



Yahoo fica verde

O Yahoo, um dos maiores portais da Internet, acrescentou um novo site com fins ecológicos. O Yahoo Green é uma iniciativa destinada a reduzir as emissões

de gases de estufa; o seu público alvo é essencialmente americano, mas está aberto a outros internautas. O endereço do site é green.yahoo.com.

Empresa brasileira desenvolve programa que aprende ao ler textos

Tecnologia reconhece várias línguas, pode ser usada para melhorar as técnicas de pesquisa *online* e é um passo a caminho da Web 3.0

João Pedro Pereira
jppereira@publico.pt

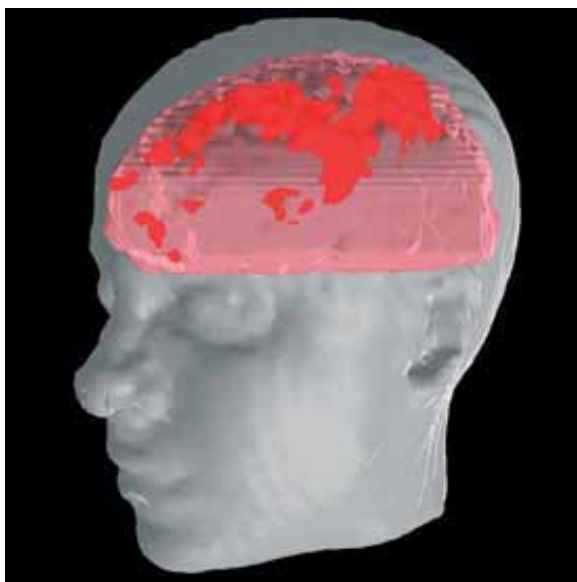
● A empresa brasileira Cortex Intelligence está a desenvolver uma tecnologia que permite a um programa de computador não apenas compreender os textos que lê, como melhorar sozinho a capacidade de leitura, à medida que tem contacto com mais textos.

Atualmente, a maioria dos programas de computador que precisam de analisar textos (como os programas dos motores de busca na Web) não é capaz de descobrir as associações complexas entre os diversos elementos de um texto nem de perceber o sentido que estes possuem enquanto conjunto.

Uma das aplicações deste tipo de tecnologia (que tem vindo a ser uma aposta de várias empresas) é precisamente tornar mais eficazes as pesquisas *online*: "Poderemos ter um sistema de pesquisa bem mais poderoso, que não procure apenas por palavra-chave, mas também pela sua semântica e relacionamentos", explica o fundador e director de tecnologia da Cortex Intelligence, Christian Aranha.

Os motores de busca, embora usem várias técnicas para apresentar e hierarquizar os resultados de uma pesquisa, ainda não são capazes de interpretar frases ou perguntas introduzidas pelos utilizadores. Escrever no Google a expressão "estimativa da quantidade de documentos na Internet" não devolve qualquer resultado, exemplifica Aranha. Isto acontece porque o programa procura páginas que tenham essa expressão exacta e não interpreta qual é a intenção do utilizador.

A tecnologia de interpretação de texto poderá abrir caminho não apenas para uma pesquisa mais inteligente, mas para o que muitos chamam a Web semântica ou a terceira fase do mundo *online* - a Web 3.0. Chegar a este estágio é, aliás, um dos objectivos do próprio fundador da Web, Tim



A leitura feita pelo computador

A tecnologia está disponível para ser experimentada por qualquer pessoa em www.cortex-intelligence.com/engine. O utilizador pode pedir ao sistema para sortear um texto dos que possui na base de dados e para o analisar.

Uma das opções identifica aquilo a que o programa chama as "entidades" do texto - basicamente, todos os substantivos. Ao passar o cursor sobre essas entidades, o utilizador recebe informação adicional, que inclui muitas vezes uma breve definição retirada da Wikipedia.

Já ao seleccionar a opção "acções", o sistema identifica os verbos presentes no texto. Clicar sobre cada um deles permite descobrir o relacionamento entre os elementos do texto. Ou seja, o programa identifica os diversos componentes de um enunciado (sujeito, predicado, complemento) respondendo, por exemplo, a questões como Quem? O quê? Onde? ou Quando?

O sistema tem ainda algumas falhas e não é permitido ao utilizador (a página serve apenas para demonstrar as potencialidades da tecnologia) introduzir os seus próprios textos.

Berners-Lee.

Nesta Web semântica, os computadores seriam capazes, por exemplo, de perceber que num site de vendas uma imagem de um DVD corresponde à descrição que é feita desse objecto e que os números que estão ao lado são o preço.

Quando estão ainda armazenados numa base de dados, os computadores conseguem perceber o que significa cada pedaço de informação, porque cada um está associado a campos descritivos. Mas, quando a página Web é formada a partir desses conteúdos, perdem-se os dados descritivos. Como consequência, a Web tem milhões de páginas cuja informação os computadores não conseguem interpretar satisfatoriamente.

"O que falta para a Web 3.0 chegar de vez é justamente alguma tecnologia que possa gerar os meta-dados automaticamente", explica Aranha. "Essa é a contribuição que a nossa tecnologia pode dar para o lançamento da Web 3.0: enriquecer os textos da Web 1.0 e 2.0 com metadados."

No caso da Cortex Intelligence, a tecnologia foi concebida de forma a que a base do sistema possa ser usada em qualquer língua. A partir daí, a empresa desenvolveu já aplicações para português, inglês e espanhol. Para além disso, o programa é capaz de se tornar mais inteligente à medida que analisa mais dados: "Não temos uma equipa de pessoas a inserir manualmente informações. A nossa tecnologia é capaz de aprender novos conhecimentos lendo os textos".

● www.cortex-intelligence.com/engine