

NOTA DE IMPRENSA

Projeto da FCT | FCCN permite reencontrar *websites* do passado

12 de janeiro, 2016

Passado um ano da publicação, 80% das páginas *online* desaparecem ou são alteradas. O serviço de preservação digital prestado pelo Arquivo.pt é essencial para conservar uma parte da herança cultural do país e contribuir para que informação valiosa não se perca. Muitos dos conteúdos arquivados são únicos no mundo e são ainda preservados *websites* de interesse internacional, como a mais antiga página da [Library of Congress](#).

Sempre que seja necessário visitar informação perdida na web no seu estado original, é possível encontrá-la em <http://arquivo.pt>.

O Arquivo.pt disponibiliza um serviço de pesquisa público e gratuito sobre dados arquivados da *web* desde 1996. O desafio de preservar a *Web* requer um esforço contínuo e é na recuperação do acesso a estes endereços que o serviço prestado pelo Arquivo da Web Portuguesa (Arquivo.pt) se propõe ser útil, para que o conhecimento nela contido esteja acessível a gerações futuras.

Iniciativa da unidade FCCN da FCT, esta é uma infraestrutura de investigação focada na preservação de conteúdos de interesse para a comunidade portuguesa. Contudo, poderá servir utilizadores internacionais, uma vez que preserva páginas em várias línguas e oferece a possibilidade de acesso a conteúdos históricos em português a utilizadores estrangeiros, através de ferramentas de tradução automática.

Como funciona o Arquivo.pt

A preservação realiza-se de forma automática e recorre a um sistema informático de larga escala. Diariamente, é feita uma recolha de 300 publicações selecionadas e, a cada 3 meses, todos os sites sob o domínio .PT são guardados. São ainda preservados sites sob outros domínios como .org, .com ou .eu. Após a recolha, a informação é processada para que se torne pesquisável e acessível.

Encontram-se arquivados cerca de 2 700 milhões de ficheiros, (95 TB de informação) e qualquer pessoa pode sugerir endereços interessantes para preservação, através da página [arquivo.pt/sugerir](#).

Como surgiu o serviço

O Arquivo.pt teve origem no projeto “TUMBA!”, realizado na FCUL entre 2001 e 2006. O resultado desta investigação nacional na área da preservação digital gerou o primeiro protótipo de arquivo da web portuguesa.

Em 2007 a FCCN lançou o projeto “Arquivo da Web Portuguesa”, tendo sido realizada, em 2008, a primeira recolha da web portuguesa. Em 2012 foi disponibilizado o serviço público de pesquisa e acesso a páginas do passado.

O Arquivo.pt no Presente

Em 2015, o Arquivo.pt recolheu para preservação cerca de 580 milhões de ficheiros e o serviço de pesquisa registou, em média, 3 692 utilizadores por mês (90% novos utilizadores).

Estes resultados devem-se a um esforço de fortalecimento do serviço, operado nos últimos anos e assente na aposta em servidores mais adequados para tarefas especificamente relacionadas com a preservação da *Web*.

Com o lançamento de uma nova versão do Arquivo.pt (versão Fénix) verificam-se melhorias de acesso e interoperabilidade do serviço de pesquisa do Arquivo.pt com outros arquivos da web, projetos de investigação ou aplicações web que possam vir a ser desenvolvidas (por exemplo, por alunos de Informática).

Para ajudar a perceber como funciona o serviço que permite pesquisar o passado na *web* foi desenvolvido um vídeo que é atualmente apresentado na página principal.

Colaborações e desafios futuros

O Arquivo.pt realizou uma primeira tentativa de arquivar *sites* alojados sob o domínio .EU, um projeto que surgiu no âmbito de atividades do RESAW (*Research Infrastructure for the Study of Archived Web Materials*), a rede europeia que tem como objetivo a criação de uma infraestrutura dedicada ao estudo de materiais arquivados da *web*.

Está planeada uma atividade de arquivo de páginas dedicadas às Eleições Presidenciais de 2016, replicando o esforço já feito nesse sentido para as Eleições Legislativas de 2015, contando com a ajuda de voluntários que ajudaram a identificar 184 sites relacionados com a temática.

Para 2016 o Arquivo.pt tem grandes desafios, nomeadamente vir a ser mais conhecido e utilizado pelo público como uma ferramenta útil. Estará em desenvolvimento um protótipo de pesquisa de imagens arquivadas, função requisitada pelos utilizadores, e há ainda planos para potenciar a reprodução de conteúdos arquivados de forma a evitar falta de imagens na visualização das páginas arquivadas.

Sobre o Arquivo.pt

Em <http://sobre.arquivo.pt> poderá ficar a saber mais acerca do serviço.

Contacto:

c.imagem@fccn.pt