

A Ciência da Web: oportunidades de investigação

Reinaldo Ferreira ¹, Isabel Seruca ²

1) Planeta Virtual, Portugal

rdf@planetavirtual.pt

2) Universidade Portucalense, Portugal

Centro Algoritmi, Universidade do Minho, Guimarães, Portugal

iseruca@upt.pt

Resumo

A Web completou a sua segunda década de existência e, apesar da sua indiscutível presença e influência na sociedade, apresenta sinais de estar longe de estabilizar. Este artigo aborda o momento atual da evolução da Web e identifica algumas oportunidades para investigação e desenvolvimento.

Palavras chave: World Wide Web, Web 2.0, Web 3.0, Semantic Web, Linked Data

1. Introdução

Chegou-se a um momento da história onde a World Wide Web se transformou no mais importante meio de comunicação. Não há dúvidas quanto ao papel fundamental que tem hoje na vida dos jovens ou dos adultos de todas as idades, dos estudantes ou dos profissionais de todos os setores, das empresas ou organizações de todas as dimensões.

Foi na Suíça, há pouco mais de 20 anos, no final de 1990, que o inglês Tim Berners-Lee criou as primeiras aplicações precursoras dos atuais browsers e servidores web [Connolly 2000]. Atualmente acedem à Internet mais de 2700 milhões de pessoas (39% da população mundial), a partir de mais de 750 mil casas ligadas à Internet (cerca de 41% do total), 2100 milhões utilizando o seu telefone móvel [International Telecommunications Union 2013].

Se o crescimento da sua utilização é evidente, a Web está ainda longe de atingir um patamar de saturação. Em primeiro lugar, porque a assimetria entre países desenvolvidos e países em desenvolvimento deixa ainda muita margem de crescimento: acedem à Internet 75% dos Europeus, mas apenas 16% dos africanos [International Telecommunications Union 2013]. Em segundo lugar, porque a banda larga ainda não está disponível para a maioria dos utilizadores, seja em acessos fixos, seja em acessos móveis: a possibilidade de descarregar um vídeo de alta definição enquanto se assiste ao mesmo (a funcionalidade apelidada de vídeo clube nos pacotes de televisão sobre IP) só é possível com velocidades de acesso pelo menos na ordem dos 10Mbps [International Telecommunications Union 2010].

Neste trabalho é feita uma revisão da evolução da Web até ao momento atual, identificando-se as oportunidades de investigação e desenvolvimento da Ciência da Web em vários domínios.

2. As gerações Web

Os fundamentos básicos da World Wide Web referiam-se à teia de documentos e respetivas hiperligações construídos em HTML (Hyper Text Markup Language) [Connolly 2000], publicados em servidores acessíveis a especialistas, destinados apenas a serem consultados pelo público em geral. A primeira geração da Web, ou usando a metáfora geracional habitualmente utilizada no software, a Web 1.0, baseava-se na capacidade de expor o conhecimento, sem pretensão de o organizar, deixando-o ao dispor do critério cognitivo individual [Fuchs et al. 2010]. A dimensão enorme que a Web atingiu originou dificuldades na sua utilização, pelo que as principais aplicações que então surgiram foram não apenas o browser como também o motor de pesquisa, de que é exemplo o Google, atualmente líder de mercado e que responde a mais de 5000 milhões de pesquisas por dia (informação da empresa).

A Web 2.0 traduz essencialmente o conceito de comunidade, de livre adesão a interesses partilhados, uma Web Social como sociais são todas as interações assumidas entre indivíduos distintos, um passo na aproximação entre a Web e o mundo como um todo. Mas ainda assim, apenas uma Web de comunicação [Fuchs et al. 2010]. Sem pretensão de menosprezar o sucesso das redes sociais, por ventura a imagem mais atual da Internet, em muito devido a um fenómeno designado por Facebook, com 1000 milhões de utilizadores em todo o mundo, esta geração da Web contribuiu apenas para a sua democratização ou mesmo liberalização, transformando-a num espaço onde todos publicam para todos. Outros exemplos caracterizam bem esta geração, como o da rede social LinkedIn, dedicada a profissionais, com 200 milhões de utilizadores, ou do Technorati, o indexador de blogs, que indexa 1,3 milhões de blogs de um total estimado que

ultrapassará largamente os 100 milhões [Shadbolt & Berners-Lee 2008]. É possível, assim, considerar que esta geração atingiu atualmente a sua maturidade.

A Web 3.0 é apresentada como uma Web de colaboração, de interações individuais que contribuem para um bem comum [Fuchs et al. 2010]. Talvez o conceito de Wiki pudesse ter sido o que melhor traduz esta ideia, mas apesar dos mais de 4,2 milhões de artigos em inglês e 800 mil em português da Wikipedia (informação da empresa), devido à dificuldade em verificar a fonte de algumas informações, não é considerada a melhor expressão desse desígnio e, por isso, é relegada para a geração 2.0. No entanto, argumenta-se que merece uma classificação de 2.5 pelo papel fundamental que tem na rapidez de acesso a um nível de conhecimento talvez superficial, mas seguramente de uma transversalidade única. Para verificar este mérito, basta procurar alguns termos sobre os quais se possam ter dúvidas neste documento, começando pela própria definição de Web 2.0 [Wikipedia Users 2011] para confirmar que são abordados de forma bastante satisfatória, embora simultaneamente questionável.

Para que a Web 3.0 se concretize plenamente, espera-se a contribuição da Web Semântica e, novamente, de Tim Berners-Lee [Shannon 2006]. Os próximos anos serão testemunha do seu crescimento [Spivack 2010]. A Web 3.0 será para os dados o que foi a Web 1.0 para os documentos, interligando e dando significado às mais diversas fontes de dados, disponibilizando-as a cada vez mais dispositivos. A linguagem base da Web, o HTML, deu origem a várias evoluções, com destaque para o XML (eXtended Markup Language) e o RDF (Resource Description Framework), que permitem que os dados possam ser identificados, publicados, reutilizados e interligados com recurso aos mesmos endereços URI (Universal Resource Indicators), utilizados para identificar a localização de um documento. Na essência, se somos autores de um determinado documento, o RDF permite-nos criar um conjunto de dados que descrevem que um determinado “Indivíduo” é “autor de” um “Documento”. Se elaborarmos um pouco e pensarmos que um “Documento” foi “publicado” numa “Revista” e “tem como tema” a “Web 3.0”, podemos rapidamente pensar no potencial que podemos extrair de análises como “Todos os Autores de Documentos publicados sobre Web 3.0”.

A infância da Web 3.0 contribui, no entanto, para que alguns autores ainda coloquem dúvidas quanto à sua importância [Ahmed & Gerhard 2010] enquanto outros salientam o seu enorme potencial [Bizer, Heath, Idehen, & Berners-Lee 2008].

Um dos exemplos de utilização potencial de dados disponíveis na Web é a abertura de dados à utilização pública proporcionada pela iniciativa do governo britânico (disponível em <http://data.gov.uk>), que apresenta mais de 5600 conjuntos de dados sobre as mais variadas matérias [data.gov.uk 2011]. Para além da disponibilização desses conjuntos de dados, a

iniciativa publica ainda as aplicações que são criadas por empresas ou indivíduos para aceder facilmente a esses dados, nomeadamente uma aplicação que lista e localiza todas as escolas no Reino Unido.

Poder-se-ia, assim, inferir que a Web se encontra na sua geração 3.1, uma quantificação que reflete não apenas algumas realizações da disponibilização de dados abertos à utilização por outras aplicações, mas também o facto de a esmagadora maioria das aplicações estarem ainda por explorar. Mas que esta quantificação simplista, uma espécie de média estatística, não esconda a dispersão que a realidade apresenta, pois é possível seguramente verificar que inúmeras presenças web são meramente de primeira geração, como também é possível aceder a aplicações da nova geração completamente funcionais.

Finalmente, antecipa-se que no futuro surgirá a Web 4.0, provavelmente com a aplicação da Inteligência Artificial à utilização individual, servindo de suporte maciço e exclusivo a todo o software, como um WebOS (Sistema Operativo Web). Será provavelmente uma grande evolução a ocorrer nos próximos 10 anos e que transformará definitivamente a Web na rede omnipresente que começamos a detetar [Spivack 2010].

Situadas no caminho para esta geração, algumas aplicações começam a dar os seus resultados. Um desses exemplos é o Wolfram Alpha (<http://www.wolframalpha.com>). Se forem colocadas perguntas simples como “distance from porto to rome” ou “age of napoleon”, ao invés de receber uma lista de páginas web onde eventualmente se encontra a resposta, serão disponibilizadas as próprias respostas às questões.

Esta evolução da Web permite perceber que o seu crescimento começa também a ser “invisível”, pois a “Internet das pessoas” começa a dar lugar à “Internet das coisas”.

3. A Internet das coisas

Com cerca de 126 telefones móveis ativos em cada 100 habitantes, mais de 75% das pessoas com acesso à Internet, 68% das quais com acesso móvel [International Telecommunications Union 2013], os Europeus lideram globalmente na utilização da Internet. O cenário que já é possível experimentar é o de uma rede global de acesso móvel, com uma cobertura extensiva, um conjunto de aplicações de utilização generalizada, com acesso às mais variadas fontes de dados, com praticamente todos os indivíduos ligados. A única diferença entre o que temos hoje e o que teremos daqui a poucos anos estará na quantidade de dados a que poderemos aceder e na forma como serão utilizados.

Perante o aparecimento da “Internet das coisas” (Internet of Things), onde tudo pode estar ligado, onde os mais improváveis equipamentos comunicam permanentemente e geram ainda mais dados, é provável e esperado um novo crescimento de utilização da Web que ultrapassa a limitação natural do número de pessoas. Em “A Internet das Coisas – um Plano de Ação para a Europa”, a Comissão Europeia estabelece prioridades para a política europeia em relação ao desenvolvimento esperado da Internet [European Commission 2009]. Como exemplos atuais são citados a utilização de telemóveis para identificar produtos a partir da imagem obtida pela câmara incorporada, a utilização de etiquetagem individualizada para rastreio (utilizando Radio Frequency Identification, RFID), a implementação de contadores inteligentes no fornecimento de energia elétrica para monitorização instantânea ou os “objetos inteligentes” utilizados para aumentar a eficácia do ciclo produtivo e de logística.

4. Dados Universais

A Web Semântica propõe-se permitir que o significado dos dados seja conhecido, identificando assim as “coisas” referidas. Por um lado, a Web das coisas tomará forma tirando partido da Web dos documentos, tal como a conhecemos, que por sua vez tira partido da web dos computadores, que por sua vez tira partido da rede de cabos (e ondas) [Berners-Lee 2010]. Por outro lado, os dados não têm que estar presos no interior dos documentos e podem por si construir ligações que lhes permitam ser úteis sem o suporte documental, recorrendo a um “browser de dados” e dando origem ao conceito de “Linked Data” [Berners-Lee 2006].

O caminho para a disponibilização dos dados de forma universal passa por várias fases e Berners-Lee criou mesmo uma classificação com cinco estrelas que o caracteriza [Berners-Lee 2006].

A primeira fase passa por tornar os dados disponíveis na Internet, para utilização livre. Tendo em conta a habitual confusão entre as diferentes autorizações de utilização, por livre entenda-se mesmo sem qualquer restrição. Alguns desafios se colocam nesta matéria, como sejam os da utilização indesejável (espionagem, terrorismo) ou da utilização excessiva (sobrecarga).

A segunda fase consiste em disponibilizar esses dados de forma a que as máquinas os entendam. Um exemplo clássico será o de substituir a imagem de uma tabela por uma tabela no formato do Excel. Os desafios passarão então a ser estruturais, quer quanto à codificação da estrutura, quer quanto à própria qualidade da mesma.

A terceira fase será a de disponibilizar os mesmos dados numa estrutura não proprietária. No mesmo exemplo, poder-se-ia substituir a tabela de Excel por uma estrutura de dados simples, mas generalizada, como a CSV (Comma Separated Values).

A quarta fase consiste em adicionar a capacidade semântica, utilizando standards do W3C como o RDF e o SPARQL (acrónimo recursivo de Sparql Protocol And Rdf Query Language, a linguagem que permite extrair dados do formato RDF). Os dados publicados podem, assim, passar a ser referidos por terceiros, usando endereços que indicam não apenas a sua localização mas também a sua identidade.

Finalmente, a quinta fase consiste na ligação desses mesmos dados aos dados de terceiros, proporcionando contexto e validação. Cria-se, neste último patamar, um cenário de dados interligados reciprocamente que corporizaria as ambições da Web Semântica.

Ao designar por desafio este percurso, está a ter-se em consideração a necessidade de investigar a forma de integração das diversas e fechadas fontes de informação existentes [Berners-Lee, Hendler, Hall, Shadbolt, & Weitzner 2006].

É essencial identificar e caracterizar a Web Semântica na sua vertente tecnológica, pois tal como a Web se confunde com a Internet tendo em conta a sua importante dimensão, a tecnologia que suporta a Web Semântica também se confunde com as suas realizações ou aplicações.

A génese da Web Semântica confunde-se com a da própria Web, pois a proposta original de Tim Berners-Lee que levou à criação da *World Wide Web* referia diferentes tipos de ligações entre diferentes tipos de documentos. O W3C publicou a primeira recomendação para a representação genérica de meta dados em 1999, criando assim o RDF [W3C 2004a]. Conforme já referido, o RDF permite que os dados possam ser identificados, publicados, reutilizados e interligados com recurso aos mesmos endereços que se utilizam para identificar a localização de um documento e, por isso, está na base da concretização da Web dos dados. Concretamente, o RDF é uma linguagem para representação de modelos de dados que utiliza afirmações constituídas por triplos: um sujeito, um predicado e um objeto. Para uma melhor compreensão da base de sustentabilidade dos dados na Web Semântica, é importante perceberem-se os seus fundamentos e a sua estrutura, que aqui se pretendem apresentar.

Praticamente todos os indivíduos reconhecem um formato de dados com base numa tabela, onde cada coluna representa um campo ou propriedade e cada linha representa um registo ou um elemento. Considere-se o exemplo simples da Figura 1, uma tabela de indivíduos que inclui os respetivos nomes e datas de nascimento.

Indivíduos

Nome	Data de Nascimento
Antônio	13/09/1948
Beatriz	05/07/1972
Camilo	21/02/1976
Daniela	10/10/1941
Eduardo	29/06/1966

Figura 1: Modelo de Dados - Indivíduos

Considere-se agora uma segunda tabela, na Figura 2, onde se indicam documentos publicados por esses indivíduos.

Publicações

Nome	Data Nascimento
Antônio	13/09/1948
Beatriz	05/07/1972
Camilo	21/02/1976
Daniela	10/10/1941
Eduardo	29/06/1966

Nome	Documento	Data Publicação
Antônio	"What about RDF?"	11/09/1998
Beatriz	"RDF triples make the Web"	15/02/2004
Beatriz	"One tale, two Webs"	26/04/2008
Daniela	"Mining the Global Data Space"	17/06/2002
Eduardo	"Understanding OWL"	19/11/2009
Eduardo	"SPARQLing: Water"	13/08/2011

Figura 2: Modelo de Dados - Publicações

Estas representações são bastante comuns e intuitivas. O modelo Entidade-Relação de representação de informação toma por origem estas representações, onde se percebe a Entidade em causa em cada tabela e se definem as Relações entre as Entidades de forma a tornar o modelo de dados representativo. Neste caso, um Indivíduo pode publicar vários documentos. O objetivo desta descrição não é obviamente o de tentar explicar todos os conceitos envolvidos no modelo, mas antes de salientar as características fundamentais que o distinguem do RDF, pelo que não se entra em mais pormenores nem se discute porque não poderia um documento ser publicado por mais que um indivíduo.

Ao alterar esta estrutura por uma estrutura semelhante, podem antecipar-se várias consequências, conforme se pode ver na Figura 3.

Publicações

Entidade	Predicado	Objecto
António	nasceu em	13/09/1948
Beatriz	nasceu em	05/07/1972
Camilo	nasceu em	21/02/1976
Daniela	nasceu em	10/10/1941
Eduardo	nasceu em	29/06/1966
António	publicou	"What about RDF?"
Beatriz	Publicou	"RDF triples make the Web"
Beatriz	Publicou	"One tale, two Webs"
Daniela	Publicou	"Mining the Global Data Space"
Eduardo	Publicou	"Understanding OWL"
Eduardo	Publicou	"SPARQLing: Water"
"What about RDF?"	publicado em	11/09/1998
"RDF triples make the Web"	publicado em	15/02/2004
"One tale, two Webs"	publicado em	26/04/2008
"Mining the Global Data Space"	publicado em	17/06/2002
"Understanding OWL"	publicado em	19/11/2009
"SPARQLing: Water"	publicado em	13/08/2011

Figura 3: Modelo de Dados - Publicações, tabela única

A informação disponível ficou agora menos intuitiva. Não apenas um indivíduo tem agora mais dificuldade em analisar a informação disponível, como se torna mais difícil responder a perguntas simples, como por exemplo, quantos documentos foram publicados antes de 1 de Janeiro de 2000. O esforço computacional necessário para processar esta estrutura também aumentou na mesma proporção.

No entanto, a estrutura tornou-se mais flexível. Veja-se na Figura 4 o que seria necessário fazer para adicionar as disciplinas lecionadas pelos indivíduos ao primeiro cenário.

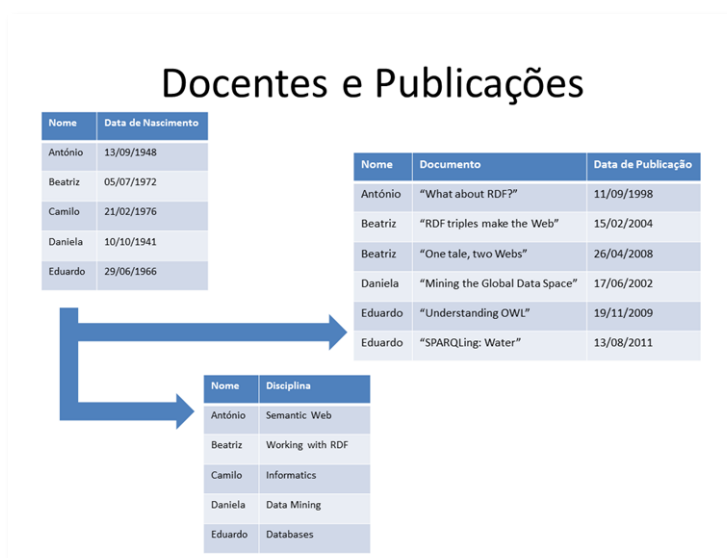


Figura 4: Modelo de Dados - Docentes e Publicações

Veja-se agora na Figura 5 o impacto da alteração no segundo cenário.

Docentes e Publicações

Entidade	Predicado	Objecto
António	nasceu em	13/09/1948
Beatriz	nasceu em	05/07/1972
Camilo	nasceu em	21/02/1976
Daniela	nasceu em	10/10/1941
Eduardo	nasceu em	29/06/1966
António	publicou	"What about RDF?"
Beatriz	Publicou	"RDF triples make the Web"
Beatriz	Publicou	"One tale, two Webs"
Daniela	Publicou	"Mining the Global Data Space"
Eduardo	Publicou	"Understanding OWL"
Eduardo	Publicou	"SPARQLing: Water"
"What about RDF?"	publicado em	11/09/1998
"RDF triples make the Web"	publicado em	15/02/2004
"One tale, two Webs"	publicado em	26/04/2008
"Mining the Global Data Space"	publicado em	17/06/2002
"Understanding OWL"	publicado em	19/11/2009
"SPARQLing: Water"	publicado em	13/08/2011
António	docente de	Semantic Web
Beatriz	docente de	Working with RDF
Camilo	docente de	Informatics
Daniela	docente de	Data Mining
Eduardo	docente de	Databases

Figura 5: Modelo de Dados - Docentes e Publicações - tabela única

Agora, a estrutura permanece inalterada e a simples adição de informação à estrutura permite estender o modelo. Aparentemente, a simplicidade da forma garante-lhe flexibilidade mas parece tornar o conteúdo cada vez menos claro.

Entretanto, o compromisso entre esforço computacional e flexibilidade tem vindo a pender em favor do segundo lado, tendo em conta a verificação da projeção ditada pela Lei de Moore, pois mais capacidade de computação disponível leva a menos preocupação com a simplificação ou otimização. Por outro lado, a economia tem beneficiado da inovação e deu origem a um ambiente de elevada competitividade onde flexibilidade e velocidade assumiram um papel por vezes mais importante que robustez e confiança. Pode assim questionar-se a validade económica atual dos modelos de normalização de dados em comparação com a flexibilidade e promessa de universalidade da Web Semântica [Segaran, Evans, & Taylor 2009].

Talvez a principal característica do RDF seja a identificação universal de todos os recursos por um *Universal Resource Identifier* (URI), fazendo uso generalizado da Web não apenas para obter informação acessível em *Universal Resource Locators* (URL) mas também para identificar qualquer coisa, seja ela possível de aceder eletronicamente ou não [Segaran et al. 2009; W3C & IETF 2001; W3C 2008]. No caso específico de *Linked Data*, são utilizados HTTP URI em detrimento de outras possibilidades, permitindo que qualquer cliente de HTTP possa obter uma descrição do recurso identificado pelo endereço [Heath & Bizer 2011]. Tal como no Modelo Entidade Relação se teria identificado cada indivíduo com um identificador único, como por exemplo com um número inteiro diferente para cada indivíduo, no RDF espera-se confiar nos URI como forma de identificação universal. Confiança que termina no ponto de poder aceitar-se que o mesmo indivíduo tenha mais que um URI desde que cada um

dos URI referencie os restantes como sendo indicadores do mesmo indivíduo, eliminando a duplicação e facilitando a extensibilidade.

O nosso exemplo passaria a ter então um formato um pouco distinto, como na Figura 6, em versão incompleta para melhor visualização.

Entidade	Predicado	Objecto
http://example.org/people/antonio	tem nome	António
http://example.org/people/beatriz	tem nome	Beatriz
http://example.org/people/camilo	tem nome	Camilo
http://example.org/people/daniela	tem nome	Daniela
http://example.org/people/eduardo	tem nome	Eduardo
http://example.org/people/antonio	nasceu em	13/09/1948
http://example.org/people/beatriz	nasceu em	05/07/1972
http://example.org/people/camilo	nasceu em	21/02/1976
http://example.org/people/daniela	nasceu em	10/10/1941
http://example.org/people/eduardo	nasceu em	29/06/1966
http://example.org/documents/what_about_rdf	tem titulo	"What about RDF?"
http://example.org/documents/rdf_triples_make_the_web	tem titulo	"RDF triples make the Web"
http://example.org/documents/one_tale_two_webs	tem titulo	"One tale, two Webs"
http://example.org/documents/minig_the_global_data_space	tem titulo	"Mining the Global Data Space"
http://example.org/documents/understanding_owl	tem titulo	"Understanding OWL"
http://example.org/documents/sparqling_water	tem titulo	"SPARQLing: Water"
http://example.org/people/antonio	publicou	http://example.org/documents/what_about_rdf
http://example.org/people/beatriz	Publicou	http://example.org/documents/rdf_triples_make_the_web
http://example.org/people/beatriz	Publicou	http://example.org/documents/one_tale_two_webs
http://example.org/people/daniela	Publicou	http://example.org/documents/minig_the_global_data_space
http://example.org/people/eduardo	Publicou	http://example.org/documents/understanding_owl
http://example.org/people/eduardo	Publicou	http://example.org/documents/sparqling_water
http://example.org/documents/what_about_rdf	publicado em	11/09/1998

Figura 6: Modelo de Dados - Docentes e Publicações - tabela única com URIs

Representando o mesmo exemplo em RDF, obter-se-ia uma estrutura como a representada na Figura 7, onde se representa apenas 20% da estrutura para melhor legibilidade.

```

Docentes e Publicações

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:example="http://example.org/rdf/example/1.0/">
  <rdf:Description rdf:about="http://example.org/people/antonio">
    <rdf:type rdf:resource="http://example.org/rdf/example/1.0/Person"/>
    <example:name>António</example:name>
    <example:birthdate>13/09/1948</example:birthdate>
    <example:publication>http://example.org/documents/what_about_rdf
    </example:publication>
    <example:lecturer>Semantic Web</example:lecturer>
  </rdf:Description>
  (...)
  <rdf:Description rdf:about="http://example.org/documents/what_about_rdf">
    <rdf:type rdf:resource="http://example.org/rdf/example/1.0/Document"/>
    <example:title>What about RDF?</example:title>
    <example:published>11/09/1998</example:published>
  </rdf:Description>
  (...)
</rdf:RDF>

```

Figura 7: Modelo de Dados - Docentes e Publicações - RDF

A representação do RDF pode tomar vários formatos, e, neste caso, o RDF/XML [W3C 2004b] usado acima é a sintaxe mais representativa. Outros formatos como N-Triples, N3, Turtle ou RDF/JSON permitem a representação do mesmo modelo de dados. Não se trata de um assunto

crítico neste trabalho, uma vez que todas as representações são facilmente dominadas com uma biblioteca específica e a transformação de uma representação noutra é um exercício dominado na maior parte das linguagens de programação.

Salienta-se ainda uma representação adicional, o RDFa [W3C 2012], “*RDF in attributes*”, um formato que pode ser embebido em HTML, permitindo que o mesmo documento inclua a parte legível ao indivíduo e a parte legível às máquinas. Este formato é muito utilizado na geração atual da Web, com exemplos proporcionados pelo Facebook, pelo Google ou pelo Yahoo, entre muitos outros.

5. A Ciência da Web

A breve história da Web não impede que se encontrem factos relevantes para o seu estudo e em conjunto com a sua evolução potencial originam uma área de estudo que se pode designar por Ciência da Web, uma combinação de Matemática, Física, Ciência dos Computadores, Psicologia, Ecologia, Sociologia, Direito, Ciências Políticas, Economia, entre outras disciplinas [Shadbolt & Berners-Lee 2008].

Na origem da Ciência da Web estiveram obviamente trabalhos no âmbito da Ciência dos Computadores, que contribuíram para a sua génese, bem como a sua componente Matemática. O trabalho dos fundadores da Google teve origem em Stanford e deu origem ao motor de busca mais utilizado em todo o mundo [Brin & Page 1998]. De forma mais discreta, um americano de origem húngara tem desenvolvido a teoria das redes livres de escala (scale-free networks), explicando a sua formação através da expansão contínua e da adesão preferencial [Barabasi 1999].

Atualmente, são muitas as áreas que podem contribuir para o desenvolvimento da Web. Aliás, o teor interdisciplinar desta nova ciência constitui por si só uma oportunidade, talvez abraçando mesmo temas no âmbito da Biologia ou Ecologia, entre outras [Shadbolt & Berners-Lee 2008]. A ciência é necessária para analisar e sintetizar e, frequentemente, a capacidade de analisar e perceber ultrapassa o potencial da criação pura. A Ciência da Web é, certamente, interdisciplinar o que proporciona a identificação de algumas das questões mais desafiantes e intrigantes do século XXI [Berners-Lee, Weitzner, et al. 2006].

6. Oportunidades de Investigação

Então o que pode um indivíduo fazer para participar neste desafio? A pergunta, sendo pertinente, tem uma resposta tão vasta quanto as áreas que envolve, conforme foi possível antecipar.

No âmbito do documento já referido “A Internet das Coisas – um Plano de Ação para a Europa”, a Comissão Europeia especifica, na Linha de ação 7, Investigação e desenvolvimento, que *“continuará a financiar, no âmbito do sétimo Programa-Quadro, projetos de investigação no domínio da Internet das Coisas, privilegiando, por um lado, domínios tecnológicos importantes, como a microeletrónica, os componentes sem silício, as tecnologias de "colheita" de energia ("harvesting technologies"), as tecnologias ubíquas de determinação da posição, as redes de sistemas inteligentes que comunicam sem fios, a semântica, as tecnologias que, logo na concepção, prevejam a proteção da privacidade e a segurança e o software que rivalize com o raciocínio humano, e, por outro, aplicações inéditas”*.

Neste mesmo documento, antecipa-se a iniciativa “Future Internet”, uma Parceria Público-Privada de integração dos esforços de Investigação e Desenvolvimento em Tecnologias de Informação e Comunicação em relação ao futuro da Internet. No âmbito desta iniciativa, mais de 150 projetos de investigação associaram-se no sentido de reforçar a competitividade europeia no mercado global [European Commission 2013].

Alguns autores referiram a necessidade de abordar problemas específicos e de encontrar soluções que permitam construir o futuro da Web. Na lista que se segue é feita uma síntese dessas opiniões sobre as oportunidades de investigação da Ciência da Web, em diversos domínios.

No campo da Ciência dos Computadores existem ainda inúmeras oportunidades de investigação, nomeadamente:

- Como analisar eficazmente uma Web constituída por repositórios interligados? [Berners-Lee, Hendler, et al. 2006]
- Como mapear e compatibilizar diferentes modelos de dados? [Berners-Lee, Hendler, et al. 2006]
- Como visualizar e navegar o enorme grafo de informação conectada na Web? [Berners-Lee, Hendler, et al. 2006]
- Quais são as características dos repositórios colaborativos de informação (p.e. Wikipedia)? [Shadbolt 2008]
- Como contribuem os indivíduos para os repositórios colaborativos? [Shadbolt 2008]
- Como suportar a inferência à escala da Web? [Shadbolt 2008]
- Como suportar a confiança e a proveniência na Web? [Shadbolt 2008]

- Como identificar partes não funcionais, atrofiadas ou redundantes da Web? [Shadbolt 2008]
- Que contribuições para o desenvolvimento da Web Semântica podem fornecer as Aplicações Web quanto à sua utilização, nomeadamente Web Site, Extranet, Intranet, Web App ou Web Service?
- Que tipos de Sistemas de Informação poderão evoluir consideravelmente com a adoção da Web Semântica?
- Que atividades na Web beneficiam com a introdução da Web Semântica?

No campo específico da Matemática, as seguintes oportunidades de aplicação são relevantes:

- Como distinguir a Web permanente da Web efémera? [Shadbolt 2008]
- Como usar representações de incerteza na Web? [Shadbolt 2008]
- Qual a estrutura da Web e como representá-la? [Shadbolt 2008]
- Como é afetada a estrutura aparente da Web em cada pesquisa realizada por um utilizador? [Shadbolt 2008]
- Como medir o nível de complexidade da Web? [Shadbolt 2008]

No campo do Direito, a interação entre a lei e a Web abre novas e vastas oportunidades. Trata-se de saber em que medida uma pode influenciar ou restringir a outra. Algumas possibilidades de trabalho podem incluir:

- Como adaptar o direito de autor aos conteúdos digitais? [Shadbolt & Berners-Lee 2008]
- Como definir os direitos de autor das obras digitais com contribuições individuais de reduzida dimensão? [Shadbolt & Berners-Lee 2008]
- Como incorporar regras de utilização nos conteúdos? [Shadbolt & Berners-Lee 2008]
- Qual o relacionamento entre as regras de utilização e a distribuição da informação? [Shadbolt & Berners-Lee 2008]
- Qual o impacto do Direito na formação da Web? [Shadbolt 2008]
- Deverá o Direito ser um catalisador ou um observador reativo da mudança económica, social e tecnológica? [Shadbolt 2008]
- Qual a aplicação do direito de autor a conteúdos que tendem a ser crescentemente gerados por computador? [Shadbolt 2008]

- Que tecnologias devem ser mantidas abertas e quais as consequências dessa decisão? [Shadbolt 2008]
- Até que ponto serão os prestadores de serviços os únicos representantes da política pública de proteção quanto a conteúdos ilegais ou danosos? [Shadbolt 2008]
- Quais os desafios de privacidade que se colocam a um ambiente de informação de complexidade crescente? [Shadbolt 2008]
- Quais as novas formas de regulação emergentes? [Shadbolt 2008]

No campo da Sociologia, a Web tem já provado que a sua influência é determinante, mas muitas possibilidades de estudo se mantêm disponíveis:

- Quais serão os métodos mais eficazes na determinação de confiança dos conteúdos? [Shadbolt & Berners-Lee 2008]
- Qual a relação entre confiança nos indivíduos e nos seus avatares? [Shadbolt & Berners-Lee 2008]
- Como realizar a integração digital de serviços sociais, médicos, financeiros ou educacionais? [Shadbolt & Berners-Lee 2008]
- Como prever o roubo de identidade, perseguição, agressão ou espionagem com base digital? [Shadbolt & Berners-Lee 2008]
- Como perceber o fenómeno social-tecnológico da Web? [Shadbolt 2008]
- Quais as relações entre a estrutura e a utilização da Web? [Shadbolt 2008]
- Quais os conceitos sociológicos e psicológicos a estudar para entender a forma como as pessoas usam a Web? [Shadbolt 2008]
- Quais as implicações da Web na diversidade? [Shadbolt 2008]
- Como se posiciona a Web em relação às redes de poder? [Shadbolt 2008]
- Qual o poder da Web? [Shadbolt 2008]
- Como poderá mudar a Web assim que novas populações lhe acedam? [Shadbolt 2008]

A Economia desempenha um papel preponderante no desenvolvimento da Web e apresenta diversas oportunidades de investigação:

- Quais as oportunidades económicas da Web 2.0? [Shadbolt 2008]

- Quais as implicações económicas da produção colaborativa de conteúdos em redes ad-hoc? [Shadbolt 2008]
- Quais as forças económicas que modelam a formação das redes sociais? [Shadbolt 2008]
- Quais são as propriedades fundamentais das redes sociais? [Shadbolt 2008]
- Qual a relação entre as estruturas social, económica e matemática da Web? [Shadbolt 2008]
- Quais os incentivos comerciais proporcionados pela Web? [Shadbolt 2008]
- Quais os incentivos industriais proporcionados pela Web? [Shadbolt 2008]
- Será a Web propensa à concentração, a um domínio de poucos e grandes intervenientes? [Shadbolt 2008]
- Quais serão as forças que permitem a coexistência entre grandes e pequenos intervenientes? [Shadbolt 2008]
- Quais os argumentos económicos a favor e contra as plataformas abertas na Web? [Shadbolt 2008]
- Deverá a política pública e económica influenciar a abertura de plataformas? [Shadbolt 2008]
- Que mecanismos económicos e sociais poderão melhorar o desempenho da Web? [Shadbolt 2008]
- Como pode a economia resolver problemas como os da pirataria, privacidade e identidade? [Shadbolt 2008]
- Que aplicações podem contribuir para uma adoção rápida e generalizada da Web Semântica nas Empresas?
- Que tipo de empresas poderá mais beneficiar com a adoção da Web Semântica?
- Como pode a Web Semântica contribuir para a Inovação?
- Que resultados da adoção da Web Semântica nas Empresas podem ser reaplicados na adoção pelos Estados e respetivos organismos?
- A utilização da Web Semântica nas Empresas contribuirá definitivamente para a sua disseminação?

- Como pode ser criada uma abordagem sistemática ao ativo Informação com base na Web Semântica?
- Quais as ferramentas necessárias para a medição adequada da utilização da Web Semântica em ambiente empresarial?

Finalmente, conjugando todas as disciplinas e tendo em conta a estrutura particular da Web, podem antecipar-se ainda as seguintes oportunidades de investigação:

- Como identificar normas sociais que podem promover a utilização da Web? [Shadbolt & Berners-Lee 2008]
- Como identificar tendências de fracionamento da Web? [Shadbolt & Berners-Lee 2008]
- Como pode ser promovida a inteligência coletiva de um ponto de vista tecnológico? [Shadbolt 2008]
- Qual o efeito da língua, raça, género e outras diferenças na criação de uma inteligência coletiva? [Shadbolt 2008]
- Quais são as razões socio económicas para a participação individual em empreendimentos coletivos? [Shadbolt 2008]
- Quais são os mecanismos de identificação psicológica de um indivíduo com uma comunidade? [Shadbolt 2008]
- Qual deve ser o papel das entidades públicas no processo de criação colaborativa? [Shadbolt 2008]
- Qual a relação entre a abertura da Web e a inovação? [Shadbolt 2008]
- Serão abertura da informação e requisitos de segurança compatíveis? [Shadbolt 2008]
- Qual o processo para avaliar um modelo de publicação em relação a um modelo de negócio proprietário? [Shadbolt 2008]
- Como identificar padrões de comportamento na Web? [Shadbolt 2008]
- Poderão a incerteza, a insegurança e a invasão de privacidade acabar com a Web? [Shadbolt 2008]
- Poderão os utilizadores encontrar formas de se protegerem quanto à incapacidade da Web o fazer por eles? [Shadbolt 2008]
- Como podemos continuar a pesquisar e interrogar a Web tendo em conta o seu crescimento? [Shadbolt 2008]

- Como representar e suportar o contexto da informação semântica? [Shadbolt 2008]
- Como utilizar a Web como fonte de informação para a aprendizagem? [Hall & O'Hara 2008]

7. Conclusão

A Web oferece atualmente um enorme potencial de investigação científica, quer porque se afastou da sua juventude e propósito iniciais, quer porque está a assimilar a transformação que provocou na sociedade. As oportunidades de investigação nas mais variadas áreas de conhecimento permitem antever alguns anos de estudo e experimentação entusiástica.

Este trabalho serviu de suporte à identificação de oportunidades de investigação no campo da Web Semântica e que são desenvolvidas em [Ferreira & Seruca 2013].

8. Referências

- Ahmed, Z., & Gerhard, D. (2010). Role of Ontology in Semantic Web Development. *Knowledge Management*.
- Barabasi, A. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509–512. doi:10.1126/science.286.5439.509
- Berners-Lee, T. (2006). Linked Data - Design Issues. *W3C*. Retrieved February 06, 2011, from <http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2010). Abstractions in Web architecture - Design Issues. *Design Issues*. Retrieved December 22, 2010, from <http://www.w3.org/DesignIssues/Abstractions.html>
- Berners-Lee, T., Hendler, J., Hall, W., Shadbolt, N., & Weitzner, D. J. (2006). Computer science. Creating a science of the Web. *Science*, 313(5788), 769–771.
- Berners-Lee, T., Weitzner, D. J., Hall, W., O'Hara, K., Hendler, J. a., & Shadbolt, N. (2006). A Framework for Web Science. *Foundations and Trends® in Web Science*, 1(1), 1–130. doi:10.1561/1800000001
- Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. (2008). Linked data on the web (LDOW2008). In *WWW2008 Workshop on Linked Data on the Web* (Vol. 2008, pp. 1265–1266). ACM New York, NY, USA. doi:10.1145/1367497.1367760

- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. In *Proceedings of the 7th International World Wide Web Conference* (Vol. 30, pp. 107–117). doi:10.1016/S0169-7552(98)00110-X
- Connolly, D. (2000). A Little History of the World Wide Web. W3C. Retrieved February 05, 2011, from <http://www.w3.org/History.html>
- data.gov.uk. (2011). data.gov.uk | Opening up government. Retrieved February 12, 2011, from <http://data.gov.uk/>
- European Commission. (2009). *Internet of Things — An action plan for Europe*. Europe.
- European Commission. (2013). Future Internet - Internet-enabled innovation in Europe. Retrieved from <http://www.fi-ppp.eu/>
- Ferreira, R., & Seruca, I. (2013). Modelo de Implementação da Web Semântica nas Empresas. In Á. Rocha, L. P. Reis, M. P. Cota, M. Painho, & M. C. Neto (Eds.), *Actas da 8ª Conferência Ibérica de Sistemas e Tecnologias de Informação (CISTI'2013), Lisboa, Portugal, 19-22 Junho 2013* (pp. 615–620). Lisbon: AISTI.
- Fuchs, C., Hofkirchner, W., Schafranek, M., Raffl, C., Sandoval, M., & Bichler, R. (2010). Theoretical Foundations of the Web: Cognition, Communication, and Co-Operation. Towards an Understanding of Web 1.0, 2.0, 3.0. *Future Internet*, 2(1), 41–59. doi:10.3390/fi2010041
- Hall, W., & O'Hara, K. (2008). Web Science. *Association of Learning Technologies Newsletter*.
- Heath, T., & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. In J. Hendler & F. Van Harmelen (Eds.), *Proceedings of the 17th international conference on World Wide Web WWW 08* (Vol. 1, pp. 1–136). Morgan & Claypool. doi:10.1145/1367497.1367760
- International Telecommunications Union. (2010). *The World in 2010*.
- International Telecommunications Union. (2013). *The World in 2013: ICT Facts and Figures*.
- Segaran, T., Evans, C., & Taylor, J. (2009). *Programming the Semantic Web*. (M. Treseler, Ed.) (p. 300). O'Reilly.

- Shadbolt, N. (2008). Research Roadmap - Web Science Trust. *Web Science Trust*. Retrieved February 20, 2011, from <http://webscience.org/research/roadmap.html>
- Shadbolt, N., & Berners-Lee, T. (2008). Web Science Emerges. *Scientific American*, 299(4), 76–81.
- Shannon, V. (2006). A “more revolutionary” Web - The New York Times. *International Herald Tribune*. Retrieved February 05, 2011, from http://www.nytimes.com/2006/05/23/technology/23iht-web.html?_r=1&scp=2&sq=Tim Berners-Lee&st=cse
- Spivack, N. (2010). The Evolution of the Web: Past, Present, Future | Nova Spivack - Minding the Planet. Retrieved January 31, 2011, from <http://www.novaspivack.com/uncategorized/the-evolution-of-the-web-past-present-future>
- W3C. (2004a). Resource Description Framework (RDF): Concepts and Abstract Syntax. Retrieved from <http://www.w3.org/TR/rdf-concepts/>
- W3C. (2004b). RDF/XML Syntax Specification (Revised). Retrieved from <http://www.w3.org/TR/rdf-syntax-grammar/>
- W3C. (2008). Cool URIs for the Semantic Web. Retrieved from <http://www.w3.org/TR/cooluris/>
- W3C. (2012). HTML+RDFa 1.1. Retrieved from <http://www.w3.org/TR/rdfa-in-html/>
- W3C, & IETF. (2001). *URIs, URLs, and URNs: Clarifications and Recommendations 1.0*.
- Wikipedia Users. (2011). Web 2.0.