

Preservação da Web – Análise do Arquivo da Web Portuguesa (AWP)

fevereiro 26, 2010 alexdasilveira 1 comentário

Continuando a análise de arquivos da web que o Bibliotecno iniciou (veja as análises do Bibliotecno para o [UK WEB ARCHIVE](#) da Biblioteca Britânica e do [LCWA/MINERVA](#) da Library of Congress) chegamos ao [Arquivo da Web Portuguesa \(AWP\)](#) que ainda está em fase experimental, mas tem seu histórico com raízes em 2001 com o projeto tumba (temos um motor de busca alternativo) desenvolvido por um grupo de investigação da Universidade de Lisboa.

Recentemente o AWP liberou uma versão experimental do arquivo, contudo, antes mesmo desta versão experimental já era possível verificar ações interessantes sobre o projeto. A principal inovação é o rARC (Replicador de Arcs), sistema de salvaguarda da informação arquivada do AWP, que tem como novidade utilizar um determinado espaço dos computadores dos próprios usuários como Backup do arquivo. A idéia é de que mesmo que o arquivo central se perca por completo ainda assim será possível recuperá-lo. Cada usuário que tiver interesse em colaborar com a preservação do AWP poderá ceder no mínimo 100mb de seu disco rígido (espaço para guardar um arquivo arc). A idéia não é utilizar o espaço do HD do usuário para sempre, possibilitando a este apagar o conteúdo arquivado e aumentar ou diminuir o espaço destinado a colaboração, assim, o AWP irá redistribuir os arquivos para computadores de outros usuários, mantendo este como um processo em plena mutação.

Webpage Screenshot

Experimental
ARQUIVO DA WEB PORTUGUESA

entre 01.01/1996 e 25.02/2010
dd/mm/aaaa dd/mm/aaaa

Pesquisar

Pesquisa Avançada

Pesquisar no Arquivo da Web Portuguesa

Com este motor de pesquisa poderá pesquisar páginas da web portuguesa recolhidas ao longo de vários anos. Poderá ajustar o intervalo de tempo sobre o qual a pesquisa será feita de forma a obter resultados temporalmente mais precisos.

Exemplos de pesquisa:

- [FCCN](#)
- [Internet Portuguesa entre 1998 e 2001](#)

entre 01.01/1996 e 25.02/2010
dd/mm/aaaa dd/mm/aaaa

FCCN Fundação para a Computação Científica Nacional

POS CONHECIMENTO Programa Operacional Sociedade do Conhecimento

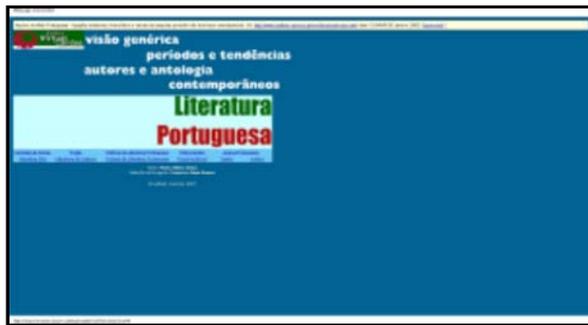
UMIC Universidade Nova de Lisboa

União Europeia FEDER

<http://experimental.arquivo.pt/vtchman/>

Tela inicial experimental do AWP

A tela inicial da versão experimental do AWP é bem simples, com um campo de pesquisa por termo livre e um diferencial em relação aos arquivos analisados: a possibilidade de demarcação de período cronológico. Mas há um link para uma segunda interface que apresenta a pesquisa avançada com a segregação dos campos por: palavras (contendo palavras, frase exata ou exclusão de termos), data (permitindo combinar o período com refinamentos de ordenação), formato de arquivo, URL e quantos resultados serão exibidos por página.



Página do Centro Virtual Camões arquivada em 2002



Mesma página do Centro Virtual Camões em 2010, mudanças na linguagem não verbal



Link ficção da página do Centro Virtual Camões em 2002



Em 2010 o link "ficção" da imagem acima não encontra-se disponível

Em outras pesquisas realizadas o resultado do arquivamento pode ser considerado satisfatório, sendo possível acessar novamente páginas antigas conforme elas eram no ato de sua publicação ou quando existiam (finalidade do conceito de arquivamento da web) até mesmo em páginas com uma quantidade maior de atualizações, contudo, é possível encontrar páginas onde os elementos visuais (linguagem não verbal) não foram preservados, principalmente nos conteúdos mais antigos.

É possível concluir que para um arquivo em fase experimental o AWP já está nivelado em relação aos demais arquivos existentes, possibilitando constatar o bom trabalho da Fundação para a Computação Científica Nacional, do Programa Operacional Sociedade do Conhecimento e da Agência para a Sociedade do Conhecimento. Diferente dos demais arquivos já apresentados no Bibliotecno este não é mantido por uma Biblioteca Nacional. Na página do AWP é possível encontrar várias outras informações: <http://arquivo-web.fccn.pt/>