

Serviço AMIS – Publicação, edição e pesquisa em repositórios de informação.

F. Miguel Oliveira, Manuel Lemos (*)

(*) Upper Design, S.Paulo, Brasil

Resumo – O propósito deste artigo é o de documentar o desenvolvimento de uma plataforma de *software* que facilite, a utilizadores pouco versados na utilização de ferramentas de edição multimédia, um serviço de composição e publicação de documentos num servidor *WWW*, disponibilizá-los numa interface *Web* sob classificação temática, e apoiar nas buscas relativas aos documentos e outros recursos informativos residentes numa base de dados. Pretende-se propor novas técnicas de distribuição de informação e de aquisição de informação sobre os perfis dos utentes do serviço.

A plataforma de *software* que serve de suporte a este serviço baseia-se na utilização de agentes de *software* devidamente adaptados para o efeito e o seu desenvolvimento foi feito no âmbito do projecto AMIS (Agentes Móveis e Inteligentes) a decorrer no âmbito das actividades do Instituto de Telecomunicações, pólo de Aveiro.

Abstract – The purpose of this paper is to document the development of a software platform that makes easy, for users with few technical knowledge, the composition and publication in a *Web* server, makes documents available in a *Web* interface under theme classification, supports searches related to documents and other info resources in the database. The work is also targeted towards the identification of new techniques for information distribution and for the acquisition of personalized user profiles. The software platform under development is based on appropriately tailored software agents paradigms and is being developed in the framework of the activities of the Institute of Telecommunications, Aveiro branch.

I. INTRODUÇÃO

A *World Wide Web* (doravante designada apenas por *Web*) é actualmente reconhecida como um dos meios mais efectivos para disponibilizar documentos e informação em geral através da Internet. O conteúdo deste documentos é normalmente disponibilizado numa linguagem hipertextual que lhes confere as capacidades de interligação com outros documentos localizados em qualquer parte da plataforma telemática que é a *Web*. Assim, para que essa informação possa ser disponibilizada é necessário que os ficheiros dos documentos (os conteúdos ...) sejam compostos numa hiperlinguagem apropriada para depois serem guardados no disco de um servidor de *WWW* para, a partir daí, o servidor os fornecer a quem os pretender aceder.

Para os autores¹ que produzem documentos destinados a ser disponibilizados através da *Web*, existem várias alternativas para compor e publicar esses documentos. A composição de documentos pode ser feita através de dois tipos de programas: os **editores manuais** e os **editores visuais**.

Os editores manuais são simples programas de edição de texto através dos quais um autor pode definir o conteúdo dos seus documentos directamente através da linguagem *HTML*.

Os editores visuais são programas que permitem ao autor ter visualmente uma ideia do aspecto do documento que está a elaborar à medida que o vai compondo.

Os editores visuais são mais apropriados para autores com poucos conhecimentos técnicos dado que não requerem a aprendizagem prévia da linguagem *HTML*.

No entanto, os editores manuais são preferidos pelos autores mais experientes porque ao permitirem escrever directamente na linguagem *HTML* dão todas as possibilidades a esses autores de tirar o máximo proveito das capacidades desta linguagem.

Depois de compor um documento, um autor necessita de o publicar, ou seja, de enviar o documento de alguma maneira para um servidor. Existem várias maneiras para enviar um documento para um servidor de *WWW*: **correio electrónico** (*E-mail*), **transferência de ficheiros** (*FTP*), **publicação através da *Web*** (*WWW*).

A publicação por correio electrónico não é muito comum porque requer a cooperação de um operador humano ou de um sistema automático que receba mensagens de correio electrónico e delas extraia os ficheiros dos documentos nelas contidos.

A publicação através da transferência de ficheiros para um servidor de *FTP* é mais comum. No entanto requer a existência de um servidor de *FTP* à parte do servidor de *Web* que os vai disponibilizar. Por outro lado, a publicação de documentos através de *FTP* requer por parte dos autores, conhecimento de utilização de programas próprios para transferir os ficheiros dos documentos.

A publicação através da *Web* é o método mais simples para utilizadores com poucos conhecimentos técnicos. A publicação pode ser feita através de uma página *HTML*, destinada a esse efeito, ou através de editores visuais. A

¹ Neste artigo a designação "autor de documento" é utilizada para denotar a pessoa que produz a versão hipertexto do documento, que poderá ser, ou não, o autor do seu conteúdo.

maior parte dos editores visuais existentes já são capazes de publicar documentos através da *Web*. Isto permite a um autor poder compor e publicar documentos a partir de um só programa sem ter que tomar conta da quantidade e dos nomes dos ficheiros que devem ser transferidos para o servidor *Web*.

A. Ferramentas de composição e publicação de documentos em servidores de *Web*

As ferramentas de composição e publicação de páginas mais conhecidas usam as primitivas POST e PUT para publicar documentos em servidores *Web*.

Concretamente, o programa *Microsoft Frontpage* usa a primitiva POST para enviar para o servidor os ficheiros das páginas compostas pelos utilizadores. Para suportar a primitiva usada por este programa é necessário ter instalado extensões no servidor de *Web* de modo a permitir a descodificação dos ficheiros enviados.

Os programas *W3C Amaya* e *Netscape Composer* (parte do programa *Netscape Communicator*) usam a primitiva PUT para publicar os documentos compostos pelos utilizadores. O suporte à primitiva PUT só pode ser proporcionado por um servidor *Web* que suporte o protocolo *HTTP 1.1*, como é o caso do programa servidor *Apache 1.3* que é usado neste projecto.

Como a primitiva PUT é a recomendada pela norma do protocolo *HTTP 1.1* para publicação de documentos em servidores *Web*, é sobre esta primitiva que se irá basear a plataforma de publicação de documentos que foi desenvolvida. Através desta primitiva é possível publicar documentos *HTML*, embora limite o serviço a um único formato de publicação. Além disso, a interface de publicação do *Netscape Composer* obriga à introdução de dados pouco *user friendly* para autores, o que dificulta esta acção a utilizadores com poucos conhecimentos técnicos. Através da experiência acumulada durante algumas semanas de testes com utilizadores piloto, verificámos uma deficiente compreensão da parte de autores em publicar documentos, o que resultava num número insignificativo de *uploads* para o repositório do serviço. Criou-se então uma página no serviço com a funcionalidade de publicar documentos através da *Web* pela primitiva *POST*, mais amigável e funcional do ponto de vista dos autores, que se revelou ser capaz de atrair um considerável volume de documentos publicados. Estas funcionalidades serão descritas mais à frente neste artigo.

De notar também que qualquer dos programas de composição e publicação mencionados que suportam a primitiva PUT, podem ser obtidos gratuitamente para a generalidade das plataformas.

II. ESPECIFICAÇÃO E IMPLEMENTAÇÃO

A ideia base do serviço AMIS é simples: disponibilizar um serviço informativo temático, que permita a publicação a utilizadores registados (autores, i.e., utilizadores que se registam e são aprovados pelo(s)

administrador(es)) de documentos (*Word, PDF, PS, PowerPoint, etc.*) para consulta e apropriação por parte de qualquer utilizador, anónimo ou não. O serviço deve ser dotado de funcionalidades amigáveis e pouco complexas, para publicação, edição, classificação e pesquisa de informação por parte dos seus utilizadores.

Para a criação de uma base de documentos de interesse é necessário desenvolver uma plataforma que permita a composição e a publicação deste tipo de conteúdos num servidor *Web*.

Assim, a plataforma desenvolvida deve cumprir os seguintes requisitos segundo dois pontos de vista: o da **utilização do sistema** e o da **administração do sistema**:

A. Utilização do sistema

A forma de utilização do sistema deve ser fácil de compreender e de usar por utilizadores com poucos conhecimentos técnicos.

A.1 - Base de dados de autores

Todos os autores de documentos têm de estar registados para que possam ser identificados pelo servidor sempre que queiram publicar novos documentos ou alterar os seus documentos anteriormente publicados.

Portanto, existe um sistema que permite o registo de novos autores numa base de dados de modo a ser disponibilizada uma conta de acesso única para cada um.

Para assegurar que cada documento publicado só possa ser alterado pelo respectivo autor, cada autor possui um nome de utilizador único e uma senha de acesso individual e secreta.

No nome de acesso de cada utilizador consta apenas de uma palavra escrita apenas com letras minúsculas (sem acentos e sem cedilhas) baseadas no nome do autor de modo a ser mais fácil de memorizar.

No momento do registo o autor deverá sugerir uma senha de acesso e também escolher uma expressão que lhe permita lembrar a senha quando se tiver esquecido dela.

O autor tem também a possibilidade de alterar a sua senha de acesso mesmo depois da conta de acesso ter sido criada, mas sem necessidade de intervenção de um operador humano.

Cada autor tem uma área de acesso restrito na qual serão colocados os documentos que publica (Fig. 1). Essa área deverá ter o mesmo nome de utilizador do autor de modo a facilitar a localização no servidor *Web* dos documentos que publica.

A.2 - Criação e publicação de documentos

A criação de um documento implica a criação de um ou mais ficheiros que têm de ser enviados para o servidor de *Web* para serem publicados. A publicação de cada documento deverá ser feita através dos programas de composição já mencionados através das primitivas atrás descritas.

Para evitar a eventual confusão de ficheiros de diferentes documentos, é criada uma área distinta para alojar os ficheiros que dizem respeito a cada documento dentro da área de trabalho do respectivo autor (Fig. 2).

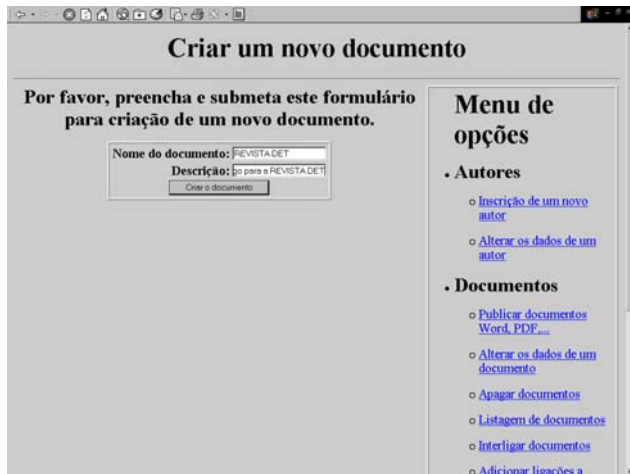


Fig. 2 – Criação de um novo directório (pasta) para publicação de files.

Para tal é necessário que cada autor crie previamente uma área (directório ou pasta) para cada documento que cria antes de o poder publicar (Fig. 3).

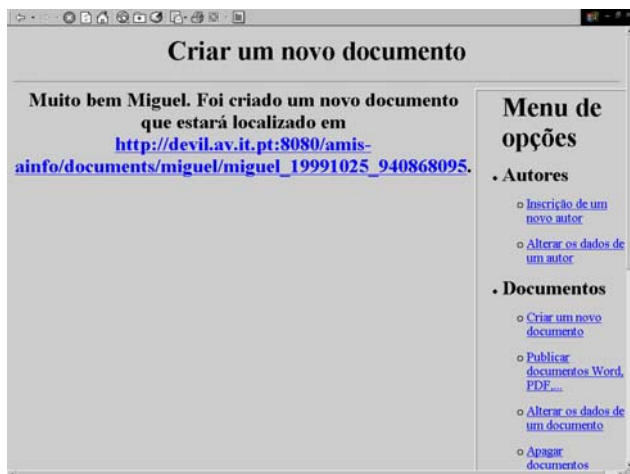


Fig. 3 – Informação relativa à área (URL) criada.

O sistema é dotado de um menu de opções com funcionalidades diversas. O menu encontra-se subdividido em Autores, Documentos, Recursos Informativos e Administração. O sub-menu “Autores” permite a inscrição e alteração de dados dos utilizadores que se querem registar e dos utilizadores registados. O sub-menu “Documentos” é detentor de funções de gestão de documentos (criar, publicar, alterar, apagar e interligar documentos). Além disso o menu possui outros itens de adicionar e procurar outros recursos informativos (*links*, eventos) e funções de administração para gerir e manipular dados de utilizadores e de documentos.

Essa possibilidade deve ser dada a cada autor através de um formulário solicitando a informação necessária, como o título do documento e uma descrição sucinta do seu contexto (Fig. 2). O nome dessa área fará parte do URL do

documento (Fig. 3). Como já foi referido anteriormente, uma das possibilidades de publicação é através de uma página HTML destinada para esse efeito. O autor, através da visualização do disco do cliente, pode seleccionar a(s) *file(s)* que pretende publicar no directório que criou (Fig. 4).



Fig. 4 – Directory browsing ao disco cliente do autor para a publicação de files através da Web.

A.3 - Classificação de documentos publicados

De modo a ajudar a organizar os documentos disponibilizados num servidor, cada documento é classificado pelo próprio autor. Para assegurar que a classificação de todos documentos é sempre feita, todos os documentos não classificados, estão disponíveis numa classe denominada “Sem Classificação” (Fig. 5).

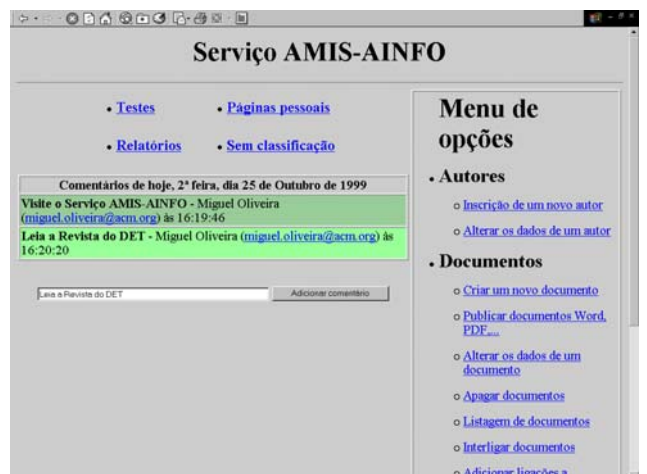


Fig. 5 – Página inicial do serviço. Das várias classes, existe uma denominada “Sem classificação” para documentos não classificados. Esta classe só é visualizada quando existem documentos não classificados.

A classificação de um documento deverá ser feita a dois níveis: uma classificação temática e outra através da

indicação de expressões chave relacionadas com o contexto do documento de modo a ajudar à sua pesquisa.

A classificação temática consiste na indicação de um tema ou mais temas dentro dos quais o assunto do documento se insere, dentro de uma lista de temas pré-definidos. Se não existir nenhum tema apropriado para classificar um documento, o autor poderá sugerir um novo tema.

O autor também poderá classificar um documento como sendo de um sub-tema abrangido dentro de um ou mais temas já existentes expandindo assim uma hierarquia de temas (Fig. 6).

A classificação por expressões chave consiste na indicação de uma ou mais expressões constituídas por uma ou mais palavras que estejam associadas aos assuntos abordados em cada documento (Fig. 6).

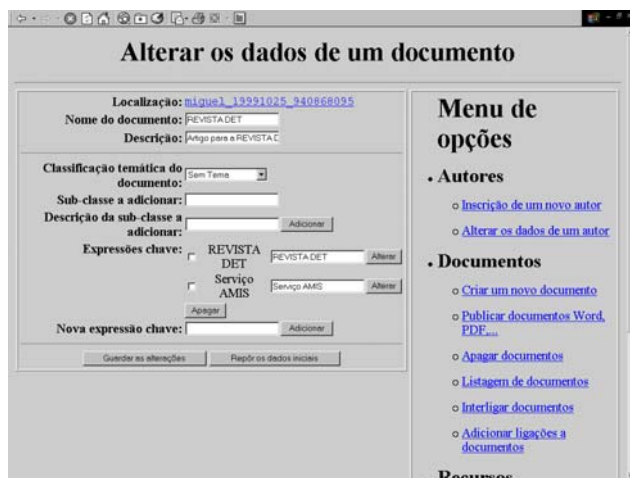


Fig. 6 – Página de alteração/remoção de dados de um documentos. Nesta página o autor pode alterar o nome, descrição, classe temática e expressões-chave do documento.

As expressões chave inseridas são adicionadas à base de dados. Cada expressão chave possui uma ou mais palavras, sendo salvaguardada a expressão e o documento a que pertence. No momento de adição de uma expressão-chave, esta é partida (*tokens* consecutivos entre palavras separadas por espaços) por palavras, e cada palavra é também inserida na base de dados juntamente com o seu valor em *ascii* e chave *soundex* (palavras pronunciadas similarmente podem produzir a mesma sonoridade i.e., o mesmo valor de uma chave *soundex*) [5] (Fig. 7). A salvaguarda dos valores *ascii* e *soundex* de palavras são de extrema importância para pesquisas e para adicionar links a um documento, como veremos a seguir.

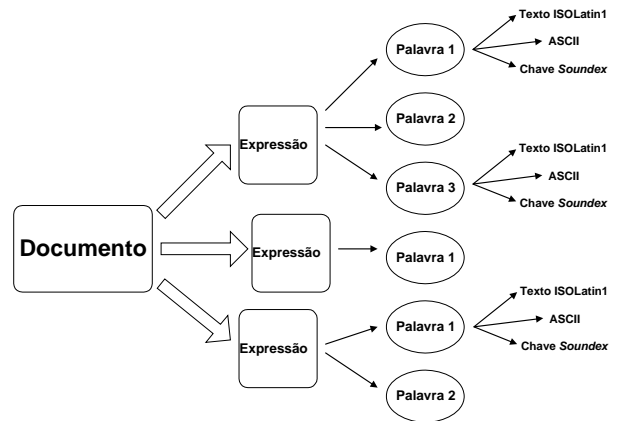


Fig. 7 – Esquema de extração e conversão de expressões e palavras para a base de dados AMIS.

A.4 - Interligação de documentos publicados

Depois de um autor ter classificado um documento que já publicou, este encontra-se disponível para consulta por qualquer utilizador. Mas para que o documento possa ser enriquecido existe uma funcionalidade que permite estabelecer a interligação do documento publicado com outros documentos existentes sobre temas relacionados.

O estabelecimento de *links* entre documentos é feito a pedido do autor de um documento depois deste ter sido publicado.

O sistema analisa os ficheiros seleccionados em formato *HTML* de cada documento e tenta encontrar expressões do documento que foram usadas para classificar outros documentos anteriormente publicados. No caso do serviço não encontrar expressões iguais no documento, converte-as para *ascii* (sem acentos e cedilhas) e determina a sua chave *soundex*, tentando de seguida encontrar na base de dados valores correspondentes. Esta funcionalidade é bastante útil dado que tolera eventuais erros ortográficos.

No caso do serviço não encontrar expressões coincidentes (iguais, *ascii* e *soundex*) de uma expressão procede à partição da mesma em *tokens* consecutivos (Fig. 8).

Expressão a ser partida em várias

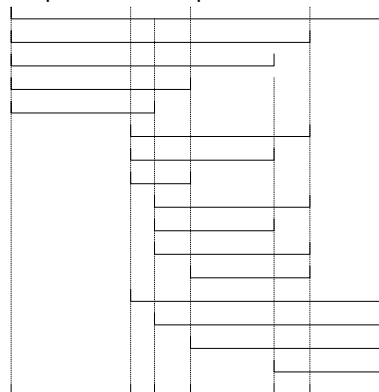


Fig.8 – Esquema de *tokens* utilizados pelo serviço AMIS para encontrar eventuais expressões ou palavras.

O sistema apresenta a lista de expressões encontradas assinalando-as directamente no documento para que o autor possa ver claramente que interligações lhe estão a ser sugeridas (Fig. 9).

O utilizador terá a possibilidade de aprovar ou não qualquer uma das interligações sugeridas (Fig. 9). As interligações aprovadas pelo autor deverão aparecer no documento como qualquer hiperligação num documento de *HTML*, mas o *URL* apontado por essa hiperligação não direcciona imediatamente o *browser* para o documento. Em primeiro lugar verifica a autenticidade da ligação e a sua existência na base de dados. Se esta não for válida informa o autor de que o *link* não está disponível, se for válida redirecciona o *browser* para o *link* a apontar para o documento.

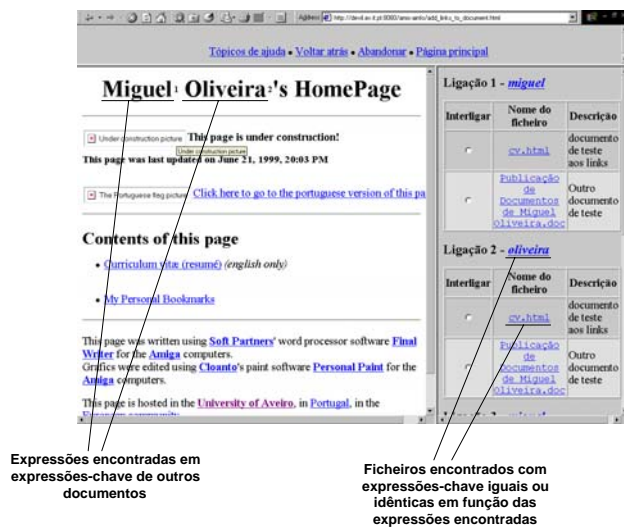


Fig. 9 – Análise da *file* de um documento a interligar. Na *frame* esquerda é visualizada a página com os *links* numerados em *superscript* a adicionar. Na *frame* direita encontram-se tabelas de documentos em função das expressões encontradas.

A.5 - Pesquisa de documentos

Para facilitar a localização de documentos existentes no servidor, foi criada uma página que permite um utilizador efectuar pesquisas sobre a base de dados do serviço, indicando palavras que podem fazer parte das expressões-chave, títulos, descrições ou classificação temática dos documentos existentes.

O procedimento para encontrar resultados na pesquisa é idêntico ao de encontrar expressões no documento para interligar (Fig. 8 – *tokens* consecutivos entre palavras).

Se não forem encontrados documentos com expressões chave iguais às indicadas pelo utilizador, a pesquisa é repetida novamente, tentando encontrar expressões com sonoridade idêntica, de modo a dotar o sistema de pesquisa de alguma tolerância a erros ortográficos que os utilizadores possam cometer.

O serviço tem duas modalidades de pesquisa: pesquisa simples e complexa. A diferença entre as duas modalidades reside no facto de na pesquisa simples o

utilizador não ter a necessidade de seleccionar quais as tabelas da base de dados que pretende pesquisar.

Na pesquisa complexa o utilizador selecciona quais as tabelas da base de dados em que pretende fazer incidir a sua pesquisa (Fig. 10) e, além disso, o utilizador também pode efectuar pesquisas nos motores de pesquisa SAPO (<http://www.sapo.pt>) e CUSCO (<http://www.cusco.pt>). A ferramenta implementada para a pesquisa no SAPO e no CUSCO é idêntica aos serviços de múltiplas pesquisas em listas classificadas e em motores de pesquisa – *socket connections* a servidores destino, guardar os resultados e extrair a informação desejada.

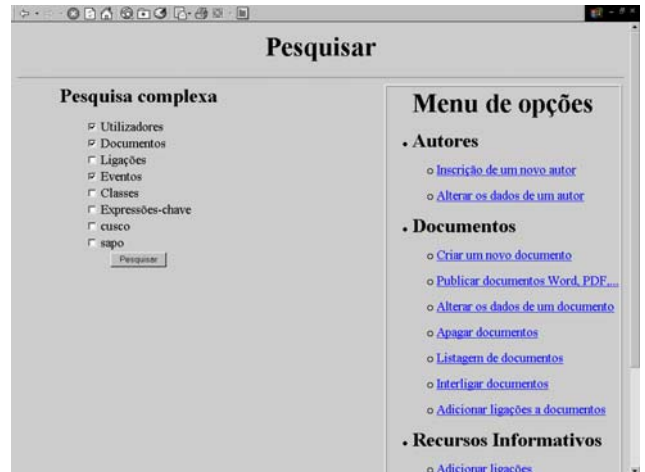


Fig. 10 – Selecção de campos a pesquisar na base de dados do serviço AMIS.



Fig. 11 – Resultados obtidos numa pesquisa.

III. VALIDAÇÃO DO PROJECTO AMIS - REPOSITÓRIO DE INFORMAÇÃO DO DCTR

A. Introdução

Tendo em vista a validação de ideias e propósitos a atingir no projecto AMIS, foi criado um Repositório de Informação – Desenvolvimento Científico Tecnológico e Regional, para apoio ao projecto PRAXIS. A este

repositório de informação foi dada a designação de RI-DCTR.

O RI-DCTR (Fig. 12) apresenta as mesmas funcionalidades do projecto AMIS acrescidos e adaptados aos requisitos do projecto DCTR.



Fig. 12 – Página principal do Repositório de Informação DCTR.

Incidiu-se particularmente sobre a problemática da divulgação da cultura científica usando os recursos proporcionados pela Internet e pelas redes telemáticas. Foi dedicada especial atenção à comunicação entre os profissionais da Educação e Investigação (Ensino Superior e Secundário). O RI-DCTR apresenta-se como um sistema de informação *on-line* desenvolvido na Universidade de Aveiro que fazendo uso dos mecanismos de comunicação proporcionados pela Internet e pelas redes telemáticas, procura incorporar as metodologias clássicas de avaliação de conteúdos de publicações científicas (*peer review*), permitindo o livre acesso à informação armazenada no sistema e também a publicação de documentos, links e eventos de carácter científico por parte de comunidades científicas de interesses comuns.

B. Objectivos do RI-DCTR

A construção de um sistema de informação deve ser capaz de operacionalizar e enquadrar a passagem para o meio envolvente de resultados e da informação construídos, produzidos ou recolhidos pelo próprio projecto. Este sistema de informação é encarado como um instrumento de trabalho essencial para a prossecução das actividades do projecto DCTR: Base de Dados – Sistema de Informação.

Em termos específicos este sistema de informação deverá proporcionar as seguintes funcionalidades [6]:

- Repositório de informação construída, produzida ou recolhida pelo projecto
- Mecanismo de difusão e divulgação dos resultados do projecto.
- Mecanismo de interacção com entidades com interesses afins.

Este sistema de informação será acessível via Internet ficando desta forma disponível a todas as pessoas e instituições com interesses afins aos da problemática do projecto.

IV. POSSÍVEIS FUTURAS IMPLEMENTAÇÕES DO PROJECTO AMIS

A. Personalização do serviço para utilizadores

Apesar dos seus primeiros passos no mundo da indústria de *software*, as tecnologias de personalização começam a ganhar algum relevo no que toca à promoção de serviços adaptados ao perfil de cada utilizador [7].

Embora seja crescente a oferta de serviços na Internet, estes não tem desenvolvido e integrado sistemas que consigam cativar, não só utilizadores com poucos conhecimentos técnicos, como também aqueles que, quando se registam num determinado serviço, desejam um leque de funcionalidades mais ou menos abrangente, pretendem ou tem por hábito aceder aos mesmos assuntos temáticos e que cometem por defeito um conjunto de acções consecutivas quando acedem ao serviço. Assim, é “marginalizado” um certo grupo de possíveis e potenciais utilizadores de determinado serviço. Além disso, muitas empresas que disponibilizam este género de serviços limitam-se a criar sistemas idênticos a outros concorrentes e o panorama nacional não é excepção como se tem vindo a verificar.

A.1 – Registo de gostos, experiências e acções comuns dos utilizadores do serviço AMIS

Pretende-se implementar no serviço AMIS uma base de conhecimento relativa aos vários utilizadores que armazene gostos e preferências dos mesmos e registe as acções mais comuns praticadas por estes.

Do ponto de vista esquemático, um perfil paradigmático para o serviço AMIS deverá ser assim (Fig. 13):



Fig. 13 – Esquema de um possível modelo de personalização do serviço AMIS.

O primeiro passo será a identificação do autor por parte do sistema. Este processo é comum por um nome de acesso e palavra-passe (*login, password*), mas limita o serviço a utilizadores registados (autores). Uma possível

implementação será através de *cookies*, em pormenor na secção seguinte.

Um agente de *software* deve registar na base de conhecimento acções, gostos temáticos e experiências concordantes, ou seja, gestos rotineiros comuns praticados pelo utilizador. O agente serve-se da base de conhecimento para apresentar ao utilizador resultados e ou acções que podem ser novos recursos temáticos ou tarefas rotineiras que este tem por hábito praticar (suponha-se o caso de um utilizador que tem por hábito diário publicar um boletim informativo e ver novos recursos informativos na classe de “comunicação social”. O serviço poderia direccionar-se, quando este se identifica, para a página de publicação de documentos, e ao mesmo tempo apresentar-lhe os últimos recursos informativos dentro da classe “comunicação social” que foram introduzidos desde a última vez que utilizou o serviço).

É nesta linha de pensamento que o agente “aprenderia” com o utilizador. Outro paradigma possível seria o agente aprender as acções do utilizador através de redes neuronais, mas não o vamos fazer porque não se encontra no âmbito deste projecto.

O serviço deve assumir sempre por defeito um perfil geral para todos os novos utilizadores (não autores, autores ou administradores) e deve permitir ao mesmo, definir o seu próprio perfil, no caso de este não desejar um perfil elaborado pelo próprio serviço. Um perfil genérico pré-definido (perfil *default*) é o primeiro passo, ao nível da programação e implementação, para a personalização por parte de um serviço, sendo o ponto de partida para uma robusta e consistente implementação.

Outra possibilidade que poderia ser usada para melhorar este serviço seria a de dar a oportunidade ao utilizador para definir novas áreas de pesquisa que ainda não estavam contempladas, permitindo assim aos sistemas aprender com as necessidades do utilizador.

O trabalho dum possível agente inteligente seria de voltar a analisar a informação classificada existente de modo a determinar que informação corresponde às novas áreas de pesquisa indicadas pelos utilizadores.

Esta tarefa não é trivial, mas com alguma cooperação do utilizador é possível evoluir o sistema de modo que este possa executar pesquisas baseadas em critérios mais personalizados que trariam resultados mais orientados para os interesses dos utilizadores.

Um problema que deriva da necessidade deste tipo de agentes de coleccionar informação sobre o perfil psicológico do utilizador, é o problema da privacidade.

Muitos utilizadores não se sentem confortáveis com a ideia de que estão a ser observados e a ser classificados. Seria sempre necessário informar que os dados recolhidos são para beneficiar o seu tipo de utilização.

Mesmo assim há sempre utilizadores que desconfiam, principalmente porque ignoram até que ponto um agente inteligente os pode ajudar. Um exemplo bem conhecido disso é o facto de que muitos recusam aceitar os *cookies* recebidos dos servidores de *WWW*, normalmente porque não sabem ao certo se é bom ou mau.

Como não compreendem, assumem que se o programa de navegação lhes perguntam se aceitam, é porque pode ser mau e à falta de mais informação negam os *cookies* sem saberem as facilidades e funcionalidades que lhes podem permitir.

A.2 *Cookies como ferramenta de reconhecimento de utilizadores do serviço AMIS.*

Um *cookie* é uma file de texto (não pode exceder os 4K) que é enviada pelo servidor *Web* para ser armazenada pelo *browser*, para que mais tarde possa ser lido do *browser*, na próxima vez que este utilizador “único” volte a navegar no mesmo servidor *Web*.

Os *cookies* são bastante úteis, para permitir a um *browser* ter informação sobre o seu utilizador tal como, último acesso ao servidor, tempo dispendido, preferências do utilizador, etc.. O número limite de *cookies* é de 300 numa pilha *FIFO*.

Os *cookies* são a única forma corrente para um serviço identificar utilizadores anónimos e de “compreender” a sua navegação durante a visita. Além disso dão a possibilidade de oferecer funcionalidades semelhantes às dos utilizadores registados (suponha-se o caso de um utilizador que não deseje registar-se, através de um *cookie*, se o serviço achar conveniente, pode permitir a publicação de documentos ao utilizador anónimo).

Na maioria dos casos um utilizador anónimo continua a ser um utilizador anónimo. Um sistema como o serviço AMIS deve sempre garantir a confidencialidade dos dados do utilizador.

B. *Federação de repositórios e sistema de pesquisa distribuída.*

O crescimento exponencial da Internet implica um maior número de utilizadores e por sua vez a uma maior quantidade de recursos informativos.

Um servidor *Web* que albergue documentos e recursos informativos não deve abarcar uma quantidade “colossal” de informação não só pelas limitações de unidades de armazenamento, nem pela sobrecarga que poderia causar aos vários servidores do sistema. Além destas limitações, as diferenças das redes de comunicação são limitadas e a dependência de *proxies* e *firewalls* atrasa o acesso a determinados *sites*. O potencial do serviço AMIS pode ser ampliado, se minimizarmos estas dificuldades existentes em qualquer infraestrutura de telecomunicações.

Múltiplos repositórios de informação disponibilizavam informação distribuída por vários servidores a pouca ou grande distância física, partilhando recursos informativos e minimizando tempos de acesso (Fig. 14).

Supondo vários repositórios de informação regionais, os utentes do serviço poderiam aceder ao serviço da sua área e, não só acederem a informação desse repositório como também obterem resultados de pesquisas colectados a partir de repositórios através de um sistema de pesquisa distribuída.

A principal vantagem é um serviço com uma federação de repositórios que partilha bases de dados entre os vários sistemas.

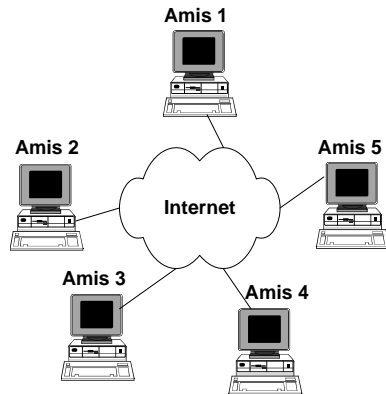


Fig. 14 – Federação de repositórios do serviço AMIS

B.2 Utilização de XML-RPC para pesquisas numa federação de repositórios.

XML-RPC é uma especificação e conjunto de implementações que permite a execução de código RPC (*Remote Procedure Calling*) em vários sistemas operativos e em diferentes ambientes na Internet [8]. XML-RPC pode ser visto como um RPC usando o protocolo HTTP como transporte e XML como código.

Uma mensagem em XML-RPC usa a primitiva *POST* do HTTP, com o corpo da mensagem em XML. Após a execução do pedido num servidor, são retornados resultados em XML. A principal vantagem deste protocolo reside no facto dos pedidos e respostas terem como formato XML, compreendido por computadores e por programadores, que torna mais fácil a elaboração de código e interpretação de resultados.

Para testar a eficiência de um sistema de pesquisa distribuída em repositórios de informação, uma das possibilidades de implementação será com XML-RPC para procura e pesquisa de informação nos vários repositórios. Após uma fase inicial, poderão ser implementados agentes móveis para minimizar os tempos de resposta.

Definindo um *standard* de comandos e padrões de comunicação baseado em XML para a partilha de informação permitiria que agentes comunicassem entre servidores (ou mesmo entre agentes) minimizando tempos de acesso e garantindo o acesso à informação distribuída.

V. CONCLUSÕES

O sistema de informação desenvolvido disponibiliza um repositório de informação a autores que desejem publicar documentos e adicionar outros recursos informativos e permite o acesso aos mesmos por utilizadores anónimos.

Para *hardware* de pequeno porte, torna-se necessário não sobrecarregar o serviço existente, desperdiçando tempo de acesso e capacidade de armazenamento. Além disso, é de

interesse para qualquer serviço que disponibiliza informação, tanto no campo empresarial como no ramo académico, a expansão do mesmo para um maior número de possíveis interessados em publicar ou adquirir informação, sejam utilizadores anónimos ou registados.

Este artigo descreveu, por ordem cronológica, os passos de elaboração e planeamento do serviço, a execução de ideias e do projecto proposto, um projecto piloto como ambiente de validação do serviço e, por último, futuras implementações e possíveis paradigmas que se encontram em estudo para um melhor desempenho e eficiência do serviço.

VI. AGRADECIMENTOS

Os autores deste artigo agradecem a todos aqueles que colaboraram neste projecto, em especial ao Prof. Dr. A. Manuel de Oliveira Duarte, Eng. Amaro de Sousa, Vasco Lagarto e Fausto Carvalho da PT Inovação, à equipa do projecto DCTR em especial ao João Tedim, ao Eng. Jorge Ferraz de Abreu, ao Eng. Fernando Azevedo, ao César Miranda e ao Miguel Jardim.

VII. REFERÊNCIAS

- [1] T. Berners-Lee, R. Fielding and H. FryStyck, "Hypertext Transfer Protocol – HTTP/1.0", RFC 1945, Network Working Group, SRI International, May 1996
- [2] R. Fielding, J. Gettys, J. Mogul, H. FryStyck and T. Berners-Lee, "Hypertext Transfer Protocol – HTTP/1.1", RFC 2068, Network Working Group, SRI Standards Track, January 1997
- [3] N. Freed and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME)", RFC 2045, RFC 2046, RFC 2047, RFC 2048, RFC 2049, Network Working Group, SRI Standards Track, November 1996
- [4] E. Nebel and L. Masinter, "Form-based File Upload in HTML", RFC 1867, Network Working Group, SRI Experimental, November 1995
- [5] Sitg Sæther Bakken, Alexander Aulbach, Egon Schmid, Jim Winstead, Lars Torben Wilson, Rasmus Lerdorf and Zeev Suraski, "PHP3 Manual", PHP Documentation Group, 1999-10-19
- [6] Prof. Dr. A. Manuel de Oliveira Duarte, "Explicação da Actividades 4 (Sociedade da Informação: acesso à informação – incidência na investigação, escola, público, empresas) e 7 (Base de Dados — Sistema de Informação) do projecto DCTR", Abril de 1999
- [7] Richard Murch and Tony Johnson, "Intelligent Software Agents", Prentice Hall PTR, 1999
- [8] "XML-RPC Specification", UserLand Software, Inc., Semptember 1999