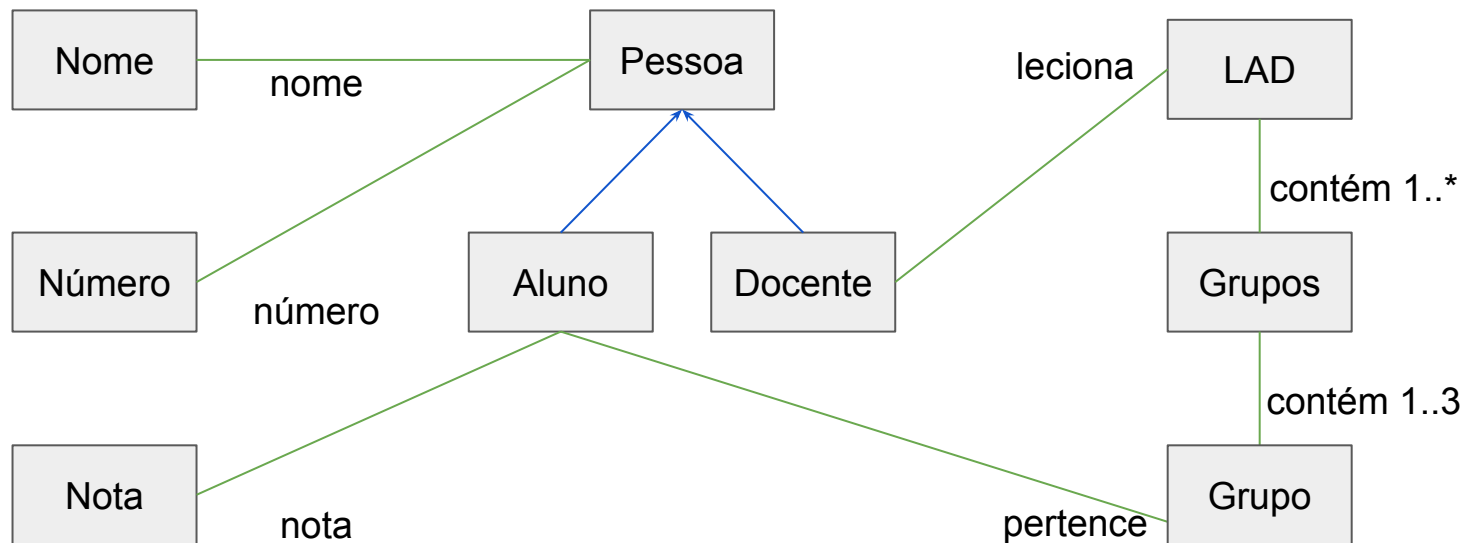
Exemplo: Grupos



Exemplo: Grupos



Exemplo: Grupos



Representação de Ontologias

- Uma ontologia em termos gerais não tem uma notação ou linguagem fixa, são conceitos abstratos, lógicos
- Também não têm um propósito concreto, são apenas modelos conceptuais
- Diferentes implementações têm diferente expressividade

Representação de Ontologias

- E.g., os esquemas DTD podem ser vistos como uma ontologia
 - As classes são os elementos
 - Atributos são os atributos dos elementos
 - Relações entre elementos e os seus filhos
- Bastante limitados, estrutura rígida
- Linguagens mais avançadas como o XSD são mais poderosas, mas ainda limitadas

Web Semântica

- A *web* hoje em dia é essencialmente sintática: recursos ligados entre si por hiper-ligações
- Os *browsers* interpretam anotações sobre como apresentar uma página, mas não a interpretam: essa parte compete a quem as desenvolve
- Visão proposta por Tim Gerners-Lee (inventor da *web*) para a evolução da *web* (sintática) atual
- Explorar os componentes comuns da WWW para promover a transmissão de dados que possam ser automaticamente interpretados por máquinas

Web Semântica

- E.g., um serviço de vendas pode usar uma ontologia para classificar os produtos e as relações entre eles
- Pode depois ser usada para responder a consultas avançadas do utilizador de forma automática
- Além disso, se a ontologia for partilhada pelos serviços, pode ser usada por, por exemplo, agregadores / comparadores de preços

Web Semântica

- E.g., um serviço noticioso pode usar uma ontologia para classificar os artigos e o seu conteúdo
- Permite categorizar artigos e facilitar a navegação entre outros artigos relacionados
- Além disso, se a ontologia for adotada por vários serviços, pode ser usada por agregadores de notícias

Web Semântica

- Tornou-se necessário criar uma tecnologia que permitisse criar ontologias poderosas, facilmente partilhadas pela *web*, e que pudessem ser automaticamente interpretadas pelos vários serviços

Resource Description Framework (RDF)

- Tecnologia proposta pela W3C para descrever recursos no contexto da *web* semântica
- Modelos conceituais contruídos através de frases *sujeito - predicado - objeto*
 - todos os alunos são pessoas: aluno - sub-classe - pessoa
 - o josé é um aluno: josé - tipo - aluno
- Os elementos do triplo são recursos identificados unicamente (URIs)
 - aluno, pessoa, josé, tipo, sub-classe, ...
- Serializados e transmitidos como XML

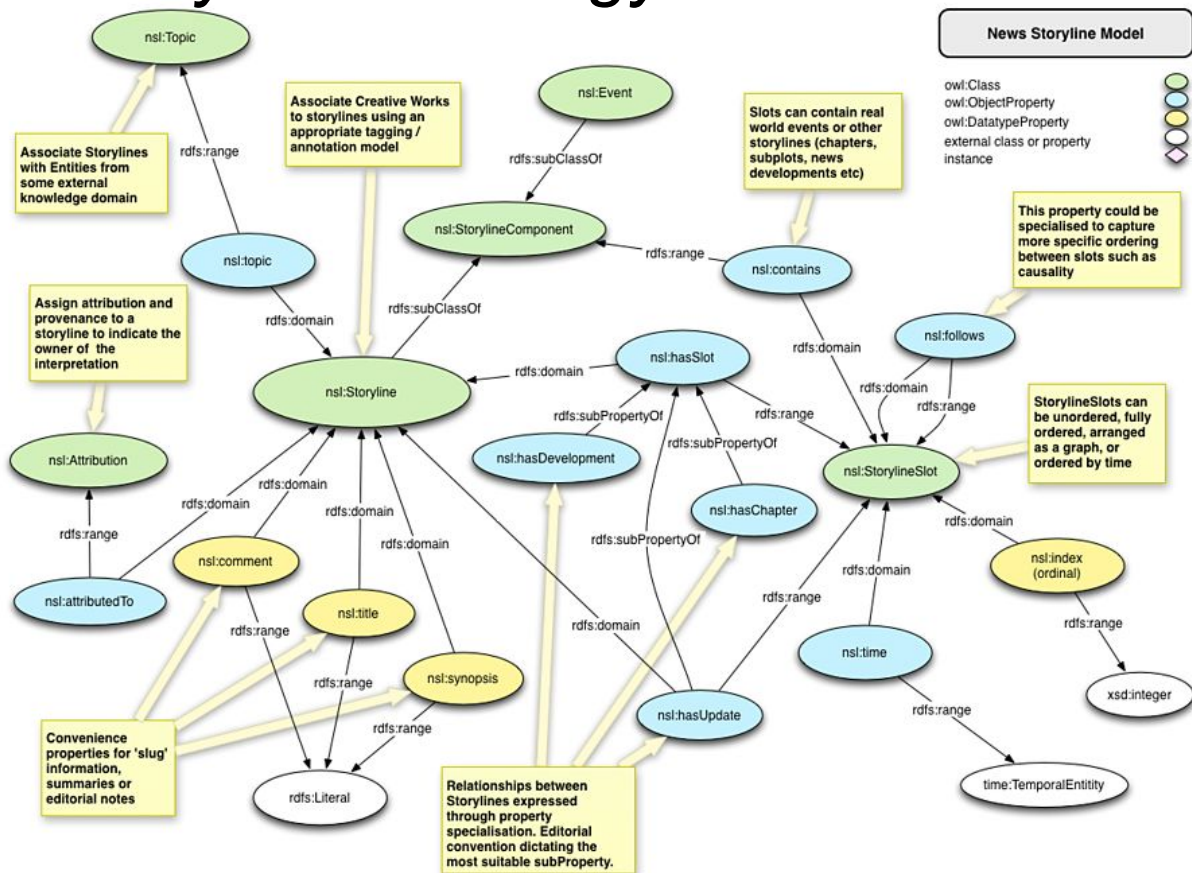
RDF Schema (RDFS)

- O RDF define apenas regras para criar frases sobre os recursos disponíveis
- Para representar ontologias, é necessário identificar alguns recursos com os conceitos básicos vistos atrás (classes, propriedades, sub-classe, tipo, ...)
- O RDFS disponibiliza o vocabulário para lidar com estes conceitos básicos (e.g., a simples ontologia de grupos)
- Semântica, o significado formal dos recursos, é ainda ambígua

Web Ontology Language (OWL)

- Última camada da *web* semântica, introduz o vocabulário necessário para criar ontologias complexas e robustas
- Construída sobre as tecnologias existentes (XML, RDF, RDFS, URIs, ...)
- Têm um significado formal matemático, que permite automaticamente testar a validade dos dados e extrair nova informação
- Suficientemente expressiva para a maioria dos domínios

Exemplo: *Storyline Ontology*



Take-home Lesson

- As linguagens para estruturar informação que temos vindo a ver são extremamente úteis mas também têm limitações
- São maioritariamente sintáticas sendo que a semântica tem que ser codificada nos sistemas que vão processar a informação
- As ontologias são modelos conceptuais que abstraem esse tipo de estruturas, focando-se em informação semântica
- Isto é particularmente relevante na *web* onde vários serviços independentes partilham e interpretam informação de origens diferentes