

## Diário da Assembleia da República Electrónico.

Joaquim Sousa Pinto, Pedro Almeida

**Resumo** – Neste artigo apresentamos o projecto do Diário da Assembleia da República Electrónico, que foi desenvolvido no âmbito da disciplina de projecto do curso de Licenciatura em Engenharia Electrónica e Telecomunicações, leccionado na Universidade de Aveiro.

**Abstract** – This article is to present the project Diário da Assembleia da República Electrónico, that was developed as part of the course Engenharia Electrónica Telecomunicações at the University of Aveiro.

### I. INTRODUÇÃO

Hoje em dia o computador é cada vez mais uma ferramenta de trabalho que não podemos desprezar.

O seu uso como fonte de informação está a implementar-se rapidamente e torna-se necessário criar sistemas de informação avançados.

Estes sistemas podem funcionar quer a nível local, dentro de uma intranet, quer a nível global, através de um site da Internet.

Um aspecto interessante destes sistemas é o facto de ser distribuído, permitindo assim que a informação esteja dividida por vários computadores.

No âmbito deste trabalho foi desenvolvido um sistema que possibilita a pesquisa de expressões dentro de um conjunto de documentos. Após apresentada a listagem dos documentos que satisfazem a expressão de pesquisa, podemos então seleccionar um dos documentos e visualizar o seu conteúdo. Quando se apresenta o documento as palavras que foram utilizadas na expressão de pesquisa aparecem destacadas para possibilitar uma rápida identificação do texto que se desejava encontrar.

O sistema permite também a possibilidade de ver a imagem digitalizada do documento.

### II. IMPLEMENTAÇÃO

Para construir este sistema foi necessário utilizar várias linguagens de programação e de serviços que são disponibilizados pela MICROSOFT, estando estes incorporados no sistema operativo WINDOWS 2000 SERVER. De seguida vamos descrever muito brevemente as tecnologias que foram utilizadas para uma melhor compreensão do trabalho.

#### *Tecnologias utilizadas:*

- HTML ( Hypertext Markup Language)

O HTML é uma linguagem de marcas utilizada na World Wide Web que permite visualizar hipertexto. É necessário um programa, um browser, para poder interpretar e apresentar correctamente o conteúdo de um documento HTML.

- URL (Uniform Resource Locator)

O URL indica-nos o endereço de um recurso acessível na rede.

- ASP (Active Server Pages)

ASP's é uma tecnologia que permite que se possam executar scripts do lado do servidor. Do lado do cliente visualiza-se somente HTML puro, permitindo assim que os detalhes da implementação de um sistema só sejam “vistos” do lado do servidor. Através do uso de ASP's podemos aumentar as potencialidades de um servidor web pois é possível comunicar com bases de dados registadas no servidor, executar scripts, e utilizar objectos criados pelo programador.

- Linguagens de Script (JavaScript, VBScript)

O JavaScript é uma linguagem de programação que permite dar alguma funcionalidades a uma página HTML. É uma linguagem derivada da linguagem C, não possuindo, no entanto, a rigidez associada com uma “typed language”, em que as variáveis possuem um tipo definido e são previamente definidas. A capacidade de processamento de JavaScript depende dos browsers. Com JavaScript podemos, por exemplo, validar informação que é introduzida em formulários, libertando assim o servidor de verificar o conteúdo dos vários campos do formulário. Este processo permite que o servidor não seja constantemente sobrecarregado com pedidos do cliente para validar dados. O JavaScript tem

muitas mais funcionalidades como aproveitar certos eventos para alertar o utilizador de alterações que estão a ser feitas no documento, ou simplesmente animar um documento. O VbScript é também uma linguagem de script derivada da linguagem Visual Basic.

- COM/ACTIVEX™

**COM** (Component Object Model) é uma arquitectura de software standard baseada em interfaces, e desenhada para ter código separado em objectos “self-contained”. Estes componentes binários (nível mais baixo de troca de informação/ dados /código) estão preparados para funcionar com qualquer cliente COM. A tecnologia COM é independente da linguagem de programação. Quando componentes COM trabalham entre vários computadores ligados em rede, os componentes são denominados como distributed COM ( **DCOM** ).

Sendo assim, **ActiveX** é uma tecnologia que integra componentes de software, num ambiente de rede, independentemente da linguagem em que foram criados. Um control ActiveX, anteriormente conhecido como control OLE ou control OCX, é implementado tendo como suporte a tecnologia COM. No nosso caso, desenvolvemos componentes em C++, e integramo-los em páginas web.

Utilizando controls ActiveX, permitiu encapsular o código, protegendo deste modo a propriedade intelectual dos componentes desenvolvidos.

Para desenhar componentes ActiveX, utiliza-se a ajuda de Class Libraries. Existem duas livrarias para desenhar controls ActiveX.

- Microsoft foundation classes (MFC)
- Active Template Library (ATL) ( O objectivo desta livraria é executar apenas o código para a aplicação, sendo por isso desnecessário, associar-se a outras livrarias.)

- IIS (Internet Information Service)

O IIS permite criar um site da Internet com todas as funcionalidades que lhe estão associadas. Através da utilização deste serviço é possível utilizar ASP's na construção de sites WEB.

Este serviço disponibiliza ainda a possibilidade de criar directorias virtuais, que podem estar localizadas em outro disco ou computador, que podem ser utilizadas com se fizessem parte do nosso site.

- Indexing Service

Este serviço permite indexar um conjunto de ficheiros e formar um catálogo com os mesmos. Após criado o catálogo e com ajuda das linguagens de programação mencionadas acima, podemos pesquisar o mesmo. Este serviço pode ser configurado de modo a permitir que nem todas as palavras dentro de um documento façam parte do catalogo.

Quando efectuamos uma pesquisa no catálogo o serviço indica-nos os documentos que satisfazem a expressão de pesquisa utilizada.

### III – COMPONENTES

Foi necessário contruir componetes que permitissem a construção do site e a gestão do mesmo. Para a gestão do sistema foi construído um componente cujo nome é TXT2HTM. Para a construção do site foram desenvolvidos outros dois componetes com o nome SHOWPAGE, e MAGICVIEWER.

#### A. Componente TXT2HTM.

O projecto **Diário da Assembleia da República electrónico**, para o seu funcionamento, necessita das diversas páginas do diário no formato HTML, para proceder à sua indexação, e à posterior visualização.

Para que através de uma simples página se possa novamente reconstruir o diário, as páginas no formato HTML, deverão ter como marcas informativas (*meta tags*), o número do diário, da legislatura, da série, da sessão, e as datas da sessão e do diário. As finalidades destas marcas é facilitar a acessibilidade e organização nas diferentes páginas de cada diário. Estas marcas possibilitam também a navegação dentro de um diário nas diferentes páginas, e dentro das diversas legislaturas, sessões e séries nos diferentes diários.

Como os diários da Assembleia da República são publicados em papel, irá ser necessário convertê-los para ficheiros HTM. Assim o primeiro passo será converter as páginas do diário em papel, para o formato digital, ou seja digitalizar as páginas para obter uma imagem e depois fazer um reconhecimento de caracteres, converte-los assim, para ficheiros de texto. Os formatos fornecidos para a imagem e o texto são os formatos TIF e TXT, respectivamente.

O segundo passo, irá ser realizado com a ajuda do programa **TXT2HTM**, que como o próprio nome indica, será a partir dos ficheiros de extensão TXT, anteriormente referidos, criar ficheiros HTM(L), onde serão inseridas as *meta tags* legislatura, série, sessão, número do diário, data da sessão e data do diário.

Os objectivos desta componente, são então os seguintes:

- Conversão de um ficheiro no formato TXT para um no formato HTM, ou seja, pegar na informação do

ficheiro TXT, e colocá-la dentro das marcas <BODY> </BODY> de um ficheiro HTM, com o mesmo nome.

- Inserir *meta tags* no ficheiro HTM, com a informação sobre o diário (número, data sessão,...).

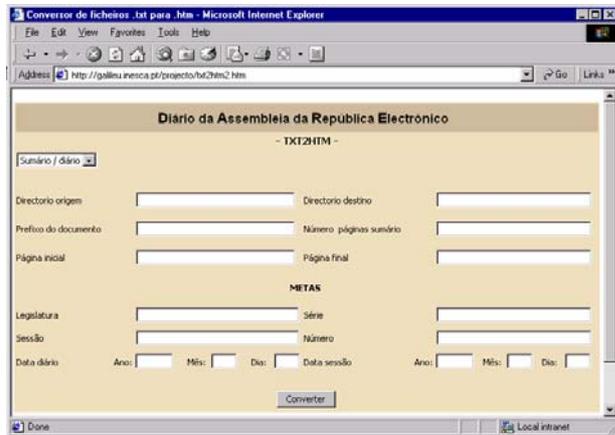


Fig.1 Interface do programa TXT2HTM.

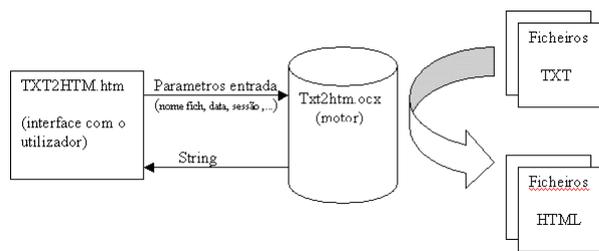


Fig.2. Modelo do Objecto.

**B. Componente SHOWPAGE.**

No âmbito do projecto **Diário da Assembleia da República Electrónico**, sempre que efectuamos uma pesquisa, é necessário apresentar o resultado da mesma de uma forma que o utilizador possa identificar rapidamente no texto pelo que pesquisou. Para tal, foi desenvolvido um módulo de software construído em visual C++, sendo este um projecto do tipo Active Template Library. Este tipo de projecto permite criar um ActiveX que vai ser incorporado dentro de uma ASP, permitindo um melhor funcionamento do sistema. Este ActiveX, em conjunto com uma stylesheet, permite realçar as palavras que desejamos encontrar. O ActiveX também permite obter valores que vão ajudar na construção de um cabeçalho que identifica o documento e que tem valores que permitem ajudar na navegação ao longo dos diários pesquisados.

Mais tarde, esta página terá também opções avançadas que permitirão imprimir um diário por completo ou seleccionar as páginas a imprimir.

Os objectivos do componente showpage são:

- Realçar uma ou mais palavras que foram utilizadas na expressão de pesquisa.
- Dado um documento de HTML com uma formatação predefinida, retirar deste documento informações que são relevantes para a apresentação da página em causa, e obter dados que permitam certas facilidades de navegação na apresentação da página. Estes dados são obtidos a partir do cabeçalho do documento, onde se encontram campos do tipo <META> com descrição dos dados que pretendemos obter (data, numero, ...).

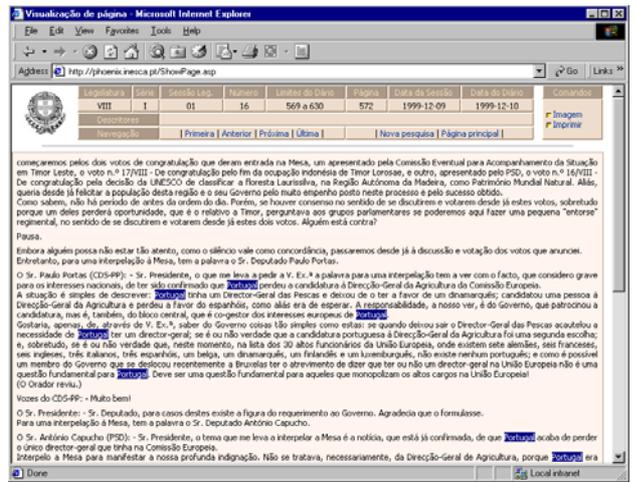


Fig. 3. Exemplo da página SHOWPAGE.ASP.

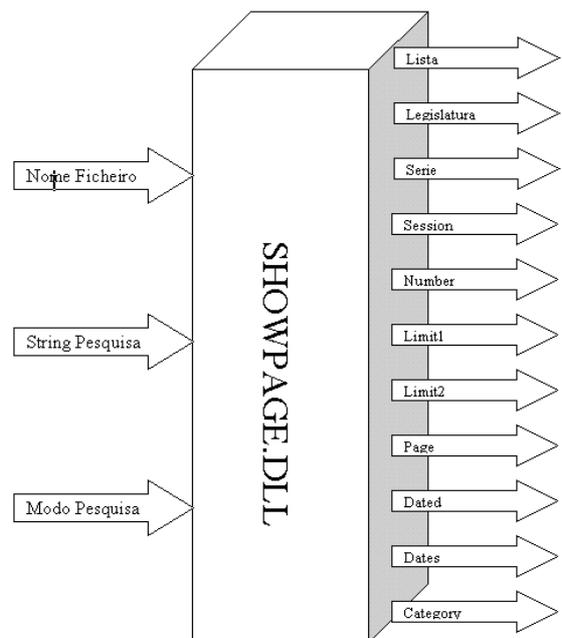


Fig. 4. Modelo do Objecto.

Este módulo de software é um ActiveX que foi desenvolvido através de uma ATL. Depois de compilado o programa, gera-se uma dll que é necessário registar no sistema, para que seja possível utilizar o objecto que foi criado. Normalmente este passo é executado automaticamente no Visual C++ 6.0, mas caso seja necessário instalar este sistema noutra computador é necessário recorrer a um ficheiro de instalação que regista a dll no sistema em causa.

O diagrama representa o modelo do objecto que foi criado. Este objecto representa o funcionamento do programa, tendo este parâmetros de entrada e parâmetros de saída. Associado a este objecto encontram-se métodos que podem ser invocados sobre ele, devolvendo cada método informação relativa a pedidos sobre o documento manipulado.

Para melhor compreensão do funcionamento deste sistema vamos descrever de seguida os vários métodos associados ao módulo.

- Método lista

Este método inicialmente preenche tabelas com expressões que vão ser pesquisadas dentro do documento HTML. Depois de preenchidas, são comparadas as palavras armazenadas, com as palavras do texto em que estamos a pesquisar. Após concluída a pesquisa, retorna o texto que foi pesquisado de forma a que seja possível realçar as palavras que se encontravam na tabela.

- Métodos legislatura, série, session, number limits1, limits2, page, dated, dates e category.

O funcionamento destes métodos é muito semelhante. Eles procuram o meta correspondente no ficheiro HTML e retornam o valor que se encontra nesse campo.

Após concluída a aplicação esta é inserida em uma página ASP. A inserção do objecto na página asp, é feita através do seguinte código:

```
Set objecto=
Server.CreateObject("showpage.mostra.1")
```

Após criado o objecto, para aceder aos vários métodos do mesmo, utilizamos a notação:

```
text_doc=objecto.lista(Nome_ficheiro,Str,Modo
_Pesquisa)
legis=objecto.legislatura()
serie=objecto.serie()
sess=objecto.session()
number=objecto.number()
limits1=objecto.limit1() * 1.0
limits2=objecto.limit2() * 1.0
page=objecto.page() *1.0
dated=objecto.dated()
dates=objecto.dates()
category=objecto.category()
```

Vamos agora descrever as funcionalidades associadas à página *showpage.asp*.

A maior parte das funcionalidades desta página estão associadas à tabela de navegação. Esta permite que o utilizador se possa situar dentro de um diário e que consulte o mesmo de uma forma eficaz, já que tem opções de navegação. Caso o utilizador pressione uma das opções do cabeçalho, como por exemplo *página seguinte*, vão ser introduzidos novos valores em campos escondidos da página, e vamos voltar a chamar a mesma página, *showpage.asp*, variando os parâmetros necessários para aparecer a página que pretendemos consultar. Na situação mencionada atrás, em que o utilizador pressiona o texto *página seguinte*, o utilizador vai voltar a ver a página *showpage.asp*, mas agora com o documento da página seguinte ao que estava a consultar. Caso apareçam nesse documento palavras que o utilizador pesquisava estas vão aparecer realçadas, isto é, a expressão de pesquisa mantém-se quando o utilizador navega ao longo do documento.

No caso de o utilizador pressionar página anterior o funcionamento é muito semelhante.

Se o utilizador pressionar página inicial então é apresentada a primeira página do diário e através de uma função em JavaScript desactivamos a opção de ir para a página anterior visto que não tem sentido voltar atrás da primeira página.

Quando o utilizador pressiona ultima página o funcionamento é semelhante, isto é desactivamos a opção de ir para a página seguinte, já que neste caso não faz sentido ir para a página seguinte.

Caso o utilizador carregue sobre imagem então vai ser chamada outra ASP que se encarrega de apresentar a imagem do diário original.

### C. Componente MAGIC VIEWER.

No âmbito do projecto **Diário da Assembleia da República Electrónico** sempre que efectuamos uma pesquisa, e visualizamos uma página, temos uma opção que nos permite ver o documento original digitalizado. Para que tal seja possível foi necessário construir um ActiveX que nos permitisse visualizar TIF's. Isto deve-se ao facto de o Internet Explorer 5.0 não possuir filtros para

abrir imagens deste formato. Este ActiveX é posteriormente incluído dentro de uma página da WEB, permitindo assim que possamos observar ficheiros deste formato dentro do Internet Explorer 5.0 . Sendo assim torna-se mais fácil a visualização dos documentos, sem ter de recorrer a programas externos para abrir as imagens. A utilização deste formato de imagem deve-se ao facto de este ser de qualidade bastante elevada e de tamanho reduzido, permitindo assim poupar recursos quer a nível de espaço de disco quer a nível de tempo de processamento. Este facto é bastante importante já que o número de imagens guardadas em disco é muito elevado. O facto deste programa obter as imagens a partir de um URL permite que o sistema seja distribuído, podendo as imagens estar guardadas numa máquina distinta de onde guardamos os ficheiros de texto.

Os principais objectivos deste componente, são então os seguintes:

- Abrir um ficheiro de imagem que se encontra num computador remoto, e apresenta-lo no computador do utilizador.
- Permitir que o utilizador visualize uma imagem em formato TIF dentro do Internet Explorer 5.0 .

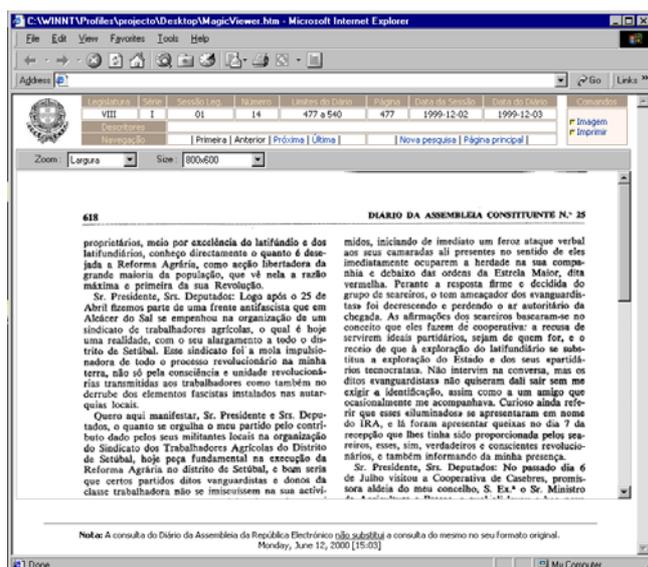


Fig 5. Exemplo da utilização do componente MAGICVIEWER.

É possível ao utilizador alterar o modo como o documento é apresentado. Disponibilizamos duas listas de opções: uma lista para zoom e outra para tamanho.

Na lista de zoom disponibilizamos seis opções diferentes. Quatro destas opções são para visualizar a página com diferentes ampliações. As outras duas servem para ajustar a página em largura ou em altura. Na lista de tamanho permitimos que o utilizador altere o tamanho do objecto de modo a poder visualizar melhor a página de acordo com as definições do browser.

Este módulo de software é um ActiveX Control que foi desenvolvido através de Visual Basic 6.0. Depois de compilado gera-se um *ocx* que é necessário registar no

sistema para que seja possível utilizar o objecto que foi criado.

Existem duas maneiras distintas de instalar este software no computador de um utilizador. Pode-se instalar o programa através de um programa de instalação normal ou através da Internet.

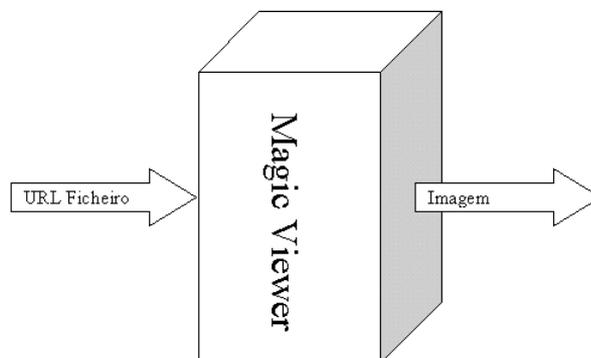


Fig. 6. Modelo do objecto

De seguida vamos descrever o funcionamento deste objecto para melhor compreensão do funcionamento deste sistema. Inicialmente fornecemos um URL ao objecto. A partir deste URL efectuamos um pedido HTTP ao servidor com o nome do ficheiro pretendido e criamos um ficheiro temporário no disco do utilizador para poder visualizar a imagem pretendida. Dentro deste objecto temos ainda incorporado um outro objecto. Este objecto é o objecto da Wang que permite visualizar TIF's e cujo nome é *imgedit.ocx* . Fornecemos então a este objecto, o nome do ficheiro que pretendemos abrir e apresentamos a imagem.

#### IV CONCLUSÕES E TRABALHO FUTURO

No geral a construção do site da Diário da Assembleia da república electrónico está concluída. As ferramentas necessárias para o seu bom funcionamento estão desenvolvidas e pensamos que o resultado global é bastante satisfatório, havendo no entanto ainda a necessidade de desenvolver algum trabalho a nível de design e de apresentação da informação.

#### REFERÊNCIAS

1. Skanson ,Jan , *C++ from the beginning*, Addison Wesley, 1997
2. Kamin, Samuel N., *Programming with class: A C++ Introduction to Computer Science*, McGraw-Hill,1996
3. Stroustrup, Bjarne, *The C++ Programming Language*, Addison Wesley, 1986
4. Skylyarov, Valery, *The Revolutionary guide to Turbo C++*, WROX Press, 1992
5. Brophy, Keith , *Teach yourself ActiveX Control programming with Visual Basic 5 in 21 days*, 1997