



Representação do Conhecimento

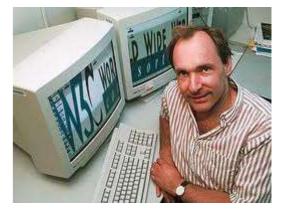
Introdução à Web Semântica

Tim Berners-Lee



Inventou a WWW em 1989

 O objetivo era oferecer acesso rápido e online a relatórios técnicos e outros documentos no CERN



A sua visão para a WWW

UA

"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A Semantic Web, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The intelligent agents people have touted for ages will finally materialize."
(Tim Berners-Lee, 1999)

Web Semântica



3

Definição

 A web semântica consiste numa "web de dados" que facilita a compreensão, pelas máquinas, do significado (semântica) da informação na web.

Na prática

- Esta estende a rede de páginas web,
 compreensíveis pelos humanos, inserindo
 metadados, compreensíveis pelas máquinas,
 acerca das páginas e das relações entre elas.
- O que pode permitir que agentes autónomos acedam à web de forma mais inteligente e executem tarefas pelos utilizadores.

Exemplo



- Organização de uma viagem consulta na web
 - 1. Transporte
 - Avião?; Comboio?; Rápida?; Barata?; Sem escalas? ...
 - 2. Hotel
 - Bom?; Barato?; Perto dos locais a visitar? ...
 - 3. Aluguer de Carro
 - Rápido?; Barato?; Confortável? ...
 - 4. Outras informações
 - Tempo; Restaurantes; Mapas; ...
 - 5. (instala-se o cansaço) ...

Qual é o problema?



- Temos de consultar uma quantidade de sites, que são todos diferentes:
 - no estilo; no propósito; na linguagem; etc.
- Temos de integrar, mentalmente, a informação disponível nesses sites, por forma a conseguirmos os nossos objetivos
- Isto pode ser um longo, aborrecido e cansativo processo

Qual é o problema? (ii)



- As páginas visualizadas, são meras pontas dos respetivos icebergues:
 - Os verdadeiros dados encontram-se escondidos algures em bases de dados, ficheiros XML, folhas excel, etc ...
 - Apenas vemos o que os designers do site permitem
- Alguns sites especializados, como o Expedia ou TripAdvisor, tentam fazer um pouco mais por nós
- E o que pensar acerca da enorme quantidade de informação dispersa por toda a web:
 - redes sociais; dados científicos, etc.

O que gostaríamos de ter?



- Usar os dados na web da mesma forma que usamos documentos:
 - ligar os dados, independentemente da sua apresentação
 - usar os dados da forma que quisermos (apresentar, explorar, ...)
 - usar scripts, programas, agentes para interpretar os dados
- Objetivo:

UA

 Estender a atual web para uma web de dados e permitir que aplicações explorem os dados diretamente

O que gostaríamos de ter? (ii)



- . E acerca dos sites *mashup*?
 - Sites que usam e combinam dados, apresentação e funcionalidades de outras fontes para criar novos serviços
 - Exemplo: www.tripit.com
- Estes sites são forçados a utilizar muitas tarefas ad hoc, porque:
 - as fontes expõem os seus dados através de diferentes web services
 - cada um, com diferentes APIs, lógicas e estruturas
 - são forçados a "reinventar a roda" muitas vezes, porque não existe um modo uniforme de fazer as coisas.

. Objetivo:

 Estender a atual web, de forma padronizada, para obter uma web de dados

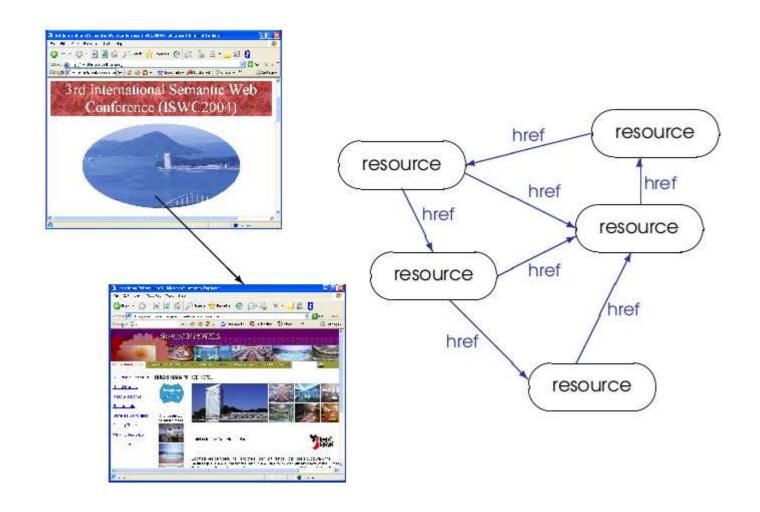
Web Atual



- As pessoas criam documentos diferentes;
- São atribuídos endereços (URIs) aos documentos para poderem ser acessíveis na web;
- Outros descobrem o site e ligam ao mesmo;
- Quantas mais ligações existem para esse site, mais importante e conhecido este se torna.
 - Parâmetro importante para oferecer resultados de pesquisa por alguns motores de busca.

Web Atual: Web Sintática





Web Sintática



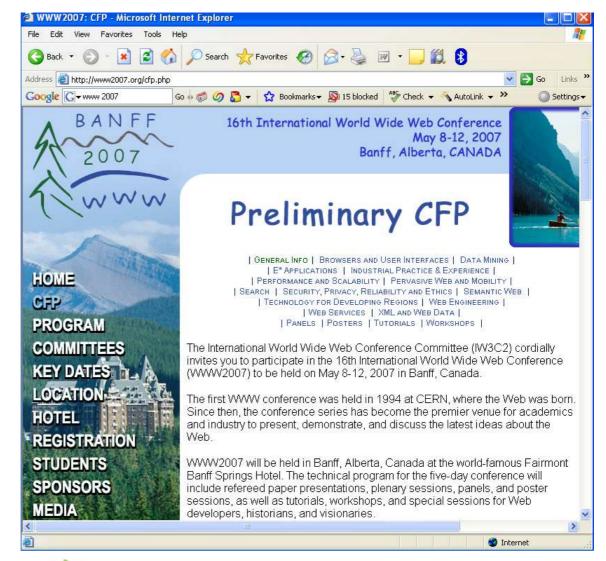
- Trata-se da web na qual:
 - os computadores fazem a apresentação
 - . a parte mais fácil
 - as pessoas estabelecem a ligação e fazem interpretação
 - . a parte mais difícil
- Porque não colocar os computadores a fazerem mais da parte mais difícil?
 - os termos utilizados nos dados possuem muitas vezes ambiguidade no seu significado
 - é necessário que esses termos sejam contextualizados
- Nesta web, os humanos são uma parte fulcral da equação
 - A semântica é facilmente acessível a estes, mas não aos computadores

Web Sintática: problema?



12

 Considere-se esta simples página web.



Web Sintática: problema? (ii)



Que informação pode ser vista?

WWW2007

The sixteenth International World Wide Web Conference

May 8-12, 2007

Banff, Alberta, Canada

Preliminary Call for Papers

The International World Wide Web Conference Committee (IW3C2) cordially invites you to participate in the 16th International World Wide Web Conference (WWW2007) to be held on May 8-12, 2007 in Banff, Canada.

The first WWW conference was held in 1994 at CERN, where the Web was born. Since then, the conference series has become the premier venue for academics and industry to present, demonstrate, and discuss the latest ideas about the Web.

WWW2007 will be held in Banff, Alberta, Canada at the world-famous Fairmont Banff Springs Hotel. The technical program for the five-day conference will include refereed paper presentations, plenary sessions, panels, and poster sessions, as well as tutorials, workshops, and special sessions for Web developers, historians, and visionaries.

Web Sintática: problema? (iii)



Que informação o computador vê?

```
******X
★★* ▲☆I▼**■▼☆ ☆■▼*□■☆▼☆□■☆◆ ★□□●☆ ★☆☆* ★≉� ÷□■☆*□★■**
** X 120000 00 00 00 X
*☆* ☆■▼*□■☆▼*□■☆● ★□□●☆ ★☆☆* ★*② ☆□■☆*□*■** ☆□○○☆▼▼** →☆★/☆・◆□
*□□**�◆●□*■◆*▼*▲ □□◆ ▼□ □�□▼***□▼* *■ ▼※* ◎★▼※ ☆■▼*□■☆▼*□■☆●
*※※ ②ぶ□▲▼ *** ※□■②※□※■※※ ▶録▲ ※※●② ※■ ◎●●● ◇ 桑▼ ・・・・・ ◇ ●※※□※ ▼※※ ▼※◎ ▶録▲
○□□□◎ **■** ▼⊕*■∅ ▼⊕* *□□⊕*□*■** ▲*□**▲ ⊕☆ ○**□○* ▼⊕* □□*○:**□
◇*■◆* ☆□□ ☆*☆☆*○**▲ ☆■* *■*◆▲▼□□▼□□□*▲*■▼め **○□■▲▼□☆▼*あ ☆■* **▲∠□
*◆▲▲ ▼光米 ●森▼米▲▼ 宋宋朱森▲ 森の□◆▼ ▼光米 ★米の⑤ ★★★•◆②②X ♪★●● ○米 光米●忠 宋田 十森田会会员
◇●◎*□▼炎炎→炎■炎虫炎 炎▼ ▼※*▶□□●虫△☆灸○□◆▲ ◆炎ホ□○□■▼ →炎■☆☆ 米□□▽■★▲ ★□▼*●◎≥
*※* ▼**※■**巻● □□□※□録○ ※□□ ▼※* ※*◆*△※蠡□*□■※*□*■** ▶★●● *■*●◆** □*※△
*□*** □爲□*□ □□*▲*■▼爲▼*□■▲※ □●*■爲□Ⅰ ▲*▲▲*□■▲※ □爲■*●▲※ 爲■* □□▲▼*□
▲*▲▲☆□■▲※ 炎▲ ▶★●● 炎▲ ▼◆▼□□☆灸●▲※ ▶□□★▲☆□□▲※ 炎■☆ ▲□※※☆灸● ▲*▲▲☆□■▲
�□□ ★*◆ ☆*◆*●□□*□▲炎 ※*▲▼□□*秦■▲炎 秦■☆ ◆*▲*□■秦□**▲◎△
```

Web Sintática: solução?



• Uso de XML?

```
<name>WWW2007
The sixteenth International World Wide Web Conference</name>
<date>May 8-12, 2007</date>
<location>Banff, Alberta, Canada</location>
<slogan>Preliminary Call for Papers</slogan>
```

<announcement>The International World Wide Web Conference Committee (IW3C2) cordially invites you to participate in the 16th International World Wide Web Conference (WWW2007) to be held on May 8-12, 2007 in Banff, Canada.

<statement>The first WWW conference was held in 1994 at CERN, where the Web was born. Since then, the conference series has become the premier venue for academics and industry to present, demonstrate, and discuss the latest ideas about the Web. WWW2007 will be held in Banff, Alberta, Canada at the world-famous Fairmont Banff Springs Hotel. The technical program for the five-day conference will include refereed paper presentations, plenary sessions, panels, and poster sessions, as well as tutorials, workshops, and special sessions for Web developers, historians, and visionaries.

Web Sintática: solução? (ii)



O computador continua a ver:

```
<■⑤*>****•◊
* 掛米 ▲羽▼米米■▼車 ☆■▼米□■@▼ホ□■@● *□□●※ *ホホ米 **② 中□■※米□ホ■米米</■@○*>
<■泰米□●▲/>▲□≉□泰☆ □□命 ●●泰・□□泰爾字○中●*□☆<|泰米□●▲>
< < □□□◆■××○×□▼> *÷* ☆■▼*□■◆▼*□□◆● *□□●* **:** **○ ·□□■÷*□*■**□*■**□ ·□□○ Z□
☆■▼本□■☆▼ホ□■韓● ★□□●忠 ★北宋米 ★本○ 廿□■忠未□未■未来 → ★★★◆◆ ※ ★ X □ ▼□ ○本 朱本● 中 □■ ★ ◆Ⅰ
X上心→◆多→○○○X 中国 十春日中華 中春日中華 中春日 中華 日本中 | < | 春日日 | 中華 | 日本中 | |
▼光本 ★本〇 ▶6 ▲ ○□□■◎ 米字■本本 ▼光本■前 ▼光本 水□■光本□本■本本 ▲本□字本▲ 光春▲ ○本本□○本 ▼光本
□□≉○★★□ ◆★■◆★ ☆□□ ☆☆☆☆★○☆☆▲ ☆■☆ ☆■☆ ★■☆◆▲▼□Ⅰ▼□ □□★▲★■▼∅ ☆★○□■▲▼□☆▼☆∅
╬■☆ ☆☆▲★◆▲▲ ▼☆★ ●☆▼★▲▼ ☆☆★☆▲ ☆☆□◆▼ ▼☆★ ★★◆◆ ※※★★ ◆ ※※★★ ★■☆ ☆★★★★■☆ ★★★
· 中華圖会会員 ◇●○本□▼泰島・中華圖泰安泰 泰▼ ▼米米 ▶□□●☆△会泰○□◆▲ ◆泰宋□○□■▼ · 中華圖会会 米□□水圖米▲
★□▼*●◎◎◎◎ </▲▼鈴▼*○*■▼>
```

Web Sintática: solução? (iii)

UA



17

 Na melhor das hipótese, o computador poderá ver:

```
<name>******
米掛米 ▲樹▼米米■▼東 ☆■▼米□■☆▼米□■☆● 米□□●# 米米串# ★米袋 ☆□■串米□米■米米</name>
<date>**! XAA>>** ** ** X</date>
<slogan> ☆□*•*○*■◆□! ÷◆•• ∜□□ ☆◆□*□▲</slogan>
<announcement>*#* ☆■▼*□■☆▼*□■☆● *□□●* *#** **◎ +□■●*□*■** +□□○△
○中▼▼※※ → ☆※/→·◆□ ※□□※☆☆●●□ ☆■◆☆▼※▲ □◆ ▼□ □□□□▼☆※☆□□▼※ ☆■ ▼→※ ◎★▼→
☆■▼本□■☆▼ホ□■☆● ★□□●忠 ★忠忠孝 ★朱〇 む□■忠孝□孝圖孝本 子★★★●◆彡彡 X回 ▼□ ○孝 養孝●忠 □■ ★☆□
▼※* ★*◆ D& ▲ ♥□□■% ※×■** ▼※*■※ ▼※* *□■**□*■** ▲*□**▲ ※☆ ▲ ♥**□○* ▼※*
□□*○**□ ◆*■◆* ②□□ 卷*卷虫*○**▲ 巻■* *■*◆▲▼□Ⅰ▼□ □□*▲*■▼め **○□■▲▼□巻▼※め
$■$ $\psi \text{$\alpha \text
中秦■会会总众●●本□▼委员中秦■委会委 泰▼▼※米▶□□●忠西会泰○□◆▲ ◆泰宋□○□■▼ 中秦■会会 米□□宋■米▲
*□▼*●>> </statement>
```

Web Sintática: falta algo?



- Pode-se responder às perguntas:
 - A que universidade pertence?
 - Onde está localizado?
 - Quantos docentes tem?
 - Quais as área de investigação?

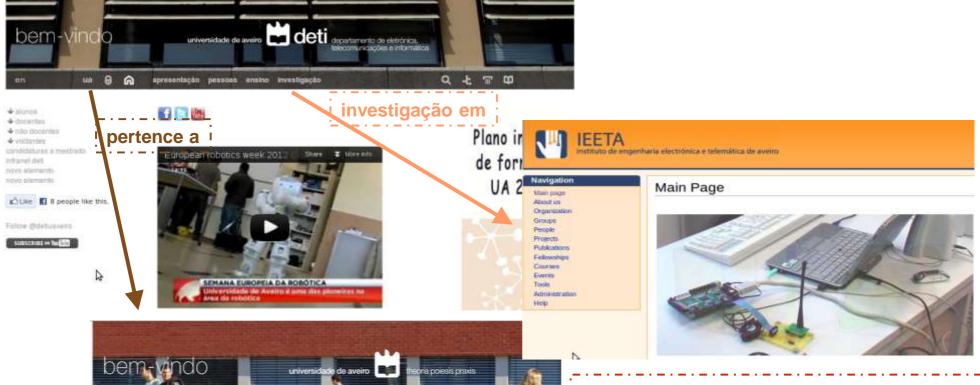


- O site possui a informação necessária para responder a estas questões, mas:
 - é <u>implícita</u>;

- conseguimos saber porque conhecemos o contexto
- As máquinas não conseguem inferir esta informação

Web Sintática: o que falta?





atunos recursos humanos antigos alunos investigadores comunicação social visitantes empresas research@ua international studenta studenta

jornal ua online

Mérito escolar dos estudantes da UA reconhecido com prémios

- é necessário adicionar informação aos links e aos dados
- esta informação liga dados a outros dados
- esta informação tem de ser compreendida pelo computador
- deve ser feito de forma padrão

Web de Dados

UA



20

- Para funcionar, é necessário:
 - publicar os dados na web
 - usar modos padrão em vez de abordagens ad hoc
 - atribuir endereços (URIs) aos dados
 - tornar possível a ligação a esses URIs, a partir de outras fontes de dados
 - as aplicações não devem ser forçadas a implementações personalizadas para acesso dos dados
 - em vez disso, usar abordagens padrão e genéricas
 - deixar então que o efeito de rede faça o seu trabalho ...

Integração de Dados



• Exemplo:

Bookstore dataset A (relational database)

ID	Author	Title	Publisher	Year
ISBN-0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Home Page
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID Publisher		City	
id_qpr	Harper Collins	London	

Bookstore dataset F (Excel sheet)

	Α	В	С	D
1	ID	Titre	Traducteur	Original
2	ISBN0-20203886682	Le Palais des miroirs	A13	ISBN-0-00-651409-X

6 ID Auteur 7 ISBN-0-00-651409-X A12

11 **Nom**12 Ghosh, Amitav
13 Besse, Christianne

UA

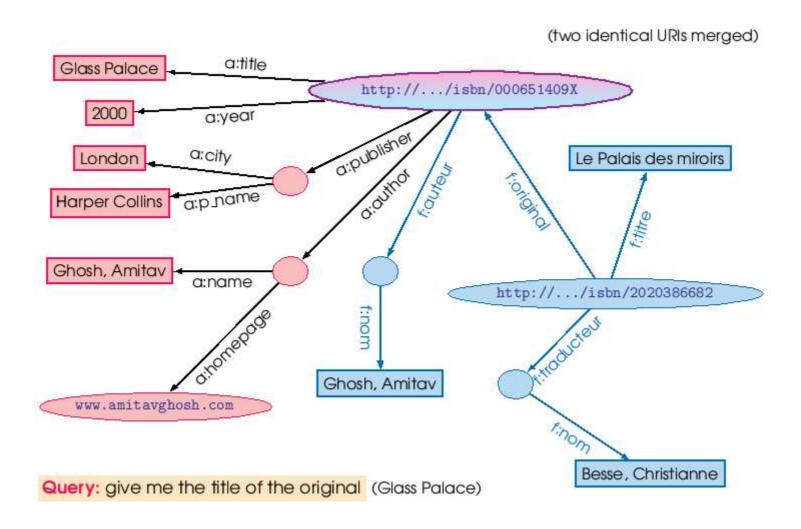
Query: what is the title of the original?

(no answer)

21

Modelo Abstrato de Dados





Adicionar Informação

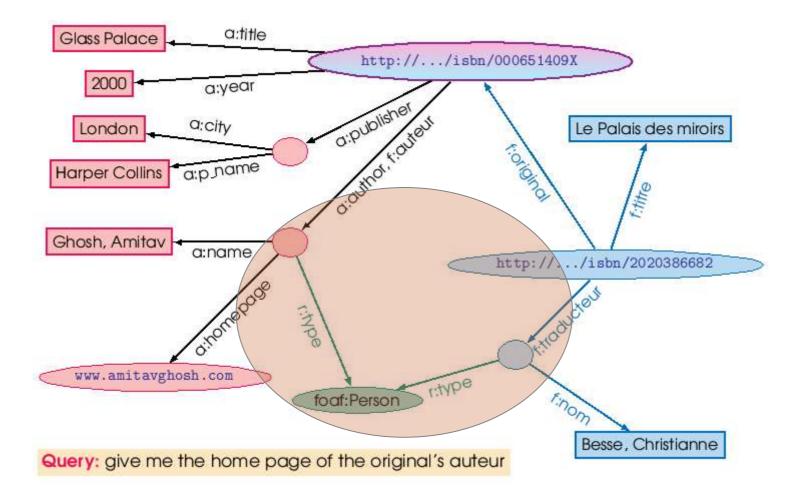


23

- A representação de dados da página anterior pode ser feita pela máquina,
 - contudo esta não sabe que "a:author" é a mesma coisa que "f:auteur"
- . Pode-se adicionar a seguinte informação:
 - a:author é o mesmo que f:auteur
 - referem-se ambos a "Person"

- O termo "Person" pode já ter sido definido pela comunidade web
- . Pode-se ainda declarar que uma "Person":
 - é univocamente identificada pelo nome e homepage
 - Pode ser usada como uma categoria para certos tipos de recursos
- . Isto vai permitir responder a mais questões, como:
 - query: dá-me a homepage do autor original
- O dataset pode ainda depois ser combinado com outras fontes como a wikipedia.

Extensão dos Dados Integrados



Resultado



25

- . Foram combinados diferentes datasets que:
 - estão algures na web
 - encontram-se em diferentes formatos (mysql, excel, html, etc.)
 - Têm nomes diferentes para as relações
- A combinação dos dados foi possível porque os URIs eram idênticos
- Foi possível adicionar informação extra, usando terminologia comum produzida pela comunidade
- Como resultado, podem ser encontradas novas relações
- Pode ainda tornar-se mais poderoso, se for adicionado conhecimento extra, como:
 - uma classificação total dos vários tipos de dados bibliográficos
 - informação geográfica
 - etc.

O que é a Web Semântica?

UA



26

- A web semântica pode ser vista como uma coleção de tecnologias padrão para conseguir a Web de Dados
- O exemplo anterior mostra que é necessário:
 - 1. linguagens formais, compreendidas pelas máquinas, para descrever os dados e suas ligações
 - 2. regras formais que permitam às máquinas a extração de informação dos dados (classificar, pesquisar, etc)
 - 3. tecnologias correspondentes e ferramentas eficientes
 - 4. e à parte, alguém tem de criar ontologias, nessas linguagens, para descrever os vários tipos de dados

Semântica?



Semântica

- do grego: semantiké →a arte da significação
- refere-se aos aspetos do significado expresso numa linguagem, código ou outra forma de representação
- Por outras palavras:
 - A semântica refere-se aos significados atribuídos aos símbolos e conjuntos de símbolos numa linguagem

. Questão:

- É difícil explicar o significado de "um fino"
 - a um humano?
 - a um computador?





Ontologia?



28

Na filosofia, ontologia consiste no estudo do ser ou existência.

 Tem por objetivo saber que entidades e tipos de entidades existem:

- O que existe?

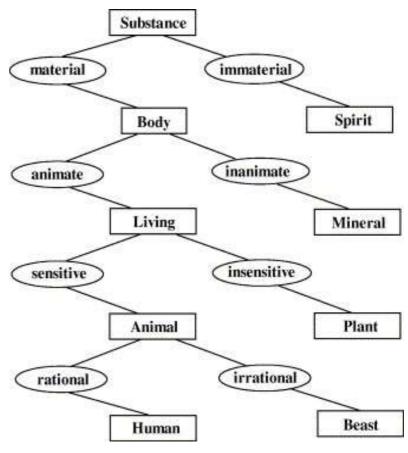
- A existência é uma propriedade?

O que é um objeto?

 Os objetos abstratos (não físicos) existem?

 Como devem as coisas serem classificadas?

Ontologia de Aristóteles



Ontologia nas Ciências da Computação



- Uma ontologia consiste num artefacto de engenharia:
 - é constituída por um vocabulário específico que é usado para descrever uma determinada realidade
 - e ainda,

- um conjunto de assunções explícitas acerca do suposto significado do vocabulário.
 - (geralmente, como os conceitos devem ser classificados)
- Assim, uma ontologia é uma especificação formal de um dado domínio de conhecimento, para:
 - o entendimento partilhado de um domínio;
 - e a manipulação pela máquina desse entendimento.