

Os Sistemas de Recuperação de Informação

Mário J. Silva

Universidade de Lisboa, Departamento de Informática

Nos sistemas de bases de dados, as interrogações processam-se sobre informação estruturada e os resultados satisfazem predicados lógicos. Nos sistemas de recuperação de informação, satisfazem critérios vagos de relevância, estimados sob a forma de medidas de similaridade ou distância. Ao mesmo tempo, os requisitos de estruturação da informação exigidos nestes sistemas são mais fracos, podendo esta ser introduzida sob a forma de textos (escritos ou falados) em língua natural, ou digitalizações de dados observados, como imagens ou filmagens.

Os sistemas de recuperação têm muito em comum com os sistemas de análise estatística ou de prospecção de dados, na medida em que as noções de similaridade se apoiam muitas vezes em métricas de natureza estatística. Há no entanto também um aspecto fundamental que os diferencia: os sistemas de recuperação de informação são primariamente interactivos. É o utilizador que submete as interrogações e é o seu julgamento da relevância dos resultados que conduz o processo de prospecção.

A recuperação de informação é, em última análise, uma solução eficaz para solução de casos particulares de questões fundamentais em IA e reconhecimento de padrões. Tira partido da capacidade de análise do conteúdo semântico de baixo nível presente na informação com os algoritmos e tecnologias de suporte à informação disponíveis no presente. À medida que estas evoluem, prevê-se a incorporação de novos métodos de processamento de linguagem natural ou compreensão de imagens nas aplicações informáticas e motores de pesquisa cada vez mais usados no dia a dia pelo cidadão comum.

O paradigma da recuperação de informação está hoje cada vez mais presente hoje nos sistemas de informação, o que pode ser ilustrado num leque muito variado de aplicações que inclui as que são mais usadas no dia a dia (o correio electrónico e os motores de pesquisa).