

MITRA: Uma solução para serviços de pesquisa em *intranets*

Jorge Machado¹, José Borbinha²

¹Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Portalegre, Lugar da Abadessa, Apartado 148
7301-901 Portalegre, Portugal
jmachado@ext.bn.pt

²INESC-ID – Instituto de Engenharia de Sistemas e Computadores, Rua Alves Redol 9
Apartado 13069, 1000-029 Lisboa, Portugal
jlb@ist.utl.pt

Resumo. Este artigo descreve o sistema MITRA, uma solução para indexação de conteúdos em linha e metadados descritivos complementares codificados em qualquer esquema XML. Esta capacidade torna este sistema uma solução ideal para serviços especializados de pesquisa em intranets. O MITRA baseia-se numa arquitectura com cinco camadas. A primeira camada é a de recolha de conteúdos que pode ser implementada por sistemas externos ou sistemas especializados de transferência de recursos, como por exemplo arquivos locais estruturados. A segunda camada cria índices invertidos dos conteúdos e dos metadados recolhidos (usando o sistema LUCENE). A terceira camada gere as relações semânticas e as associações dos metadados aos recursos. . Uma quarta camada muito recente permite a implementação de uma metodologia de análise do domínio. A última camada, a de apresentação, permite receber pesquisas estruturadas em pedidos HTTP e responder em XML ou HTML conforme sejam ou não utilizadas XSL's. O esquema de representação interna dos metadados é o Dublin Core, o qual permite ao MITRA fornecer naturalmente uma interface de SRU/SRW, mas outros esquemas podem ser também configurados. O MITRA combina assim o poder da indexação livre de conteúdos com o poder do processamento de metadados estruturados, oferecendo o melhor dos dois mundos. Esta solução é usada como suporte a vários serviços efectivos, reportados no texto.

Introdução

É normal uma organização publicar hoje em dia conteúdos através de múltiplos serviços e portais. Neste artigo descreve-se o MITRA, uma ferramenta para indexar espaços dessa natureza e fornecer serviços de pesquisa e análise desse mesmo espaço ou domínio. Este sistema pode indexar conteúdos publicados em linha de forma livre (“sítios” HTML) ou conteúdos estruturados aos quais podem ser associados quaisquer esquemas de metadados codificados em XML. Esta solução é usada como suporte a vários serviços efectivos no contexto da Biblioteca Nacional.

Jorge Machado¹, José Borbinha²

Este artigo prossegue com uma explicação do que se entende por recursos e colecções no contexto do MITRA, e que implicações isso traz na criação dos índices.

O artigo prossegue com a descrição da arquitectura do MITRA. De seguida descrevemos as técnicas usadas para otimizar as pesquisas e as apresentações de resultados, assim como a forma como os metadados são utilizados para melhorar essas apresentações. Em sequência explica-se que serviços remotos são fornecidos pelo MITRA para facilitar a interoperabilidade com outros sistemas.

Por fim propomos uma metodologia suportada pelo MITRA para análise de contextos, onde apresentamos um caso de estudo da revista DLIB e a forma como se está a desencadear a análise da mesma. Esta capacidade é uma forma inovadora de usar estes sistemas, podendo suportar processos de extracção de conhecimento em Intranets ou colecções de conteúdos concretas.

Colecções e Recursos no MITRA

O MITRA destina-se a indexar sítios perfeitamente delimitados e estruturados. Para o MITRA cada sítio é um recurso, que pode ser associado não só com metadados descritivos ou estruturais de codificados em XML, só mas também com colecções.

No MITRA uma colecção é um conjunto de recursos que pertencem a um espaço virtual bem definido. As colecções no MITRA não são no entanto factor restritivo em nenhum aspecto. Em termos de pesquisa o agrupamento de recursos que se faz com uma colecção pode ser feito com qualquer outro índice vindo ou não de um esquema XML. No entanto pensamos ser importante introduzir as colecções para marcar uma posição mais forte que um mero índice que pode ser invocado numa pesquisa. As colecções têm a associados vários dados que de alguma forma podem vir a influenciar a arquitectura futura do MITRA. Exemplos disso são a localização do INDICE, sendo que no futuro o MITRA poderá vir a ter múltiplas camadas de indexação, usando assim sistemas de indexação diferentes para cada colecção. A colecção está integrada no esquema de pesquisa e resposta do MITRA sem ter que estar especificamente dentro da *query* efectuada, etc. Cada recurso é identificado por um URN e pode estar associado a várias descrições na forma de metadados. Quando uma pesquisa é lançada o MITRA retorna os resultados organizados em grupos de recursos (*clusters*) nos quais se podem identificar os metadados, a colecção, a informação de estado e a informação básica acerca do resultado principal dentro de um recurso (Figura 1).

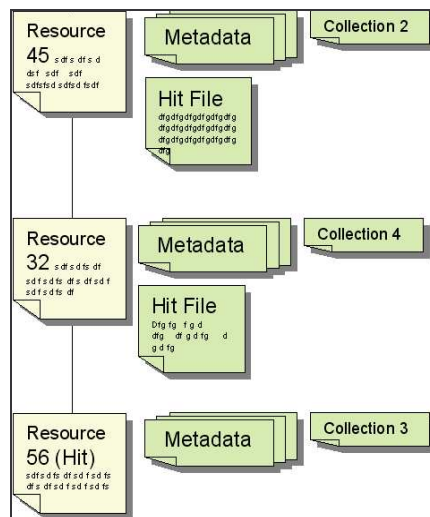


Figura 1 – Resposta estruturada em XML

Um exemplo poderá ser visto no serviço de Busca da Biblioteca Nacional [2], bastando para isso aceder à seguinte ligação (como neste caso nenhuma opção XSL é especificada, o resultado será retornado em XML):

<http://busca.bn.pt/Handler?verb=search&query=lusiadas>

As colecções permitem ao MITRA dividir os seus alvos de trabalho por áreas distintas o que lhe confere um ainda maior poder de pesquisa. Assim como nos restantes casos é criado um índice especial para identificar os documentos de cada colecção sendo que facilmente se filtra uma pesquisa.

Arquitectura do MITRA

O sistema MITRA tem uma arquitectura de cinco camadas codificadas em JAVA, tal como se mostra na Figura 2.

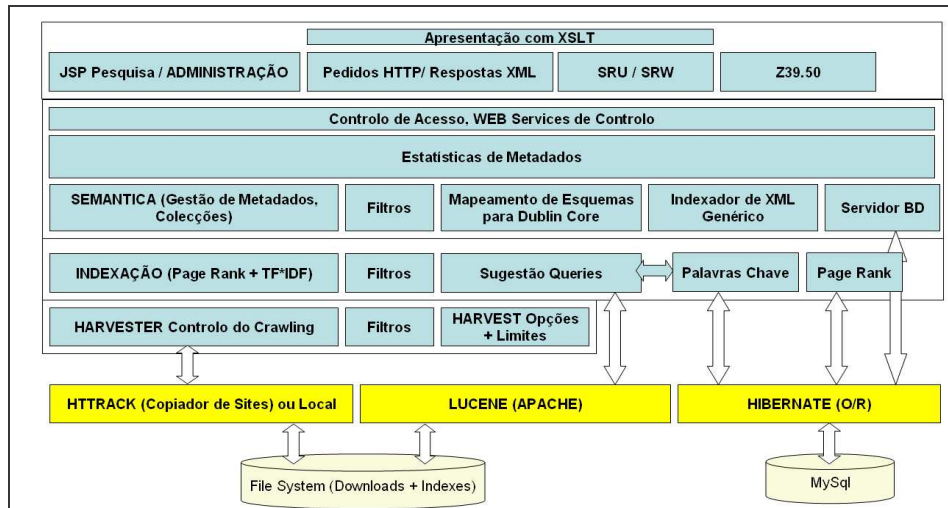


Figura 2 – Arquitectura do MITRA.

Neste sistema reutilizam-se vários componentes e soluções já existentes, conforme se pode constatar na descrição da arquitectura no ponto seguinte. Por exemplo, a camada de indexação usa o LUCENE [10] para a criação de índices, enquanto que a camada de recolha pode usar o HTTRACK [8] para recolha de recursos em linha. Podem ser desenhados ainda componentes de interfaces especializados para arquivos locais de conteúdos estruturados. No caso do serviço BUSCA [2] da Biblioteca Nacional, as descrições bibliográficas codificadas em XML são recuperadas da PORBASE [18], em esquemas como o UNIMARC [27], que é depois convertido em Dublin Core [25]. Isto permite, por exemplo, pesquisar a Biblioteca Nacional Digital [1] usando não apenas pesquisa no texto livre mas também sobre índices estruturados construídos a partir das descrições bibliográfica, tais como autor, assunto, título, data de publicação, resumo, etc.

O sistema compreende ainda uma camada semântica para a gestão dos metadados que cada serviço possa fornecer ao MITRA.

Fluxo de trabalho do MITRA

A Figura 3 ilustra a sequência de processos no fluxo de trabalho do MITRA.

Quando um pedido de indexação é feito pela primeira vez, a camada de controlo de acesso valida o pedido, que é depois passado à camada semântica para esta criar, para o recurso a indexar, um identificador na forma de um URN. Com esse identificador um cliente poderá invocar o MITRA sempre o desejar, com pedidos de reindexação ou, por exemplo, de adição de metadados descritivos.

Os pedidos de indexação pendentes são passados, de forma assíncrona, à camada de Indexação, que por sua vez os passa à camada de *Harvesting*.

Jorge Machado¹, José Borbinha²

também teríamos de usar o *namespace* no nome do índice para garantir que não haveria nomes de índices repetidos.

Saindo um pouco da camada de indexação seria interessante de referir que o MITRA permite a adição de metadados em qualquer altura do ciclo de vida de um recurso. Quando isso é feito a indexação de metadados é chamada assincronamente ao resto do fluxo de indexação uma vez que é completamente independente.

Voltando à camada de indexação estes índices criados com base nos metadados permitem a construção de interfaces de pesquisa avançada Figura 4 como podemos ver na Figura 4 em relação ao serviço BUSCA da Biblioteca Nacional Digital.

Desta forma a camada de indexação coloca todos os conteúdos, estruturados ou livres, em vários índices sendo alguns deles muito conhecidos e usados noutros motores de busca, como por exemplo o *site*, que guarda todos os URL's indexados, a data de última modificação, etc.

Processo de Interrogação no MITRA

Todos os índices podem ser usados na construção de perguntas ao MITRA, tendo a linguagem de interrogação (*query language*), que é muito semelhante à do Google, a seguinte sintaxe que vem por defeito com o LUCENE:

Índice1:([NOT] <texto a encontrar>) [[AND OR NOT] [ÍndiceN:(<texto a encontrar>)]]
--

Apresentamos de seguida alguns exemplos do índice URN e de possíveis perguntas, todas equivalentes, aos índices:

<i>urn:MITRA.bn.pt:1</i> – <i>http://site1/index.html, http://site1/public/copy1/i.htm, ...</i> <i>urn:MITRA.bn.pt:2</i> – <i>http://site2/index.html, http://site2/public/copy1/i.htm, ...</i>
--

Pergunta: <i>URN:urn:MITRA.bn.pt:1 AND (contents:software for xml)</i> Pergunta: <i>site:http://site1 (contents:software for xml)</i> Pergunta: <i>site:http://site1 software for xml</i> Pergunta: <i>dc.creator:Saraiva AND mitologia</i>
--

O MITRA usa diversos índices que resultam da interpretação daquilo que se está a indexar, sendo que muitos deles resultam de *parsing* de HTML e outros da estrutura do material indexado. Exemplos disso são os índices *site*, *sumary*, *comments*, *image*, *imagesize*, e todos os índices resultantes da indexação de metadados.

Finalmente, na ordenação dos resultados podem-se utilizar várias formas de medição de qualificação de resultados face a uma pesquisa algoritmo de análise de ligações usado no Google, o *Page Rank* [17], ou o cálculo de *rank tfidf* [Salton88] para descobrir o peso de um documento face a uma pesquisa.

MITRA: Uma solução para serviços de pesquisa em intranets

Serviço de **BUSCA** da BN

Language/ Língua: Português

Pesquisa Avançada

Título

Autor

Colaborador

Data

Editor

Assunto

Descrição

Texto livre

Com **todas** as palavras

Com a **frase exacta**

Com **pelo menos** uma das palavras

Sem as palavras

SITE (ex: purl.pt/1)

Pesquisar

🔍 Tudo (73865 ficheiros)

- 📖 Biblioteca Nacional Digital (<http://bnd.bn.pt> e <http://purl.pt> - 56592 ficheiros)
- 🏛️ Portal institucional da BN (<http://www.bn.pt> - 130 ficheiros)
- 📄 Portal da cooperação PORBASE (<http://www.porbase.org> - 78 ficheiros)
- 📄 Serviço de teses digitais (<http://dited.bn.pt> - 11388 ficheiros)

Figura 4 – Interface de pesquisa avançada recorrendo aos diferentes índices.

Serviços WEB, Interoperabilidade e Pesquisa

Quer a interface administrativa, quer a interface de pesquisa foram desenvolvidas para permitir pedidos humanos e de outras máquinas. Assim o MITRA pode ser totalmente administrado remotamente por outros sistemas como é o caso do serviço DiTeD [5], o repositório de Teses e Dissertações Digitais da Biblioteca Nacional Digital, que o usa como motor de busca interno para indexar as suas colecções. Essa interface é baseada em pedidos POST e respostas em XML, para as quais foi desenvolvido um cliente JAVA para distribuição. Além desta interface o MITRA apresenta outras interfaces standard como a SRU ou SRW [21] e [22], sendo que a SRU/SRW, que significa Search and Retrieve via URL ou via *WEB Service* é um *standard* muito utilizado para executar *queries* CQL (*Common Query Language*) [3] e responder num esquema bastante simples contendo normalmente uma lista de registos Dublin Core como resultados mas não obrigatoriamente (Figura 5).

As interfaces WEB simulam todos os serviços do MITRA, públicos e administrativos, para facilitar a interacção com outros sistemas.

No caso da interface pública de pesquisa o MITRA define uma XSLT pela qual os resultados são passados produzindo HTML. Neste passo é tirado grande partido dos metadados para garantir que os resultados são explícitos, tal como mostrado na Figura 7. Na Figura 8 mostra-se uma comparação com a mesma pesquisa dentro da BND mas agora com o sistema Google.



Figura 5 – Interface WEB para o servidor de SRU do serviço BUSCA (http://busca.bn.pt/jsp/sru.jsp)



Figura 6 – Interface WEB para os Serviços

MITRA: Uma solução para serviços de pesquisa em intranets

Pesquisar: ☒ tudo ☐ BND ☐ BN ☐ PORBASE ☐ DITeD

Pesquisar por: [Pesquisa Avançada](#)

... tudo Results: 0 - 20 de cerca de 38 para **lusíadas ilustrações**.


[\[Os Lusíadas\]: \[ilustrações\]](#)

Autor: Bramtot, Alfred, 1852-1894

Data/Tipo: D.L. 1976 / material gráfico a duas dimensões

Descrição: Miscelânea

>>>>> Biblioteca Nacional Digital - [Os Lusíadas], D.L. 1976 - ?xml version="1.0" encoding="iso-8859-15"?> >>>>> Biblioteca Nacional Digital - [Os **Lusíadas**], D.L. 1976] Obra [Os **Lusíadas**], D.L. 1976] Ficha ...] > >BRAMTOT, Alfred, 1852-1894 >[Os **Lusíadas**] [Visual gráfico : [\[ilustrações\]](#)]. - [S.l.: s.n., D.L. 1976




[\[Os Lusíadas\]: \[ilustração\]](#)

Autor: Gameiro, Alfredo Roque, 1864-1935

Data/Tipo: [1900] / material gráfico a duas dimensões

Descrição: Monografia Atribuição de autor segundo nota ms. no verso II. da ed. da Empresa da História de Portugal, 1900

Biblioteca Nacional Digital - [Os Lusíadas], [1900] - Biblioteca Nacional Digital - [Os **Lusíadas**], [1900] Obra [Os **Lusíadas**], [1900] Ficha bibliográfica (visualização ISBD) [704429] GAMEIRO, Alfredo Roque, 1864-1935 [Os **Lusíadas**] [Visual gráfico : [\[ilustração\]](#)], [1900]. - 1 desenho : tinta-da-china



[\[Os Lusíadas\]: \[ilustração\]](#)

Autor: Macedo, Manuel de, 1839-1915 , Gameiro, Alfredo Roque, 1864-1935

Data/Tipo: [1900] / material gráfico a duas dimensões

Descrição: Monografia Atribuição de autores segundo nota ms. no verso II. da ed. da Empresa da História de Portugal, 1900

Biblioteca Nacional Digital - [Os Lusíadas], [1900] - Biblioteca Nacional Digital - [Os **Lusíadas**], [1900] Obra [Os **Lusíadas**], [1900] Ficha bibliográfica (visualização ISBD) [979522] MACEDO, Manuel de, 1839-1915 [Os **Lusíadas**] [Visual gráfico : [\[ilustração\]](#)], [1900]. - 1 desenho : tinta-da-china




Figura 7 – Pesquisa por “lusíadas ilustrações” no BUSCA

Web [Imagens](#) [Grupos](#) [Notícias](#) [Desktop](#)

Google [Pesquisa Avançada](#) [Preferências](#)

Pesquisar: ☒ a web ☐ páginas escritas em Português ☐ páginas de Portugal

Web Resultados de 1 a 7 de cerca de 11 de **purl.pt** correspondente a **lusíadas ilustrações**. (0,16 segundos)

[\[Os Lusíadas\]: <\[>ilustrações\], D.L. 1976\] - Biblioteca Nacional ...](#)

[Os **Lusíadas**] [Visual gráfico : [\[ilustrações\]](#)]. - [S.l.: s.n., D.L. 1976]. - 20 imagens : color. ; 33x24 cm. - Repr. das il. da edição de Paris, de 1890 ...

[purl.pt/12345 - 7k](#) - [Em cache](#) - [Páginas semelhantes](#)

[\[Os Lusíadas\]: <\[>ilustração\], \[1900\] - Biblioteca Nacional Digital](#)

[Os **Lusíadas**] [Visual gráfico : [\[ilustração\]](#)] [1900]. - 1 desenho : tinta-da-china e aguarela ; 35x24 cm [http://purl.pt/1229 ...](#)

[purl.pt/1229 - 7k](#) - [Em cache](#) - [Páginas semelhantes](#)

[BIBLIOTECA NACIONAL DIGITAL // : Ilustradores de Quixote : O quixote](#)

No domínio das versões, com **ilustrações** de artistas portugueses, ... tinha feito publicar Os **Lusíadas**, a obra máxima no género da épica renascentista.) ...

[purl.pt/920/1/quixote-pt-2.html - 32k](#) - [Em cache](#) - [Páginas semelhantes](#)

Figura 8 – Pesquisa por “site:purl.pt lusíadas ilustrações” no Google

Suporte ao estudo de ambientes controlados

Ainda relativamente à Figura 3, o processo “Análise de Texto (*tokens*)” está destacado porque pretendemos chamar a atenção para a sua especial importância para a metodologia apresentada de seguida.

A quarta camada do MITRA dá suporte à análise dos espaços indexados. Se esta for correcta, permite a extracção de conhecimento que poderá ser uma mais valia para outros processos de análise ou de decisão. O problema que se coloca é como analisar o domínio que estamos a indexar. Os conteúdos digitais de uma organização são uma fonte rica de informação que bem trabalhada poderá fazer surgir conhecimento escondido. O grande problema é que os sistemas de informação apenas extraem conhecimento de dados estruturados e muitas vezes apenas em bases de dados relacionais. Pedir mais que isto é ainda hoje pedir muito. O método apresentado de seguida para a análise de domínios com o suporte do MITRA está a ser testado pela primeira vez num estudo aos conteúdos da revista D-LIB [6], a qual tem acessíveis todos os artigos desde 1996, estruturados por mês e ano. Achamos no entanto que este método pode ser generalizado a qualquer domínio.

Como fazê-lo?

O grande objectivo do processo é encontrar os *termos chave* do domínio, compostos por uma ou mais palavras. Para isto, problemas como por exemplo plurais ou sinónimos têm de ser considerados. Para lidar com esta problemática de identificação de palavras-chave criámos no MITRA o conceito *keyword* que pode ser composto por um ou mais *termos chave*, podendo estes conter uma ou mais palavras. Por outro lado temos o índice de palavras existentes no domínio. O objectivo final será maximizar a lista de *termos chave* e minimizar o índice de palavras soltas do índice de palavras. Desta forma vamos identificar todos os objectos importantes do nosso universo de estudo.

Para atingir esse objectivo o sistema de suporte existem mais duas listas que podem ser definidas no MITRA, as *stop-words* e as palavras *irrelevantes*. As palavras *irrelevantes* são palavras soltas que estão numa situação intermédia em que não se tem a certeza se não poderão vir a ser *keyword*. As *stop words* são as palavras descartadas que já não conseguem formar qualquer *keyword*.

Para gerir este processo foi desenvolvido um serviço com a interface que se mostra na Figura 9.

Add new Term

Name [Add](#)

Terms

Checked: [Change](#) | [Delete](#) | [Create new Group](#)

	Name	
<input type="checkbox"/>	Digital Signature	Delete Change
<input type="checkbox"/>	Digital Signatures	Delete Change
<input type="checkbox"/>	dublin core	Delete Change
<input type="checkbox"/>	institutional repositories	Delete Change
<input type="checkbox"/>	institutional repository	Delete Change
<input type="checkbox"/>	intellectual properties	Delete Change
<input type="checkbox"/>	intellectual property	Delete Change
<input type="checkbox"/>	oai	Delete Change
<input type="checkbox"/>	oai-pmh	Delete Change

KeyWords

Join Terms: Check Them and click in the button ">>" of choosed group

	Name	
<input type="button" value=">>"/>	digital repositories	Remover do Grupo
<input type="button" value=">>"/>	digital repository	Remover do Grupo
<input type="button" value=">>"/>	dublin core	Remover do Grupo
<input type="button" value=">>"/>	institutional repositories	Remover do Grupo
<input type="button" value=">>"/>	institutional repository	Remover do Grupo
<input type="button" value=">>"/>	intellectual properties	Remover do Grupo
<input type="button" value=">>"/>	intellectual property	Remover do Grupo
<input type="button" value=">>"/>	oai	Remover do Grupo
<input type="button" value=">>"/>	oai-pmh	Remover do Grupo

Figura 9 – Interface para análise de texto.

Método de Análise

A análise do domínio terá de ser feita de forma interactiva e com intervenção humana de especialistas. O MITRA disponibiliza no entanto algumas rotinas que permitem a criação de listas de possíveis candidatos a *keywords* combinando as palavras do índice duas a duas e três a três, ordenando-as pelo número de referências ao termo composto.

Assim a cada interacção descobrem-se mais *keywords* e reindexa-se o domínio. Nesse processo o analisador de texto vai descartando as *stopwords* e o universo de palavras vai diminuindo e facilitando cada vez mais o trabalho de análise. O Analisador de Texto (Figura 3) é uma peça chave que se utiliza em todo lado na camada de indexação, incluído na análise das *queries* para as pesquisas. Ele limpa o texto de *stop words*, letras maiúsculas, diacríticos, faz o *stemming* e neste caso normaliza também as *keywords* substituindo as ocorrências dos termos pelo código da *keyword* respectiva.

Recuperação da Informação na D-LIB

Construída a base de *keywords* cada domínio terá os seus requisitos de conhecimentos, os quais irão requerer módulos de recuperação diferentes.

No caso da D-LIB todos os conteúdos estão organizados de forma cronológica. Para auxiliar o processo foram criados dinamicamente descrições de metadados, onde entre outras coisas, está descrito o Mês e o Ano de cada artigo. Essas descrições especialmente criadas para descrever uma estrutura que se evidenciava numa

Jorge Machado¹, José Borbinha²

disposição de arquivos locais, foram associadas aos registos, juntamente com as descrições em Dublin Core.

Combinado o índice de *keywords* com os índices de informação estruturada como os autores, assuntos ou títulos, poderemos chegar a informação bastante relevante para a análise das bibliotecas digitais e sua evolução. Um exemplo são as datas de publicação que nos podem ajudar a descobrir quando é que determinadas tecnologias ou assuntos foram mais abordados, quem foram os autores que os abordaram em primeira instância, ou aqueles que mais abordaram determinados temas.

A parte desta informação se considerarmos fazer dos nomes de autores *keywords* onde criamos um conjunto de todas as assinaturas de determinada pessoa podemos ainda extrair listas de artigos por autor, por tema, por data, etc.

Por observação gráfica do número de referências a certas *keywords* podemos por exemplo perspectivar os temas que irão ser mais abordados nos próximos anos e quais estão em decadência como se pode visualizar na Figura 10.

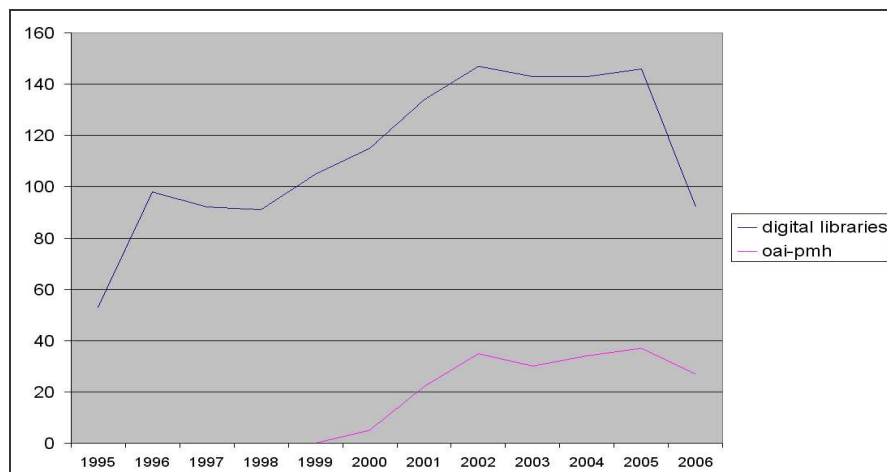


Figura 10 – Exemplo da contagem de referências a duas *keywords* na revista D-LIB

Para além de tudo isto poderemos ainda filtrar resultados usando apenas determinados autores em casos que se verifique quem são os melhores em determinados assuntos de forma a melhorar os resultados, etc.

Conclusões e Trabalho Futuro

Este artigo descreve um simples e efectivo motor de busca desenvolvido para gerir colecções de recursos estruturados em linha. Esta solução é usada como suporte ao serviço de pesquisa genérico na Biblioteca Nacional (<http://busca.bn.pt>) e ainda nos serviços de pesquisa especializados do Repositório de Teses da Biblioteca Nacional (<http://dited.bn.pt>). Foi criado ainda um serviço experimental de indexação da revista

DLIB (recurso localizado em <http://www.dlib.org>, e serviço em <http://MITRA.bn.pt/dlib>).

O serviço na BN reconhece neste momento esquemas de metadados, os quais são usados em módulos de filtragem semântica, melhoria da apresentação dos resultados, e suporte à pesquisa avançada. Entre esses esquemas destacam-se os formatos MARCXML, Dublin Core, METS e DEPTAL. Com as descrições em METS, por exemplo, conseguimos descobrir por exemplo os objectos representando obras digitalizadas que possuem conteúdos resultantes de processos de OCR automáticos (texto em ficheiros) e associar esses conteúdos às respectivas imagens digitalizadas, permitindo assim oferecer um efeito de “pesquisa de texto em imagens”. Finalmente, as funções de análise de domínios controlados estão a ser usadas em projectos de caracterização de colecções e espaços, o que se espera venha a produzir resultados bastante interessantes, a reportar no futuro.

Referências

- [1] BND. Biblioteca Nacional Digital. <<http://bnd.bn.pt>>
- [2] BUSCA. <<http://busca.bn.pt>>
- [3] CQL (Common Query Language) <<http://www.loc.gov/standards/sru/cql/index.html>>
- [4] DEPTAL. Repositório Institucional. <<http://deptal.bn.pt>>
- [5] DiTeD. Digital Thesis and Dissertations. <<http://dited.bn.pt>>
- [6] D-LIB. <<http://www.dlib.org>>
- [7] HIBERNATE. <<http://www.hibernate.org/>>
- [8] HTTRACK. <<http://www.httrack.com>>
- [9] LDAP. <<http://www.openldap.org/>>
- [10] LUCENE. Apache LUCENE. <<http://lucene.apache.org/>>
- [11] MARCXML. MARC 21 XML Schema. <<http://www.loc.gov/standards/marcxml/>>
- [12] METS. Metadata Encoding and Transmission Standard. <<http://www.loc.gov/standards/mets>>
- [13] MITRA. <<http://mitra.bn.pt>>
- [14] MYSQL <<http://www.mysql.com/>>
- [15] Nyna Plat, Search Engines for Intranets <<http://www.llrx.com/features/nina.htm>>
- [16] OAI-PMH. Open Archives Initiative - Protocol for Metadata Harvesting <<http://www.openarchives.org/>>
- [17] Page Rank <<http://www.google.com/technology/>>
- [18] PORBASE. Base Nacional de Dados Bibliográficos. <<http://www.porbase.org>>
- [19] PURL sistema de gestão dos recursos digitais da BND. <<http://purl.bn.pt>>
- [20] SCORM. The Sharable Content Object Reference Model. <<http://www.adlnet.org/index.cfm?fuseaction=scormabt>>
- [21] SRU Search and Retrieve URL. <<http://www.loc.gov/z3950/agency/zing/srw>>
- [22] SRU/SRW <<http://www.loc.gov/standards/sru/>>
- [23] SRU/SRW schema <<http://www.loc.gov/standards/sru/xml-files/srw-types.xsd>>
- [24] STRUTS. The Apache STRUTS Web Application Framework. <<http://struts.apache.org/>>
- [25] The Dublin Core Metadata Initiative <<http://dublincore.org>>
- [26] The Google Pagerank Algorithm and How It Works, Ian Rogers, IPR Computing Ltd. <<http://www.iprcom.com/papers/pagerank/>>
- [27] UNIMARC <<http://www.unimarc.info>>
- [28] Z39.50. Z39.50 Maintenance Agency. <<http://www.loc.gov/z3950/agency/>>