

Título da comunicação: *Arquivo da Web Portuguesa: uma infraestrutura para investigação científica*

Resumo:

A Web foi idealizada como um meio de comunicação rápido mas tem vindo progressivamente a substituir a imprensa como meio privilegiado de publicação porque permite que qualquer pessoa ou organização disponibilize informação acessível a todos de uma forma rápida e económica. Diariamente, são publicados milhões de conteúdos exclusivamente na Web como textos, fotografias ou vídeos. No entanto, passado relativamente pouco tempo, a grande maioria desta informação deixa de estar acessível e perde-se irremediavelmente. Cerca de 50% dos endereços disponíveis hoje, tornam-se indisponíveis passados apenas 2 meses, 80% dos conteúdos são alterados ou desaparecem passado 1 ano.

O Internet Archive é uma organização norte-americana que recolhe e arquiva conteúdos da Web à escala mundial. É difícil para uma única organização fazer um arquivo exaustivo de todos os conteúdos publicados porque a Web está em permanente mutação e muita informação desaparece antes de poder arquivada. Além disso, a documentação de acontecimentos históricos de relevância nacional para Portugal não é prioritária para o Internet Archive e grande parte da informação publicada na Web portuguesa perde-se irremediavelmente. Este problema é sentido igualmente por outras comunidades nacionais já existem pelo menos 48 iniciativas de arquivo espalhadas pelo mundo.

O arquivo da Web além de contribuir para preservar conhecimento científico e histórico, permite também que cidadãos comuns mantenham as suas memórias enquanto indivíduos. Todos os dias as pessoas tiram fotografias e partilham-nas na Web. Contudo, os cuidados mais elementares para a preservação desta informação, como copiar as fotografias para um disco, raramente são tomados. No futuro, muitas pessoas terão dificuldade em mostrar retratos das suas memórias ou famílias.

A publicação na Web quebrou barreiras económicas e geográficas ao conhecimento, mas a natureza efémera da Web poderá fazer com que se quebre o acesso ao próprio conhecimento. Cabe-nos a responsabilidade de fazer com que a informação publicada em-linha permaneça acessível para as gerações futuras.

O Arquivo da Web Portuguesa visa arquivar e preservar conteúdos da Web relevantes para a comunidade portuguesa. Este projeto destaca-se por ter desenvolvido um serviço de pesquisa textual sobre conteúdos do passado disponível em <http://www.arquivo.pt>. Através deste serviço público é possível pesquisar rapidamente informação arquivada da web desde 1996 até aos dias de hoje.

O Arquivo da Web Portuguesa fornece um serviço de preservação gratuito aos autores da web portuguesa e ao mesmo tempo um recurso valioso de investigação que já foi utilizado por investigadores, por exemplo, para medir de forma automática a acessibilidade da web portuguesa para pessoas com deficiência.

Nota biográfica:

Daniel Gomes. Actualmente coordena o Arquivo da Web Portuguesa e a unidade de desenvolvimento web da FCCN-FCT. Os seus principais interesses são gestão de equipas, usabilidade, desenvolvimento e controlo de qualidade de sistemas de informação baseados na web. Em particular, na sua tese de doutoramento disserta acerca do impacto das características da informação publicada na web sobre o desenho de sistemas de larga escala que permitam o seu processamento. A web portuguesa foi objecto de estudo frequente ao longo deste trabalho.

Daniel Gomes é investigador convidado no grupo XLDB da Universidade de Lisboa, tendo trabalhado principalmente em recolha, integração e pesquisa de informação proveniente da web.