

## Web 3.0 e Web semântica – do que se trata?

Apesar de não ser aceite por todos, corre a teoria de que a *World Wide Web* (ou, mais simplesmente, a Web) tem evoluído à semelhança de um programa de software, através de diferentes versões, com características diferentes e relativamente bem definidas. É consensual que a Web 1.0 foi a primeira geração da Web. Nesta fase, a principal preocupação era a própria construção da Rede, torná-la acessível e comercializável. As áreas de interesse centravam-se no acesso à Internet através de ISPs, nos primeiros *web browsers*, em protocolos como o http, em metalinguagens como o html e o xml, na criação de sites e portais, etc.

Contrariamente à primeira versão, o conceito de Web 2.0 revela-se mais impreciso. Refere-se a uma suposta segunda geração de serviços baseados na Internet, a qual enfatiza a colaboração *on-line* e a partilha entre utilizadores. Apesar de difuso, a maioria dos que estão ligados à indústria da Web concordaria que o conceito de “Web 2.0” se identifica com algumas realidades, como o trabalho social em rede (redes sociais<sup>1</sup>), os *folksonomies*<sup>2</sup>, *social bookmarking*<sup>3</sup>, etc. Nesta linha de raciocínio, um dos fenómenos mais emblemáticos da actual Web será a *Wikipedia*.

Com um grau de indefinição semelhante ao da Web 2.0, a Web 3.0, assim crismada por John Markoff em 2006, pressupõe uma terceira geração de serviços baseados na Internet, os quais se suportam no que podemos chamar “Web inteligente”, isto é, um tipo de Web que, por exemplo, se baseia numa maior capacidade do software em interpretar os conteúdos em rede, devolvendo resultados mais objectivos e personalizados de cada vez que se fizer uma pesquisa. Esta é uma das características da Web semântica, uma proposta do “pai” da *World Wide Web*, Tim Berners-Lee, identificável com a terceira geração da Web, em que os conteúdos podem ser registados de maneira a serem compreendidos, interpretados e processados por determinados

---

<sup>1</sup> Trata-se, fundamentalmente, da partilha de ideias entre pessoas que possuem valores, interesses e objectivos em comum. Tal realidade leva a que um grupo de discussão seja composto por indivíduos que possuem identidades semelhantes. As redes sociais estão hoje instaladas principalmente na Internet, devido ao facto deste ambiente possibilitar uma aceleração na divulgação das ideias e na captação de novos participantes.

<sup>2</sup> Prática de criar e gerir palavras-chave de uma forma colaborativa, para anotar e categorizar conteúdos.

<sup>3</sup> É uma forma de os utilizadores da Internet armazenarem, organizarem, partilharem e pesquisarem marcadores (*bookmarks*) de páginas Web. No sistema de *bookmarking*, os utilizadores mantêm links para páginas da Web que querem recuperar e partilhar.

agentes de software<sup>4</sup>, os quais passam a pesquisar, partilhar e integrar a informação disponível de uma forma mais eficiente.

O princípio é simples. Actualmente, as páginas da Internet estão feitas de modo a poderem ser lidas pelas pessoas, não pelos computadores. Um das consequências desse facto é que, em pesquisas mais simples ou mais complexas, embora os computadores devolvam resultados, são as pessoas que procedem à filtragem da informação devolvida, porque só elas é que tem a capacidade para o fazer. Quantos de nós não se depararam já com centenas ou milhares de resultados (ou, de outras vezes, demasiado poucos ou mesmo nenhuns) quando se faz uma pesquisa mais complexa na Internet, sendo obrigados a fazer uma depuração, página a página, sabe-se lá com que custos de tempo e resultados relevantes?

A Web semântica permitirá que as pessoas e computadores trabalhem em cooperação na exploração do conhecimento, uma vez que pressupõe a atribuição de significado aos conteúdos publicados na Internet e o desenvolvimento de tecnologias e linguagens que colocam esse significado ao alcance das máquinas. Ao mesmo tempo, o estabelecimento de significados e de linguagens compatíveis garante ambientes de maior interoperabilidade entre serviços. Tendo em conta essas características, não é de estranhar que os computadores possam, no futuro próximo, ser instruídos para fazer pesquisas tão complexas como apresentar uma lista de preços de televisores de alta definição, com *écran* superior a 70 cm, resolução de 1080p, à venda em lojas da cidade mais próxima, abertas até às 20H00 durante os fins-de-semana.

Para fazer face a este tipo de pesquisa, a Web semântica baseia-se, fundamentalmente, em três pilares. O primeiro deles é a meta-linguagem XML (*eXtra Markup Language*), que torna possível ao autor de conteúdos estruturá-los/organizá-los de uma forma personalizada através de anotações/marcações (*tags*) feitas sobre os dados, de que resultam metadados, essenciais para uma recuperação eficiente da informação na Web. Uma das premissas essenciais é que os computadores só empreendem tarefas inteligentes – que envolvem deduções e raciocínios automatizados – se as colecções de informação a que acedem forem estruturadas e organizadas, algo que é, desde logo, assegurado pelo XML. Por outro lado, a linguagem XML, se admite a diversidade e

---

<sup>4</sup> O conceito de agente descreve uma entidade de software complexa, capaz de actuar com um certo grau de autonomia para cumprir tarefas em nome do seu utilizador. Um agente define-se de acordo com o seu comportamento (persistência, habilidade social, reactividade e independência são alguns desses “traços de carácter”).

estimula a recuperação de dados heterogéneos na Web, também aceita que sobre ela se apliquem esquemas sintácticos que visam dar sentido lógico (uma segunda organização, mas não de significado) à informação, como é o caso do RDF (*Resource Description Framework*), que, constituindo o segundo pilar ou camada em que se suporta a Web semântica, mais não é que do que um modelo de representação para fazer, em XML, afirmações acerca de recursos disponíveis na Web. Através da fraseologia RDF, pode dizer-se que determinado recurso (uma página Web, por exemplo) foi criado por este ou aquele autor. Ao identificar e sequenciar um “sujeito” (o recurso), um “predicado” (a acção) e um valor (o autor), o RDF explicita uma sintaxe que ajuda a fazer a ponte entre aplicações diferentes, das quais se espera que partilhem dados.

Todavia, não basta haver uma linguagem (o XML) e uma sintaxe (o RDF) para assegurar uma compreensão abrangente e relacional da informação por parte dos computadores. Se uma determinada aplicação pode criar os seus próprios conceitos e para que outras possam manipulá-los, é necessário que o significado desses conceitos seja exposto de tal maneira que possa ser interpretado automaticamente. O papel de definir termos que representam objectos, conceitos e outras entidades e de definir axiomas formais que condicionam a interpretação e uso desses termos cabe ao terceiro pilar da Web semântica: as ontologias. Na prática, as ontologias servem de vocabulário consistente para a troca de informações entre aplicações, na medida em que, para que estas “falem” a mesma língua de uma forma organizada, é imprescindível que dominem o significado da informação que pretendem partilhar.

Apontando-se o ano de 2007 como o limite para a implementação da terceira geração da Web, espera-se que, entre 5 a 10 anos, a actual Web – uma rede de aplicações e repositórios de conteúdos separados – seja transformada num todo em que a interoperabilidade é a chave. Para já, são várias as iniciativas em desenvolvimento que procuram assegurar que os computadores interpretem a linguagem humana. Entre elas, destacam-se os projectos Scorpion e Indexa, ambos com o propósito de indexar de forma electrónica e automática os documentos patentes na Internet. Contando com o *apport* de técnicas biblioteconómicas, esses projectos acabam por colocar, uma vez mais, as bibliotecas na vanguarda da tecnologia. Era desejável que a área dos arquivos pudesse partilhar, tanto quanto antes, desse admirável mundo novo...