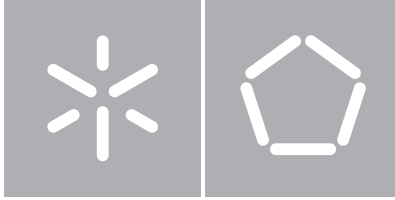


Universidade do Minho
Escola de Engenharia

Nuno Antunes Marques

Gestão de depósitos em arquivos



Universidade do Minho

Escola de Engenharia

Departamento de Informática

Nuno Antunes Marques

Gestão de depósitos em arquivos

Dissertação de Mestrado

Mestrado de Informática

Trabalho realizado sob orientação de

Professor Doutor José Carlos Leite Ramalho

Declaração

Nome

Nuno Antunes Marques

Endereços Electrónicos

pg17475@alunos.uminho.pt

namrks@gmail.com

Título da Dissertação

Gestão de depósitos em arquivos

Orientador

Professor Doutor José Carlos Leite Ramalho

Ano de Conclusão

2012

Designação do Mestrado

Mestrado de Informática

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 31/10/2012

Assinatura: Nuno Antunes Marques

Agradecimentos

Gostaria de agradecer ao meu orientador, Professor Doutor José Carlos Leite Ramalho pela orientação, motivação e apoio prestado ao longo deste trabalho.

Agradeço à empresa KEEP Solutions pela disponibilização de um espaço e condições necessárias nas suas instalações que proporcionaram um bom ambiente de trabalho. Agradeço também à equipa que compõe esta empresa, pela forma como me integraram, e pela sua disponibilidade sempre que assim precisei e sobretudo pelo *feedback* e sugestões dadas. Fica aqui um agradecimento especial ao Rui Rodrigues por toda a ajuda e esclarecimentos que me prestou, ao longo do desenvolvimento deste projeto.

Finalmente gostaria de agradecer aos meus pais e irmão, pelo apoio constante, dedicação e pela confiança em mim depositada, nos bons e maus momentos, que me deram ânimo para a concretização deste momento na minha formação académica.

A todos, o meu mais sincero agradecimento.

Resumo

Os Arquivos Nacionais e Regionais têm como objectivo principal recolher e preservar a herança arquivística nacional que de alguma forma possuam interesse histórico, como por exemplo, documentos originados em cartórios, tribunais, ou em organismos religiosos. Estes documentos estão tipicamente relacionados com registos de nascimentos/óbitos, processos cíveis, entre outros. Para além de preservar esta informação, os Arquivos disponibilizam também a consulta pública destes documentos a quem assim o desejar.

De uma forma geral, estes documentos são organizados, de forma hierárquica num modelo que comporta três níveis, orgânico, funcional e documental, e constituídos fisicamente em depósitos, que correspondem a uma determinada área física do(s) edifício(s) que representam o Arquivo (por exemplo, salas).

A aplicação DigitArq da KEEP Solutions é um *software* de gestão de Arquivos, baseada em normas internacionais, que permite a produção e exportação de registos descritivos (informação recolhida nos documentos e outros objetos presentes nos fundos), disponibilizando-os em portais agregadores de conteúdos arquivísticos. Atualmente, esta aplicação permite guardar e obter informação relativa aos documentos e à sua hierarquia lógica nos fundos em que estão inseridos, mas não permite conhecer detalhes sobre os depósitos em que estão inseridos, nem sobre os depósitos em si mesmos.

É neste contexto que esta dissertação é realizada, acompanhando o processo de criação de um novo módulo experimental para o DigitArq que permita realizar a manutenção de depósitos, incluindo a gestão do seu espaço, a identificação da localização física dos seus documentos, e que possua também funcionalidades de otimização dos seus espaços em termos de armazenamento.

Palavras-chave: Arquivos históricos, Gestão de depósitos, DigitArq

Abstract

National and regional archives main purposes are to collect and maintain national archiving heritage that, in some way, may have some historical relevance, like notary's office documents, courts or religious organizations. These documents are typically related with birth/death records, civil proceedings, among others. Besides preserving this info, archives allows for records public consultation to whoever wishes so.

Generally, these documents are organized in a hierarchical way, in a model that supports three levels, organic, functional and documental, and then constituted physically in deposits that match to a given physical area of the building(s) that belongs to the archive (such as rooms).

The software DigitArq of KEEP Solutions is an archival management application, based on international standards, that allows for production and exporting of descriptive records (data collect from the documents and other objects in the fonds), making them available on archival aggregation portals.

Currently, this application allows for saving and retrieving documents data and their logical organization in the fonds they belong, but it doesn't provide any sort of information about the deposits where they are kept nor about the deposits themselves.

It's in this context that this dissertation will be developed, following the creation process of a new experimental software module for DigitArq, that will allow for deposit management, including its space management, documents physical location identification, and that also have storage space optimization features.

Keywords: Historical archives, Deposit management, DigitArq

Índice

Declaração	iii
Agradecimentos	iv
Resumo	v
Abstract.....	vi
Lista de abreviaturas e siglas	ix
Lista de figuras	x
Lista de tabelas	xi
1. Introdução	12
1.1. Enquadramento	12
1.2. Objetivos	14
1.3. Método	15
1.4. Organização da dissertação	15
2. Conceitos arquivísticos	17
2.1. Arquivos	17
2.2. Depósitos.....	18
2.3. Fundos	19
2.4. KEEP Solutions	22
2.5. DigitArq	22
3. Caso de estudo	26
3.1. Objeto de estudo	26
3.2. Tecnologias	27
3.3. Módulo de Gestão de Depósitos	28
3.3.1. Arquitetura.....	28
3.3.1.1. Casos de utilização	28
3.3.1.2. Diagrama de estados	30
3.3.1.3. Modelo de dados	31
3.3.2. Desenvolvimento	36
3.3.2.1. Comunicação com o servidor	36
3.3.2.2. Criação de serviços.....	37
3.3.2.3. Interface da aplicação-cliente.....	39
3.3.2.4. Suporte multilingue.....	41
3.3.2.5. Gestão de depósitos	42

3.3.2.6. Registos descritivos	43
3.3.2.7. Fragmentação de registos descritivos	47
3.3.2.8. Desfragmentação/otimização de depósitos	49
4. Performance da aplicação	54
4.1. Carregamento de TreeViews	54
4.2. Otimização de depósitos	55
4.3. Associação de registos descritivos	56
5. Conclusão	57
5.1. Síntese	57
5.2. Trabalho futuro	58
6. Referências bibliográficas	60

Lista de abreviaturas e siglas

API	Application Programming Interface
CSV	Comma-separated values
EAD	Encoded Archival Description
HTTP	Hypertext Transfer Protocol
ISAD(G)	General International Standard Archival Description
REST	REpresentational State Transfer
SGBD	Sistema de Gestão de Base de Dados
SQL	Structured Query Language
XML	eXtensible Markup Language
D	Documento simples
DC	Documento composto
UI	Unidade de instalação

Lista de figuras

Figura 1. Modelo de níveis de organização de um fundo	20
Figura 2. Interface do BackOffice do DigitArq.....	24
Figura 3. Relatório de registos descritivos que precisam edição	30
Figura 4. Diagrama de estados	31
Figura 5. Diagrama de classes	32
Figura 6. Diagrama do modelo lógico da base de dados.....	33
Figura 7. Arquitetura do DigitArq.....	37
Figura 8. Interface do formulário principal do módulo.....	39
Figura 9. Diálogo usado para a abertura/carregamento de depósitos	40
Figura 10. Edição de nível de depósito	43
Figura 11. Níveis de registos descritivos e seus possíveis encapsulamentos ..	45
Figura 12. Detalhes de registo descrito antes de associação a depósito	46
Figura 13. Exemplo de fragmentação de registos descritivos	47
Figura 14. Relatório de associação de registos descritivos num depósito	48
Figura 15. Reorganização de fundos num depósito	49
Figura 16. Preparação de otimização de um depósito	50
Figura 17. Relatório de otimização de um depósito	51

Lista de tabelas

Tabela 1. Documentação dos atributos das entidades da base de dados	35
Tabela 2. Tipos de registos descritivos e respectivos encapsulamentos	44

1. Introdução

Neste capítulo será feito um enquadramento do tema desta dissertação, seguido dos seus objetivos, o método utilizado para os atingir, e uma breve descrição da forma como este documento está organizado.

1.1. Enquadramento

Atualmente, quando se fala de espaço virtual pode afirmar-se que não existem limitações, uma vez que a sua expansão é realizada sem qualquer dificuldade com a adição de novos recursos físicos aos sistemas informáticos responsáveis pela sua gestão.

Quando se fala em arquivos, bibliotecas ou mesmo museus, as tecnologias presentemente existentes oferecem um espaço virtual praticamente infinito que permitem o armazenamento de todo o tipo de informação que seja necessária para catalogar um qualquer objecto (por exemplo, papéis, documentos, livros, etc.); já no caso de espaços físicos destes locais, esta situação não se verifica. Estes tipos de instituições têm geralmente um espaço limitado para o armazenamento dos seus objetos quando comparados com a sua grande quantidade, e uma vez que a expansão dos seus espaços nem sempre é viável, maioritariamente devido aos custos envolvidos, mas por vezes também devido a impossibilidades geográficas que limitam as suas expansões, a alternativa recai, geralmente, sobre uma melhor gestão do espaço existente, por forma a aproveitar ao máximo a capacidade disponível nas infraestruturas.

Os arquivos físicos não podem ser considerados como apenas locais onde objetos, como documentos, são guardados e preservados ao longo do tempo. Os conteúdos destes objetos favorecem estes espaços, de forma cultural, social e histórica, sendo assim de extrema importância garantir que os objetos presentes nos arquivos se mantenham em bom estado, mas que, com semelhante consideração, estejam também bem organizados, permitindo assim o fácil acesso a quem os pretenda consultar. Assim, é necessário documentar corretamente os objetos, bem como as suas localizações no arquivo para otimizar o processo de catalogação e pesquisa.

Esta é a Era da Informação e portanto a sua partilha é de extrema importância, por forma a manter as fontes de conhecimento vivas e ativas. Como a disseminação da informação digital é hoje um processo bastante vulgar, é necessário assegurar que a comunicação de dados entre este tipo de entidades seja realizada uniformemente, permitindo que a partilha seja realizada de forma fluída. Este processo é assegurado a partir de normas, que englobam, por exemplo, a caracterização e catalogação dos objetos, bem como a forma como estes devem ser partilhados na rede.

O *software* DigitArq, da KEEP Solutions¹, tem como objectivo a gestão integrada de Arquivos Definitivos (também conhecidos por Arquivos Históricos ou Mortos), e está preparado para satisfazer as necessidades dos profissionais de arquivos, contendo funcionalidades como a gestão de projetos de digitalização, publicação Web, navegação, pesquisa de objetos e descrição arquivista.

A descrição arquivística consiste na criação de representações precisas e adequadas, de acordo com modelos pré-determinados, de unidades de descrições e das suas componentes, através da apreensão, análise, organização e registo de informação que permita identificar, gerir e localizar materiais arquivísticos e o contexto e sistemas de documentos que os produziram. Os procedimentos relacionados com a descrição podem ter início no momento da produção dos documentos (ou antes até) e continuar durante todo o seu ciclo de vida [11] [10].

O DigitArq assenta em várias normas internacionais para a produção e exportação de registos descritivos, permitindo desta forma a integração com portais agregadores de conteúdos nacionais (Portal Português de Arquivos²) e internacionais (Europeana³, DRIVER⁴, etc.) e permite guardar e obter informações relativas aos objetos em arquivo, bem como a sua organização lógica.

No entanto, no DigitArq não existem funcionalidades que permitam conhecer as estruturas físicas onde os objetos (representados na aplicação através de registos descritivos) se encontram, normalmente designados por

¹ <http://www.keep.pt/>

² <http://portal.arquivos.pt>

³ <http://www.europeana.eu/>

⁴ <http://www.driver-repository.eu/>

depósitos, o que asseguraria não apenas a gestão dos documentos mas também a gestão dos espaços que os contêm.

Assim, este projeto visa o desenvolvimento de um novo módulo experimental para o DigitArq, que permita realizar a gestão do espaço físico de arquivos. Essa gestão envolve a monitorização da capacidade do arquivo e a disposição da mesma, bem como a gestão do espaço ocupado pelos documentos do arquivo. Ou seja, dotar a aplicação com ferramentas que possibilitem recriar no universo digital a organização física do arquivo e a relacioná-la com a sua organização lógica.

1.2. Objetivos

A ausência de um módulo capaz de realizar a gestão de espaço físico em depósitos, por parte da aplicação DigitArq, é o mote principal para a realização desta dissertação. Desta forma, o objectivo principal desta dissertação passa pela criação de um módulo e de toda a tecnologia envolvente (base de dados, serviço *Web* correspondente e aplicação para o utilizador final) que complemente as funcionalidades já existentes no DigitArq, permitindo a gestão de espaços físicos, de forma semelhante à atualmente utilizada pelo DigitArq para registos descritivos em fundos.

Com a disponibilização deste módulo de gestão, surge uma série de funcionalidades que proporcionarão não apenas o cruzamento de informação entre as duas hierarquias descritivas (arquivista e do espaço físico), como por exemplo, indicar a posição física de um dado objecto num depósito, mas possivelmente permitir também operações de “desfragmentação” e sugestão de realocação de objetos no espaço físico. Este é um outro objectivo desta dissertação, que passa pela abordagem a métodos de optimização do espaço utilizado nos depósitos, sugerindo alterações na forma como os objetos são armazenados, e a consequente criação de algoritmos que sugiram a “desfragmentação” dos mesmos.

1.3. Método

Por forma a alcançar os objectivos anteriormente referidos, foi seguida uma abordagem que permitiu conhecer melhor o mundo dos Arquivos, da sua utilidade e conseqüente importância na sociedade.

Assim, inicialmente foi realizado um estudo exaustivo sobre os Arquivos, sobre a forma como, nas últimas décadas, tem sido feito um esforço mundial para o melhoramento da catalogação e organização dos seus conteúdos, permitindo uma disseminação mais fácil e rápida de conhecimento, com a criação de comités e normas internacionais para o efeito.

Seguidamente, a aplicação-alvo DigitArq, desenvolvida pela KEEP Solutions, foi estudada, relativamente ao seu funcionamento e as ferramentas que disponibiliza, analisando ainda o impacto que esta representa na gestão de fundos nos Arquivos distritais e nacionais onde é atualmente utilizada. Durante este período foi também realizada uma visita ao Arquivo Distrital de Braga⁵, onde foi possível verificar, de facto, como são armazenados e preservados os conteúdos em depósitos.

Finalmente, foi criado o módulo de gestão física dos arquivos, que engloba a elaboração de uma nova base de dados, criação de funções para a manipulação e otimização da informação sobre os depósitos (serviço *Web*) e a aplicação para o utilizador final.

1.4. Organização da dissertação

De seguida, é apresentada a estrutura da dissertação, acompanhada de breves descrições dos capítulos seguintes.

No Capítulo 2 são apresentados, de forma mais extensa, alguns dos conceitos mais relevantes no universo dos arquivos e da aplicação a ser desenvolvida. Assim, são abordados os conceitos de arquivos, depósitos e fundos, e depois é descrita a aplicação para o qual será desenvolvido o novo módulo de gestão de arquivos, bem como a empresa responsável pela mesma.

O Capítulo 3 corresponde à análise ao caso de estudo, que consiste no desenvolvimento de um módulo de gestão de depósitos para a aplicação

⁵ <http://www.adb.uminho.pt/>, Largo do Paço — 4704-553 Braga

DigitArq da KEEP Solutions. São descritas as tecnologias envolvidas e, de forma mais compreensiva, será descrita a arquitetura pretendida para a mesma (funcionalidades, casos de utilização, diagramas de estado e desenvolvimento de um modelo de dados), e todo o processo de desenvolvimento, incluindo a justificação de algumas das decisões mais relevantes durante o mesmo. O processo de otimização de depósitos (parte integrante dos objectivos desta dissertação) será também descrito no final deste capítulo.

O Capítulo 4 aborda alguns dos testes realizados sobre a performance da aplicação e resultados obtidos, relativamente às operações disponibilizadas pelo módulo. Os testes incidem sobretudo em operações de associação de objetos (através dos seus registos descritivos) aos depósitos, e às operações de otimização dos mesmos, uma vez que são as operações mais complexas na aplicação, quer em termos de cálculos, quer em quantidade de dados manipulados, daí a sua importância.

Finalmente, o Capítulo 5 trata a conclusão da dissertação, apresentando um resumo da mesma, as deduções retiradas de todo o trabalho efetuado e o que poderá ser realizado no futuro com o objetivo de melhorar este módulo e assim torná-lo numa ferramenta essencial para os arquivistas que já utilizam os restantes módulos do DigitArq.

2. Conceitos arquivísticos

Neste capítulo serão apresentados com maior detalhe alguns dos conceitos mais relevantes no interesse dos arquivos e da aplicação que será desenvolvida. Serão descritos os conceitos de arquivo, depósito e fundos; a aplicação DigitArq e a empresa responsável pelo seu desenvolvimento, KEEP Solutions, serão também apresentadas com mais detalhe.

2.1. Arquivos

Em Portugal, os Arquivos Nacionais (Arquivo Nacional da Torre do Tombo e Centro Português de Fotografia) e Regionais (16 no território continental), que se encontram sob a alçada da Direcção-Geral de Arquivos [1], têm como missão principal a recolha, preservação e valorização do património arquivístico nacional com que possuam determinado relevo histórico. Todos os arquivos são públicos, permitindo assim a consulta de qualquer documento neles guardados (exceptuando alguns casos particulares, como por exemplo, devido ao estado de conservação dos mesmos) [5], a qualquer pessoa, através das salas de leitura e serviços de referências que possuem.

Ao abrigo da Lei Geral dos Arquivos Distritais, estes arquivos incluem documentos produzidos pelas Conservatórias do Registo Civil (com uma antiguidade superior a cem anos), Cartórios Notariais (com mais de 30 anos) e pelos Tribunais do distrito em que se inserem (com mais de 35 anos após a conclusão dos processos).

Além destes tipos, é comum encontrar também documentos de outros organismos públicos, como Governos Cívicos, Alfândegas, Finanças, bem como conjuntos documentais pertencentes a organismos monásticos e religiosos (como antigos conventos, dioceses e ordens religiosas extintas, sendo que algumas remontam à Idade Média, entre os séc. V e XV), confrarias, irmandades, misericórdias, etc.. A título de exemplo, os tipos de documentos produzidos por paróquias, e que podem ser encontrados em arquivos, passam por registos de batismos, casamentos e óbitos. Deste exemplo, é possível observar a importância da conservação destes documentos, para a prática de genealogia, ou para certificação de identidades e relações familiares.

Em regime de doação, os arquivos distritais podem executar também a preservação documental de famílias, pessoas singulares ou colectivas que desejem que estes sejam conservados definitivamente, como fonte de compreensão da memória social e cultural.

Para além da preservação do património arquivístico, cabem aos arquivos também as tarefas de tratamento técnico documental, bem como a organização, classificação e catalogação dos documentos dentro do conjunto de edifícios que formam os arquivos distritais e nacionais. Esta tarefa torna-se algo complicada, não apenas pelo volume de documentos que poderá estar presentes em cada arquivo, mas também, conseqüentemente, pela limitação da capacidade das infraestruturas dos mesmos. Daí que seja necessário otimizar a gestão do espaço por forma a recolher o maior volume de documentos possível, mas mantendo-os devidamente organizados.

Os arquivos são organizados em depósitos, que normalmente representam as várias divisões da infraestrutura onde, de forma definitiva, serão guardados os documentos. Estes podem ter vários andares (dependendo da estrutura da divisão), e os documentos são armazenados em armários, prateleiras, estantes, etc. No entanto, o mais correto será que os documentos de um dado assunto, ou produtor, sejam armazenados de forma contígua, e aí surge novos tipos de organização nos depósitos: os fundos e as coleções.

2.2. Depósitos

Os arquivos possuem dimensões variáveis, podendo ir de uma simples prateleira num armário, até ao agrupamento de vários edifícios. Para permitir uma melhor organização dos seus documentos, estes são agrupados de acordo com a descrição arquivística associada aos fundos que serão armazenados, a partir da sua descrição funcional, que por exemplo, corresponde a séries ou secções, formando assim os depósitos.

Devido à elevada quantidade de possibilidades e ao espaço que estes podem ocupar, torna-se complicado designar cada um dos níveis (ao contrário dos fundos, em que estes estão definidos à partida; é apenas uma questão de definir o número de subníveis que existirão) da hierarquia para um dado depósito (separação por salas, pisos, armários, gavetas, etc.). No entanto, os

arquivistas (profissionais responsáveis pela reunião, organização e preservação da informação presente nos arquivos, em qualquer suporte, incluindo fotografias, documentos, sons, vídeos, etc.) desse arquivo conhecem melhor a sua estrutura e portanto serão mais capazes de identificar e classificar corretamente cada nível.

Desta forma, a construção de depósitos neste projeto terá uma hierarquia em tudo semelhante à de um fundo, ou seja, existe um nível superior (depósito), que será subdividido quantas vezes for necessário e por sua vez os subníveis também o serão. Devido à particularidade das designações em cada nível do arquivo físico, será dada a liberdade aos seus utilizadores de estabelecerem as designações e hierarquias que bem entenderem. A cada nível serão depois associados objetos (representados pelos seus registos descritivos no DigitArq) pertencentes ao arquivo: fundos, séries documentais, documentos simples que correspondem a documentos físicos ou a agrupamentos destes (papéis, livros, capas, etc.).

2.3. Fundos

Os Arquivos Definitivos representam um conjunto de objetos guardados com carácter definitivo, com a possibilidade de estes serem utilizados para fins diferentes daqueles para os quais foram inicialmente criados. Assim, estes objetos passam a ser considerados como fontes de informação e pesquisa para a administração dos arquivos ou mesmo para terceiros.

Quando os objetos são recolhidos por unidades administrativas, estes passam a estar agrupados com outros, constituindo-se assim em fundos arquivísticos.

Um fundo arquivístico é um conjunto de documentos de arquivo, independentes da sua forma ou suporte, organicamente produzido e/ou acumulado por uma ou mais pessoas, no decurso das suas atividades (mantendo assim uma única fonte de informação), guardando entre si relações orgânicas, que podem ser preservados como provas, testemunhos legais ou culturais, não devendo portanto, ser misturados com outros objetos de outros conjuntos acumulados por outras instituições mesmo sendo semelhantes [12].

Na prática arquivística, os fundos representam o nível de organização mais alto. Estes podem ser divididos em hierarquias mais pequenas, para garantir uma melhor organização dos objetos, como por exemplo, em subfundos. Por sua vez, os subfundos podem ser subdivididos em séries, que também poderão ser subdivididas, até atingirem o nível mais pequeno de descrição, as peças ou documentos simples. Em teoria, cada nível poderá assumir um valor infinito de subníveis, e estes também os poderão ter, e assim sucessivamente, mas na prática, é raro que tenham mais que um.

O diagrama seguinte (Figura 1) poderá ilustrar melhor a organização hierárquica típica de um fundo.

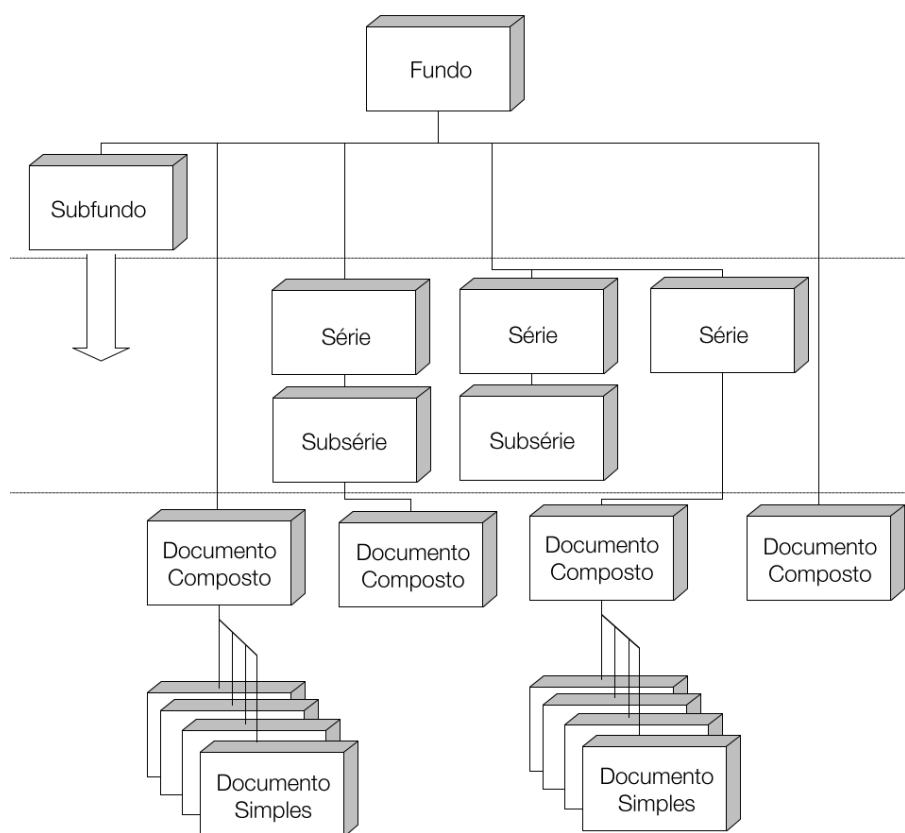


Figura 1. Modelo de níveis de organização de um fundo

Como é possível observar pelo diagrama [4], um fundo poderá ter várias configurações possíveis, dependendo do nível de organização e detalhe que lhe foi dado por quem o descreveu.

A estrutura de um fundo pode ser classificada em três tipos de descrições, que podem ser observados pelas linhas horizontais presentes na Figura 1: orgânica (inclui o fundo e subfundo), funcional (inclui séries, subséries) e documental (que inclui os documentos compostos e documentos simples).

Para uma melhor compreensão dos vários elementos de organização de um fundo, estes serão brevemente descritos de seguida [12].

Subfundo: representa uma subdivisão no fundo, correspondendo a uma subdivisão na descrição orgânica, que contém um conjunto de documentos relacionados que podem corresponder, por exemplo, a subdivisões administrativas, geográficas, cronológicas, etc. São justificados quando a estrutura hierárquica se começa a tornar demasiado complexa. A título de exemplo da sua utilização, considere-se uma empresa que possui múltiplos departamentos. Cada um desses departamentos poderia corresponder a um subfundo, sendo que o fundo representaria a totalidade da empresa.

Série: representa um conjunto de documentos (minutas, atas, arquivos de correspondência, etc.), na hierarquia funcional, organizados com um dado sistema de arquivamento e conservados como uma unidade, por pertencerem a um mesmo processo de acumulação, por terem uma determinada tipologia, ou uma outra qualquer relação resultante do processo de produção/utilização.

Sub-série: à semelhança do subfundo, representam uma subdivisão na série, para melhor definir uma hierarquia, quando esta se tornar muito complexa, e contém o mesmo tipo de documentos presente nas séries. Por exemplo, considerando uma série que possui documentos de correspondência; esta poderia ser dividida em duas sub-séries, para as correspondências recebidas e enviadas.

Unidade de instalação: consiste numa unidade básica de armazenamento e cotação de unidades arquivísticas. Os objetos que representam geralmente unidades de instalação são caixas, pastas, livros, maços, etc. [3].

Processo ou **Documento composto**: é uma unidade organizada de documentos agrupados, para uso corrente do seu produto ou decurso da organização arquivística, por tratarem um mesmo assunto, atividade ou transação.

Peça ou **Documento simples**: representa a unidade arquivística mais pequena, e é normalmente indivisível (por exemplo, cartas).

2.4. KEEP Solutions

A KEEP Solutions é uma empresa de base tecnológica que tem como missão a prestação de serviços na área de gestão de preservação de informação, dirigidos sobretudo a museus, bibliotecas e arquivos, especializando-se na prestação de serviços de consultoria em preservação digital, migração de dados, digitalização e manutenção, alojamento e suporte de repositórios digitais. Os seus principais clientes pertencem sobretudo ao sector público e educacional, como ministérios, instituições militares e académicas, museus, arquivos e fundações [14].

A empresa iniciou a sua atividade em 2008 com o estatuto de *spin-off* académica por se tratar de uma iniciativa com fortes ligações aos centros de investigação e departamentos da Universidade do Minho, permanecendo-se ativa na produção de conhecimento científico, através de participações e publicações dos seus colaboradores em eventos do foro científico, e em parcerias de projetos de investigação com várias instituições nacionais e internacionais.

Para além do DigitArq (aplicação-alvo desta dissertação), desenvolveu nos últimos anos, um conjunto de projetos também eles dedicados à preservação digital⁶, implementação de repositórios⁷, e ao desenvolvimento de portais agregadores de conteúdos⁸.

2.5. DigitArq

A plataforma de *software* DigitArq, da KEEP Solutions, tem como propósito a gestão global de documentação em Arquivos Definitivos (também conhecidos como Arquivos Históricos ou Mortos). É constituído por vários módulos, que incluem a descrição arquivística, gestão de projetos de digitalização, publicação na Web, registos de intervenções de conservação e

⁶ RODA: http://www.keep.pt/?page_id=293

⁷ DSpace: http://www.keep.pt/?page_id=283

⁸ Retrievio: http://www.keep.pt/?page_id=287

restauro, etc., que em conjunto procuram responder às necessidades de profissionais de arquivos [13].

O DigitArq encontra-se atualmente implementado em vários arquivos em Portugal, destacando-se o facto de estar na totalidade dos arquivos dependentes da Direcção-Geral de Arquivos (16 Arquivos Distritais, Arquivo Nacional da Torre do Tombo e o Centro Português de Fotografia), no Museu da Presidência da República, Ministérios, em vários Arquivos Municipais, e mais recentemente no Arquivo Histórico da Marinha.

Através de um dos seus vários módulos, o DigitArq é completamente compatível com o Portal Português de Arquivos, que se trata de um portal agregador de conteúdos das instituições ligadas à Rede Portuguesa de Arquivos⁹, permitindo o acesso e pesquisa aos mesmos. Isto significa que toda a informação armazenada nos arquivos através do DigitArq poderá estar facilmente acessível a partir de qualquer ponto do globo.

Em 2004, a Agência para a Sociedade do Conhecimento premiou este projeto pela “inovação e contributo para o desenvolvimento da Sociedade da Informação” [6].

De entre os vários módulos da aplicação, Administração (configuração da aplicação, gestão de utilizadores, etc.), Interoperabilidade OAI-PMH¹⁰ (disponibilização de registos descritivos através de protocolo com a mesma designação, em portais agregadores como o Portal Português de Arquivos, DRIVER, Europeia, etc.), FrontOffice (ponte entre arquivo, e seus fundos, e utilizador externo através da disponibilização do serviço na Web, permitindo operações de localização e pesquisa de registos descritivos), o módulo de BackOffice é o que mais importância possui no âmbito desta dissertação e portanto será descrito de uma forma mais abrangente que os restantes.

O módulo de BackOffice é o responsável pela produção e gestão de registos de descrição, de acordo com a norma ISAD(G), que se trata de uma norma aprovada pelo Conselho Internacional de Arquivos para a produção de registos descritivos [11].

⁹ <http://www.arquivos.pt/>

¹⁰ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Entre as suas funcionalidades mais importantes, salientam-se a gestão automática de códigos de referência, mecanismos de pesquisa avançada, controlo de qualidade das descrições, produção de relatórios nos mais populares formatos, importação/exportação de instrumentos de descrição em XML, EAD (sendo este um formato específico utilizado para descrição arquivística [9]), CSV, etc., e suporte para ações de conservação e restauro de documentos. A sua performance mantém-se mesmo lidando com fundos de grandes dimensões, por vezes na ordem de milhares de registos.

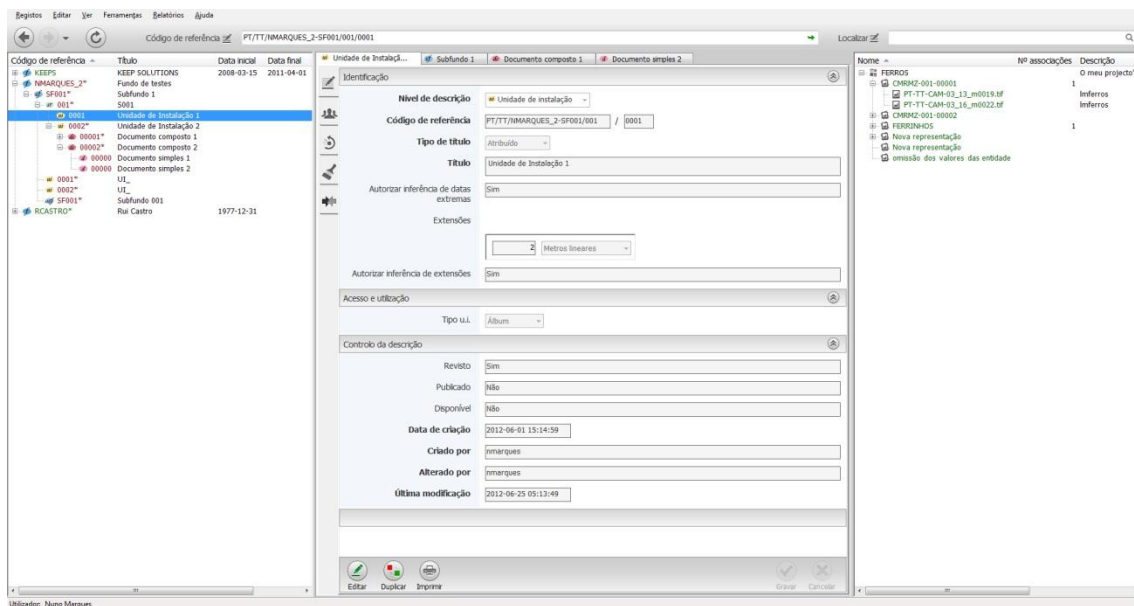


Figura 2. Interface do BackOffice do DigitArq

O BackOffice está organizado de forma bastante intuitiva, apresentando uma árvore para os níveis descritivos (onde se localiza toda a estrutura dos fundos abertos), outra para a visualização de representações (visualização de imagens e outros tipos de multimédia que representam os registos descritivos) e um painel central, onde podem ser visualizados/editados detalhes sobre registos descritivos e representações, que apresenta um comportamento similar aos *browsers* modernos, suportando separadores/*tabs* e navegação por histórico. Suporta ainda *drag-n-drop* para mais facilmente mover registos nas suas árvores de descrição e adição de multimédia aos registos.

O BackOffice possui também algumas ferramentas essenciais para a revisão de objetos e publicação *online* dos mesmos, mecanismos de inferência (cada objecto pode ter associadas extensões, que consistem em quantificações

ou unidades de medidas, e que o definem de forma mais correta; os mecanismos de inferência calculam para todo o fundo os valores associados a cada extensão, obtendo assim uma melhor percepção das suas dimensões físicas), e uma vasta gama de relatórios que darão todo o apoio necessário aos arquivistas, como catalogações, inventários, guias, revisões, etc..

3. Caso de estudo

Neste capítulo será descrito todo o processo de desenvolvimento do módulo de gestão de depósitos que mais tarde poderá ser implementado no *software* DigitArq.

Na primeira parte, será feita uma breve descrição da aplicação-alvo, DigitArq BackOffice, as suas funcionalidades, juntamente as tecnologias utilizadas neste projeto. Na segunda parte, e de uma forma mais compreensiva, será explicado o processo do desenvolvimento do módulo, descrevendo a arquitetura e a abordagem utilizadas e problemas encontrados.

3.1. Objeto de estudo

O objecto de estudo desta dissertação será o desenvolvimento de um novo módulo de *software*, capaz de realizar a gestão de depósitos (espaço físico) em Arquivos, na aplicação DigitArq, desenvolvida pela KEEP Solutions.

O DigitArq é um sistema de gestão de documentação em Arquivos, garantindo operações sobre os vários pontos de vista e necessidades dos arquivistas, incluindo registo de intervenções de conservação e restauro, gestão de digitalizações, publicações na *Web*, etc.

Um dos módulos pertencentes a este *software*, DigitArq BackOffice, lida diretamente com a descrição arquivista, e através desse módulo é possível garantir a organização lógica dos fundos guardados em depósitos, podendo identificar e gerir os locais onde determinados objetos estão armazenados.

Assim, o DigitArq BackOffice será considerado o caso de estudo, pois só apenas através desta aplicação é possível conciliar a gestão arquivística e a gestão física de depósitos, ao permitir definir representações de depósitos e associar-lhes registos descritivos.

3.2. Tecnologias

Uma vez que esta dissertação aborda a criação de um novo módulo a ser adicionado a *software* já existente, é de alguma forma conveniente manter as tecnologias previamente utilizadas nos outros módulos, para haver coerência com o desenvolvimento da restante aplicação, permitir a fácil adaptação de código previamente utilizado e/ou objetos de interface e familiarização com os mesmos e no futuro tornar mais fácil a implementação deste módulo na aplicação principal.

Uma vez que o DigitArq foi desenvolvido com ferramentas Microsoft, as mesmas serão mantidas para este projeto. Assim, o sistema de gestão de bases de dados a utilizar será o Microsoft SQL Server 2008, e o ambiente de desenvolvimento será o Microsoft Visual Studio 2008 (utilizado para o desenvolvimento do serviço *Web* necessário para a gestão de depósitos e da aplicação-cliente), especificamente com a linguagem orientada a objetos, Visual Basic .NET.

A utilização de ferramentas de desenvolvimento de um mesmo fabricante possibilita uma melhor interligação das suas funcionalidades e performance, uma vez que geralmente se complementam nos projetos em que são ambas utilizadas.

Apesar de existirem versões mais recentes de ambas as ferramentas (nomeadamente Visual Studio 2010 e 2012, e SQL Server 2012) estas não serão consideradas neste projeto, porque se tratam de versões muito recentes e portanto com menor maturidade e menos exploração pelos desenvolvedores de *software* (estando também assim mais expostas a possíveis erros e falhas de segurança comparativamente à versão 2008), e possíveis inconsistências que possam surgir com o código desenvolvido no módulo BackOffice e nos serviços (descritos nas secções 3.3.2.1 e 3.3.2.2) que será aproveitado para o novo módulo. Uma situação similar verifica-se com o SGBD SQL Server, uma vez que a informação comunicada aos serviços é obtida a partir da execução de *Stored Procedures* (sub-rotinas que utilizam comandos SQL para o tratamento de dados) e portanto podem surgir pequenas alterações que possam trazer resultados indesejados.

3.3. Módulo de Gestão de Depósitos

Nesta secção será analisado o processo de construção do módulo de Gestão de Depósitos a ser integrado no DigitArq. A análise será dividida em duas partes, correspondentes à arquitetura (onde será realizada a descrição dos casos de utilização, elaboração de diagrama de estados e o modelo de dados) e ao desenvolvimento (onde será caracterizado o processo de criação de serviços, a aplicação-cliente e respectiva interface, descrição dos mecanismos de gestão dos depósitos e processos de fragmentação, desfragmentação e otimização dos mesmos).

3.3.1. Arquitetura

Nesta subsecção será descrita, de uma forma mais compreensiva, a arquitetura do módulo de gestão de depósitos. O comportamento esperado da aplicação será explicado através da demonstração das ações suportadas, da construção do diagrama de estados, especificação do modelo de dados e finalmente da criação da aplicação-cliente, respectiva interface e seus formulários adicionais.

3.3.1.1. Casos de utilização

Tratando-se de um novo módulo cujos objectivos principais estão definidos desde o ponto de partida do projeto, o processo de identificação dos mesmos torna-se fácil. No entanto, existem algumas funcionalidades interessantes no módulo BackOffice, que seria de alguma forma proveitoso serem adaptadas às necessidades deste módulo, como por exemplo, mecanismos de pesquisa, que neste caso seriam para níveis físicos dos depósitos, e para registos descritivos associados a depósitos.

Existem outras funcionalidades que apesar de interessantes, ficam um pouco fora do âmbito deste projeto e dos seus objectivos, e poderão ser mais tarde implementados, como por exemplo, um sistema de navegação por histórico, semelhante aos utilizados pelos *Web browsers*.

De uma forma geral, estas serão as funcionalidades a serem implementadas, e respectivas descrições, para esta aplicação:

- **Autenticação**

Apenas a funcionalidade de *Login* será implementada, pois a criação de utilizadores é feita a partir de um outro módulo do DigitArq.
- **Sair**

Termina a execução da aplicação, mas sem antes guardar todo o trabalho (se o utilizador assim o desejar).
- **Criação/edição de depósitos**

O utilizador (administrador apenas) pode criar/editar depósitos e restante estrutura física dos mesmos, para assim se poderem associar os registos descritivos; o utilizador comum apenas pode adicionar contentores aos depósitos.
- **Associação de registos descritivos aos depósitos**

Aos níveis do depósito devem poder ser associados registos descritivos (que correspondem a representações de objetos) para assim ser efectuado o mapeamento do mesmo. As associações podem ser feitas de forma individual (registo a registo), grupos de registos, ou todo o fundo de uma só vez.
- **Gestão automática de espaços consumidos**

A cada alteração feita sobre os depósitos (em termos de associações de registos descritivos), a atualização sobre o espaço físico é feita de forma automática e com repercussões sobre todo o depósito, permitindo ao utilizador saber a cada momento, qual o espaço real que dispõe em cada nível.
- **Espaço disponível em níveis de depósito**

Quando o utilizador precisa associar objetos a um depósito (através dos seus registos descritivos nos fundos), precisa saber quais os níveis do depósito que possuem espaço livre suficiente para fazer a inserção. Esta funcionalidade permite que ao indicar qual o espaço necessário, sejam apresentados todos os níveis do depósito capazes de suportar os metros lineares desejados.
- **Detecção automática de registos repetidos**

Salvo raras exceções (esclarecidas mais à frente), não poderão existir registos repetidos num depósito, uma vez que os registos descritivos correspondem a objetos únicos.

Por forma a evitar esta situação, que não se pode verificar na realidade, a aplicação faz uma verificação no momento de associação do conjunto de registos ao depósito, e impede-a caso se encontrem registos nesse conjunto que estejam previamente no depósito.

- Pesquisa de depósitos

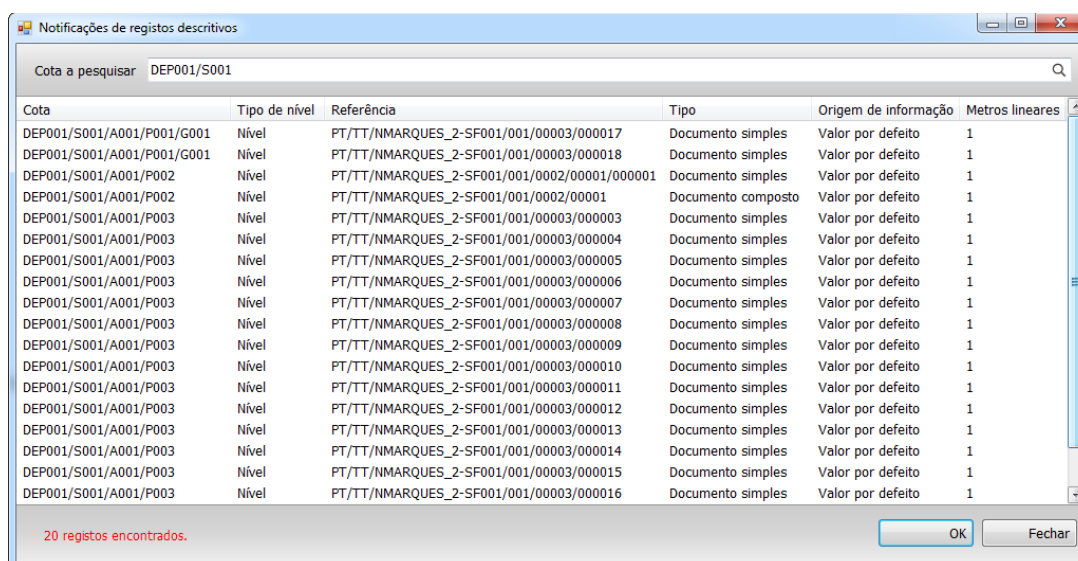
Uma das funcionalidades de pesquisa; permite indicar a cota (sistema de referência) de um depósito ou de um dos seus subníveis e assim verificar todos os detalhes e registos descritivos associados a esse nível.

- Pesquisa por registos descritivos em depósitos

Permite procurar registos descritivos e dessa forma identificar qual o depósito e nível em que o objeto que este representa está associado (apenas no caso de existir a associação dos registos).

- Filtragem de registos descritivos que precisam edição

O cálculo do espaço consumido é feito com base numa das várias extensões existentes na prática arquivística, os metros lineares. Quando estes não forem fornecidos, a aplicação atribui-lhes um valor por omissão, e portanto poderá não refletir a realidade sobre as dimensões dos objetos que representam. Esta filtragem permite ao utilizador identificar e editar os registos que estejam nesta situação.



Cota	Tipo de nível	Referência	Tipo	Origem de informação	Metros lineares
DEP001/S001/A001/P001/G001	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000017	Documento simples	Valor por defeito	1
DEP001/S001/A001/P001/G001	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000018	Documento simples	Valor por defeito	1
DEP001/S001/A001/P002	Nível	PT/TT/NMARQUES_2-SF001/001/0002/00001/000001	Documento simples	Valor por defeito	1
DEP001/S001/A001/P002	Nível	PT/TT/NMARQUES_2-SF001/001/0002/00001	Documento composto	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000003	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000004	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000005	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000006	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000007	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000008	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000009	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000010	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000011	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000012	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000013	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000014	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000015	Documento simples	Valor por defeito	1
DEP001/S001/A001/P003	Nível	PT/TT/NMARQUES_2-SF001/001/00003/000016	Documento simples	Valor por defeito	1

Figura 3. Relatório de registos descritivos que precisam edição

3.3.1.2. Diagrama de estados

Observando o diagrama de estados seguinte (Figura 4), é possível compreender o funcionamento global da aplicação e a forma como os vários diálogos disponíveis na aplicação são alcançados e utilizados.

A maior parte das funcionalidades deste módulo acontecem no formulário principal, onde são criados/editados depósitos, feitas associações de registos descritivos, etc.. Uma vez que todas essas ações decorrem num painel central

de edição localizado nesse formulário e todas elas são confirmadas/canceladas da mesma forma (botões 'Guardar', 'Cancelar' e 'Editar' disponíveis neste painel), torna-se redundante aprofundar mais o diagrama de estados da aplicação.

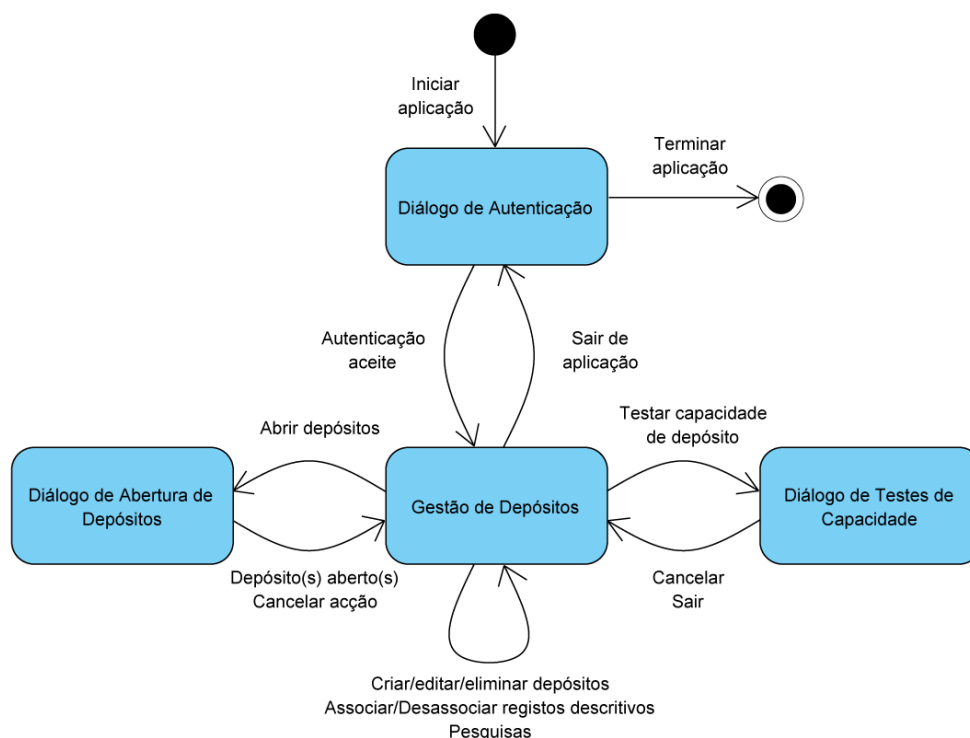


Figura 4. Diagrama de estados

3.3.1.3. Modelo de dados

O processo de criação/edição de depósitos e associação de registos descritivos necessita obrigatoriamente de comunicar com o servidor, para realizar ações de leitura e escrita. Para tal foram criadas classes de dados (ver Figura 5), que servem de ponto de comunicação intermédio entre o servidor aplicacional e o servidor de base de dados. Estas classes representam adequadamente as estruturas definidas na base de dados que servem para armazenar estruturas físicas, registos descritivos associados, tipos de contentores e identificação de níveis físicos.

A descrição de um depósito exige a participação de todas estas classes, uma vez que se complementam entre si, descrevendo assim o depósito e respectiva estrutura, bem como os registos descritivos a si associados.

Deve apenas salientar-se que um dado objecto do tipo *PhysicalStructure* tem sempre associado a si um *ObjectType* ou um *Level*, mas nunca ambos ao mesmo tempo, uma vez que o primeiro destina-se à classificação de contentores (seus tipos) e o segundo destina-se à classificação dos níveis físicos.

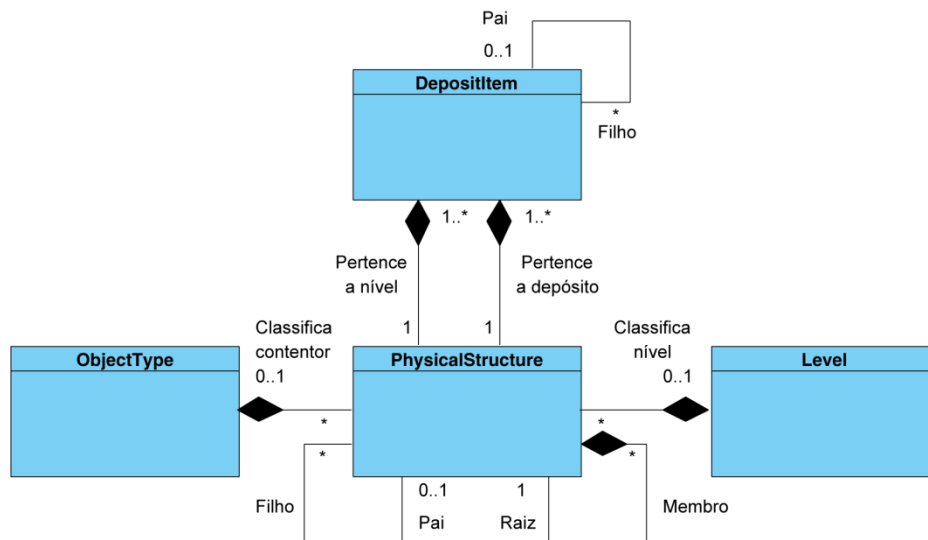


Figura 5. Diagrama de classes

Estas classes representam adequadamente as entidades definidas na base de dados (ver Figura 6) que servem para armazenar estruturas físicas (depósitos e seus níveis), registos descritivos associados, tipos de contentores (objetos existentes em arquivos com o contexto de armazenar num único local, vários registos descritivos relacionados; por exemplo, caixas, pilhas, etc.; um contentor apenas pode armazenar itens de um único fundo) e identificação de níveis físicos (classificações subjacentes à descrição da estrutura física de um depósito, ou seja a sua descrição; por exemplo, sala, armário, prateleira, gaveta).

PhysicalStructure e *DepositItem*), torna-se complicado criar um modelo de dados tradicional, em que cada nível corresponde a uma entidade no modelo, até porque é possível evitar níveis intermédios na definição de registos (como pode ser observado pela Figura 1).

A solução passa assim por utilizar uma única entidade para toda a definição da hierarquia, criando desta forma um relacionamento recursivo. Num relacionamento recursivo, a mesma entidade participa mais que uma vez no relacionamento com diferentes funções [2].

Num exemplo clássico que define esta solução, considere-se uma entidade que represente Empregados de uma empresa, em que são definidos os supervisores e os supervisionados através de um atributo adicional que corresponde ao identificador do supervisor imediato. Assim é possível definir hierarquias com uma única entidade, à semelhança do que é apresentado nas entidades *PhysicalStructure* e *DepositItem*.

Entidade	Atributo	Descrição	Tipo	Nulo?
Object Type	ID	Identificador único	Inteiro	Não
	Name	Designação do tipo de objecto	String / 35	Sim
Level	ID	Identificador único	Inteiro	Não
	Name	Designação do nível	String / 35	Sim
Physical Structure	ID	Identificador único	Inteiro	Não
	RootID	Identificador raiz do depósito	Inteiro	Sim
	ParentID	Identificador de registo pai	Inteiro	Sim
	IsStructure	Indicador de nível ou contentor	Bit	Não
	Kind	Tipo de nível	String / 2	Não
	OrderItem	Ordem do nível no depósito	Inteiro	Sim
	Reference	Cota do nível	String / 35	Sim
	CompleteReference	Cota completa do nível	String/150	Sim
	Title	Título do nível	String / 45	Sim
	LevelID	Identificador de tipo de nível	Inteiro	Sim
	ObjectTypeID	Identificador de tipo de objecto	Inteiro	Sim
	Capacity	Capacidade total do nível	Float	Sim
	UsedCapacity	Capacidade usada do nível	Float	Sim
	Creator	Criador do registo do nível	String / 50	Não
	Created	Data de criação de registo	Data	Sim
	Modifier	Modificador de registo do nível	String / 50	Sim
	Modified	Data de modificação de registo	Data	Sim
Notes	Notas adicionais	Texto	Sim	
Temporary	Indicador de registo temporário	Bit	Não	
Deposit Item	ID	Identificador único	Inteiro	Não
	ParentID	Identificador de registo pai	Inteiro	Sim
	DepositID	Identificador de depósito	Inteiro	Não
	PhysicalStructureID	Identificador de nível físico	Inteiro	Não
	ComponentID	Identificador registo descritivo	Inteiro	Não
	IsFragmented	Indicador de fragmentação	Bit	Não
	LinearMeters	Valor extensão metros lineares	Float	Sim
	InfoOrigin	Origem informação da extensão	Inteiro	Sim
	OrderItem	Ordem do registo na árvore	Inteiro	Sim
	Title	Título do registo descritivo	String / 45	Sim
	DescriptionLevel	Identificador nível de descrição	String / 45	Sim
	DescriptionLevelPT	Nível de descrição Português	String / 45	Sim
	FondsID	Identificador do fundo	Inteiro	Não
	Fonds	Designação do fundo	String / 75	Sim
Reference	Referência do registo descritivo	String / 75	Não	

Tabela 1. Documentação dos atributos das entidades da base de dados

3.3.2. Desenvolvimento

Nesta subsecção será descrita, de forma mais detalhada, o processo de desenvolvimento deste módulo, explicando os passos iniciais desta aplicação e que envolvem a comunicação e interligação de componentes previamente desenvolvidos para o DigitArq, seguida do processo de criação da sua interface, a forma como será feita a gestão dos depósitos e a descrição das funcionalidades de otimização dos mesmos.

3.3.2.1. Comunicação com o servidor

O módulo BackOffice da aplicação DigitArq, bem como o módulo criado neste projeto, fazem pedidos HTTP a um servidor REST.

A arquitetura REST é caracterizada por um conjunto de princípios que tenta minimizar a latência e comunicação de dados na rede, ao mesmo tempo que maximiza a independência e escalabilidade da implementação de componentes [7]. Ou seja, permite o desenvolvimento de serviços *Web*, independentes de plataformas e linguagens (a forma como os serviços *Web* são desenhados nesta arquitetura permite que várias linguagens de programação, mesmo diferentes da usada originalmente para os desenvolver, não tenha quaisquer problemas em aceder a estes serviços e produzir os resultados esperados), ao mesmo tempo que tenta reduzir a utilização de recursos do sistema. É atualmente considerado como o modelo predominante no desenvolvimento de serviços *Web*, devido às vantagens anteriormente citadas e também pela facilidade de utilização, comparativamente a outras arquiteturas e serviços existentes.

O BackOffice possui uma API, que garante o acesso às funções dos *Core Services* presentes no servidor, a partir do qual executa operações sobre registos descritivos, gestão de utilizadores, representações, etc.. O módulo de gestão de depósitos também irá precisar de uma API que permita realizar operações sobre os seus itens, e portanto estes *Core Services* serão expandidos para compreenderem também este módulo. Na secção seguinte, 3.3.2.2, este processo será descrito com maior detalhe.

O acesso a estas APIs e conseqüente manipulação da informação no servidor é dado através do processo de autenticação na aplicação-cliente. No

momento de autenticação, se esta for bem-sucedida, é criada uma instância da classe *CoreServices*, associada ao utilizador e à sua senha, que lhe permite o acesso a todos os serviços disponíveis, e assim realizar todas as operações existentes na aplicação cliente (criar, apagar e modificar registos).

Uma vez que a maioria dos resultados da aplicação das funções disponíveis nas APIs são instâncias das classes desenhadas para realizar a tradução direta da informação proveniente da base de dados para a aplicação e vice-versa (como foi descrito na secção 3.3.1.3), ou então tipos nativos (como inteiros ou *Strings*), não se torna necessário desenvolver mecanismos adicionais para a interpretação da informação, simplificando assim a comunicação entre ambas as partes.

3.3.2.2. Criação de serviços

O DigitArq é uma aplicação orientada a serviços. De entre os vários implementados (a totalidade dos serviços implementados corresponde aos *Core Services*; os serviços são utilizados como instâncias numa classe com esta designação, que dessa forma os encapsula, permitindo um acesso simplificado aos mesmos por parte da equipa de desenvolvimento), podem encontrar-se serviços dedicados a registos descritivos, representações, autenticação e gestão de utilizadores.

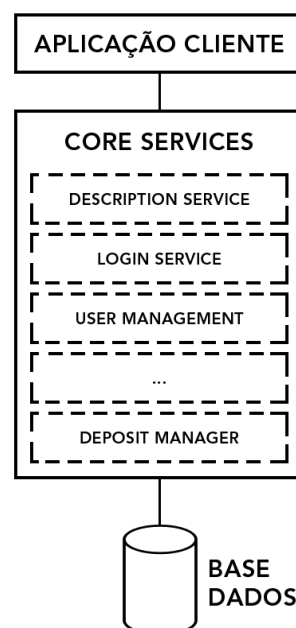


Figura 7. Arquitetura do DigitArq

Uma vez que a gestão de depósitos envolve uma temática que não fora abordada antes pela aplicação, será necessário desenvolver um novo serviço para esse efeito.

Após a definição das funcionalidades necessárias para a gestão de depósitos e a elaboração do respectivo modelo de dados, o passo seguinte consiste na criação de classes que correspondam às entidades encontradas no modelo de dados.

Uma vez que os *Core Services* já existentes utilizam adaptadores (classes responsáveis pela tradução dos atributos das tabelas na base de dados em propriedades que manipulam as variáveis de instância das classes respectivas), a mesma metodologia será utilizada para este módulo, não apenas por uma questão de coerência, mas também pela flexibilidade que estes oferecem na criação de critérios de pesquisa/filtragem dinâmicos para ambas as partes (cliente e servidor).

Terminado este passo, segue-se a implementação das funcionalidades necessárias que irão fazer parte do novo serviço, que será denominado *DepositManagerService*.

Este serviço está orientado à manipulação das classes correspondentes às entidades presentes no modelo de dados (Figura 6). O serviço possui funções para a criação, modificação, remoção e filtragem de depósitos, seus níveis e tipos de contentores, e associação de registos descritivos aos depósitos (este último item está fortemente ligado ao *DescriptionService*, que corresponde ao serviço orientado à manipulação de registos descritivos, devido à sua natureza e importância para este módulo). Para além da criação das funções em si, é necessário criar os contratos (consiste numa definição formal, independente de plataforma ou linguagem de programação, que apresenta as condições necessárias para a utilização de cada função do serviço [15]) que estas cumprem e as respectivas interfaces de utilização.

Com a conclusão destas tarefas, é dado como concluído o processo de criação do novo serviço direccionado à gestão de depósitos, e é iniciada a construção da aplicação-cliente.

3.3.2.3. Interface da aplicação-cliente

Desde que este projeto foi idealizado, ficou estabelecido que as suas interfaces iriam ter um aspecto bastante próximo das apresentadas pelo módulo BackOffice. Desta forma, haveria uma maior coerência e familiaridade relativamente à sua usabilidade por parte dos atuais utilizadores do DigitArq.

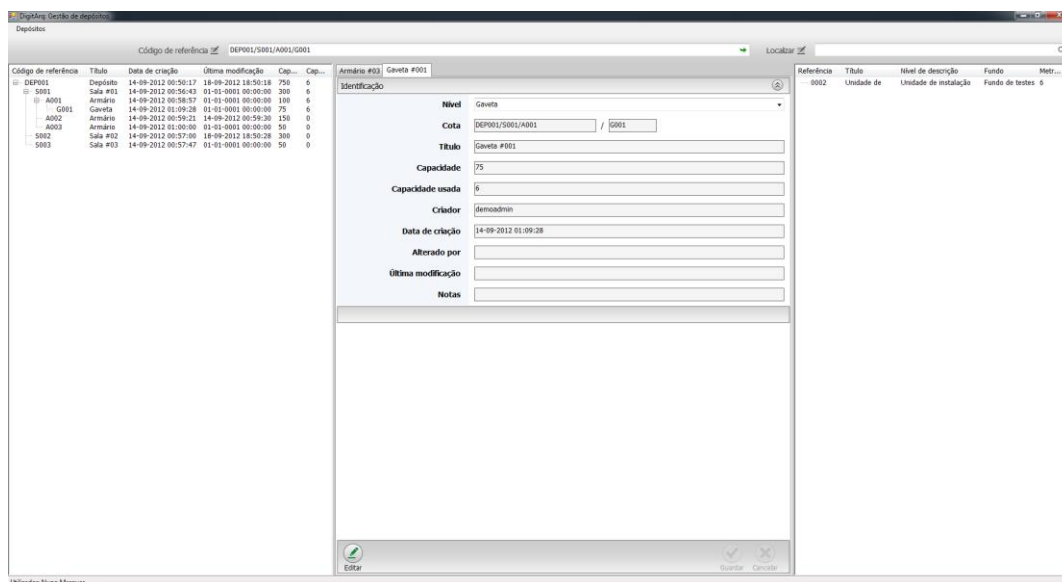


Figura 8. Interface do formulário principal do módulo

Observando a Figura 2 é possível reconhecer a familiaridade do aspecto da Figura 8. No entanto, e apesar de partilharem do mesmo ambiente de desenvolvimento, a interface deste módulo foi totalmente construída de raiz, não tendo sido aproveitado qualquer controlo personalizado previamente criado para o DigitArq BackOffice.

O formulário principal do módulo, que corresponde à Figura 8, é constituído por 3 painéis principais, sendo que o painel central corresponde à área de criação/edição/visualização de níveis de depósitos e registos descritivos, e os restantes representam árvores com funcionalidades avançadas¹¹, com a árvore da esquerda a representar as estruturas físicas dos depósitos abertos durante a execução da aplicação, e a árvore da direita para visualizar os registos descritivos associados a um dado nível de depósito, que esteja no momento a ser visualizado/editado.

¹¹ Controlo personalizado que combina as funcionalidades dos controlos nativos disponibilizados pelo Microsoft Visual Studio 2008, *TreeView* e *ListView*, denominado *Advanced TreeView*

No topo do formulário encontram-se dois mecanismos de pesquisa, sendo que o primeiro é utilizado para procurar depósitos e seus níveis, enquanto o segundo é usado para pesquisar registos descritivos que estejam associados a um qualquer depósito.

O menu do formulário, no topo, contém todas as restantes ações, já referidas: criar, abrir e fechar depósitos, realização de testes de capacidade de níveis e otimização de depósitos.

Os restantes diálogos (formulários adicionais) do módulo seguem a mesma política de interfaces descrita anteriormente, não incluindo apenas alguns dos componentes relativos à pesquisa avançada ou pré-visualização de documentos, que não trazem grande relevância aos objectivos estabelecidos para o projeto neste momento. O aspecto geral destes diálogos pode ser observado na Figura 9, apresentada de seguida.

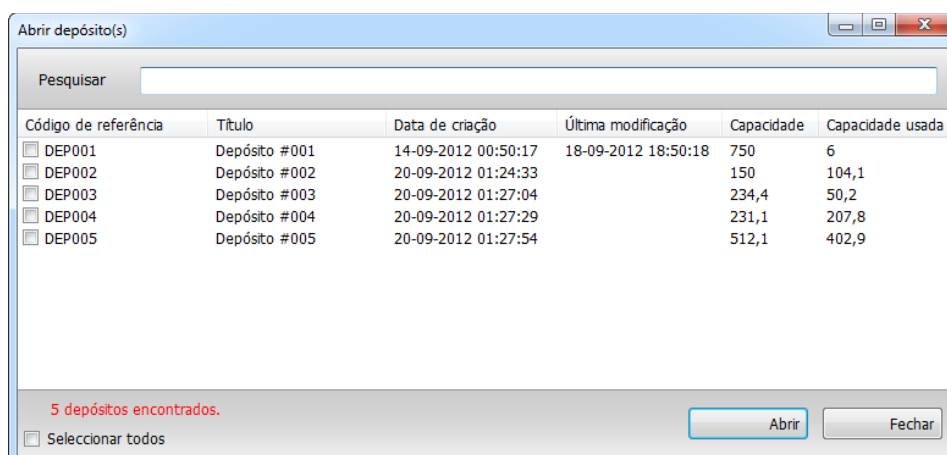


Figura 9. Diálogo usado para a abertura/carregamento de depósitos

Grande parte das funcionalidades está disponível através de menus de contexto¹², disponibilizados em ambas as árvores anteriormente referidas. Em termos de permissões, algumas das funcionalidades estão restritas a tipos de utilizadores. Por exemplo, se o utilizador atual for um administrador, tem controlo total sobre a definição dos depósitos que administra, podendo criar, editar e remover os seus níveis, enquanto um utilizador comum apenas pode associar registos descritivos ou adicionar contentores nesses depósitos, e portanto as funcionalidades a que não tem acesso não lhe são visíveis sequer.

¹² Controlo nativo, normalmente associado a um outro controlo e que disponibiliza um menu quando o utilizador pressiona o botão direito do rato sobre esse controlo

3.3.2.4. Suporte multilingue

Neste projeto, o suporte de múltiplos idiomas não foi considerado como um objectivo estritamente necessário. No entanto, foi feito um esforço desde o início nesse sentido e dessa forma, todo o processo de traduções de elementos da interface da aplicação (mensagens, botões, *labels*, etc.) é feito de forma simples.

O ambiente de desenvolvimento Visual Studio 2008 possui mecanismos nativos que auxiliam este processo, como serviços de localização e *Resources*.

Localização consiste no processo de adaptar a interface de uma aplicação globalizada (que é funcional em múltiplas culturas e locais, incluindo formatos de datas, números, moeda, medidas, etc.) às necessidades únicas de uma dada região, incluindo gráficos, formatos numéricos, idioma, etc. [8].

Em vez de modificar o código-fonte da aplicação, estas alterações (sobretudo idioma e gráficos, uma vez que os restantes formatos são da responsabilidade do sistema operativo) são realizadas em ficheiros XML adicionais, designados *Resources*, que representam o mapeamento dos valores globais para os respectivos das regiões que serão localizadas. Assim, através de simples *Strings* em ficheiros de configuração, é possível definir o comportamento da aplicação das mais variadas regiões geográficas. No caso de uma dada região ou idioma não estar ainda localizada, a aplicação assume as definições por omissão (neste caso, em Português) e carrega-as para a interface.

Existem, obviamente outras ferramentas externas que permitem agilizar o processo de tradução de aplicações .NET, como por exemplo o Multi-Language Add-In for Visual Studio .NET¹³, mas que introduzem código externo na aplicação e aumentam a complexidade de um objetivo que por si só não é relevante na situação atual deste projeto.

No projeto, todos os controlos dos formulários e mensagens (estas apenas podem ser atingidas em certas condições, daí não fazerem parte dos controlos dos formulários) foram traduzidos para Inglês, encontrando-se também disponíveis na língua por omissão, Português.

¹³ <http://www.jollans.com/tiki2/tiki-index.php?page=MultiLangVsNet>

3.3.2.5. Gestão de depósitos

Em termos da gestão dos níveis dos depósitos, existem algumas regras para a sua manutenção:

- **Estrutura hierárquica**

Dada a estrutura hierárquica dos depósitos, cada novo nível de depósito terá que ser guardado na base de dados, antes de lhe poderem ser adicionados subníveis. Ou seja, no momento de criação de um dado nível, não é possível adicionar-lhe logo de seguida subníveis, só após o utilizador o guardar pela primeira vez na base de dados. Só desta forma é possível estabelecer a condição pai-filho no modelo de dados e conseqüentemente na aplicação.

- **Gestão de espaço**

A cada nível de depósito estão associados dois atributos de espaço: atribuído e consumido. Nenhum nível pode ter o valor de espaço atribuído superior aos seus níveis superiores na hierarquia. Da mesma forma, a soma do espaço atribuído a todos os subníveis de um nível não pode ser superior ao espaço atribuído a esse nível. A mesma regra aplica-se para o espaço já consumido. O espaço consumido resulta apenas da associação de registos descritivos ao nível; não existe outra forma de manipular este valor.

A unidade de medida utilizada neste projeto é o metro linear que corresponde à unidade convencional para a determinação de espaço ocupado pelos documentos em suporte analógico, como por exemplo nas estantes [3].

- **Contentores**

Os contentores apenas podem ser definidos nos níveis que não possuam subníveis, ou seja, nos níveis que representam a definição mais fina daquele elemento da estrutura física no depósito (por exemplo, prateleira num armário). A partir do momento em que um contentor é associado a um nível físico, esse nível não poderá ter outros subníveis, a não ser que sejam removidos todos os seus contentores. Cada contentor apenas pode conter registos descritivos oriundos de um único fundo. Ou seja, num contentor não se podem misturar objetos de fundos diferentes.

- **Registos descritivos**

À semelhança dos contentores, os registos descritivos, só podem ser definidos nos níveis que não possuam subníveis (mas podem ter contentores), e da mesma forma, a partir do momento em que essa associação é feita, o nível não poderá ter outros subníveis, a não ser que os registos descritivos desse nível sejam removidos. Por exemplo, se existe um nível 'armário' que possui vários subníveis 'prateleira', a associação de objetos não poderá ser feita ao 'armário', mas sim a uma das 'prateleiras', pois só dessa forma faz sentido.

- **Remoção de níveis**

A remoção de um nível implica a remoção de todo o “ramo da árvore”, i.e., a remoção de todos os subníveis a si associados e conseqüentemente todos os subníveis destes. Desta forma, todos os contentores e registos descritivos associados em qualquer ponto desse “ramo” serão também removidos.

Identificação	
Nível	Armário
Cota	DEP001/S001 / A001
Título	Armário #001
Capacidade	95
Capacidade usada	65
Criador	demoadmin
Data de criação	22-09-2012 21:44:07
Alterado por	nmarques
Última modificação	19-10-2012 01:04:52
Notas	

Editar Guardar Cancelar

Figura 10. Edição de nível de depósito

3.3.2.6. Registos descritivos

Os registos descritivos são os elementos utilizados pelo DigitArq BackOffice para organizar as representações de objetos nos fundos. Podem corresponder a elementos de categorização de hierarquias ou a objetos que estão presentes em fundos arquivísticos e simplificam a organização dos mesmos.

O grau de descrição dos elementos do fundo está diretamente ligado ao detalhe como as hierarquias estão estabelecidas pelo seu administrador ou arquivista responsável por estas tarefas no DigitArq. Por exemplo, um fundo pode ser definido ao ponto de possuir um registo descritivo para cada documento (situação ideal, pois melhora a catalogação e detalhe da informação) ou possuir apenas registos que descrevem as séries que contêm esses documentos, indicando apenas qual o tema dos mesmos e o seu número (situação pouco aconselhável, uma vez que se perde a noção realista da dimensão do fundo e dos documentos nele contido).

Tipo	Designação	Encapsulamento
■ F	Fundo	SF, SC, ST, UI, DC, D
■ SF	Subfundo	SSF, SC, ST, UI, DC, D
■ SSF	Subsubfundo	SC, ST, UI, DC, D
■ SC	Secção	SSC, ST, UI, DC, D
■ SSC	Subsecção	SSSC, ST, UI, DC, D
■ SSSC	Subsubsecção	ST, UI, DC, D
■ ST	Série	SST, UI, DC, D
■ SST	Subsérie	SSST, UI, DC, D
■ SSST	Subsubsérie	UI, DC, D
■ UI	Unidade de Instalação	DC, D
■ DC	Documento composto	D
■ D	Documento simples	–

Tabela 2. Tipos de registos descritivos e respectivos encapsulamentos

A Tabela 2 apresenta as designações e todas as possibilidades de encapsulamento para todos os tipos de registos descritivos existentes no DigitArq. As cores correspondem a auxiliares visuais utilizados pela aplicação para uma mais fácil identificação dos registos e suas hierarquias.

Relativamente aos registos descritivos, apenas 3 tipos de registos serão armazenados, uma vez que os restantes níveis apenas representam uma organização virtual dos objetos presentes nos fundos. Por sua vez, os três níveis utilizados representam fisicamente os objetos quando estes estão no depósito. São eles: unidades de instalação, documentos compostos e documentos simples (também conhecidos como apenas documentos). As suas definições podem ser consultadas na secção 2.3.

Os documentos compostos podem encapsular apenas documentos simples. Os documentos simples, por se tratar de itens indivisíveis, não podem encapsular nenhum outro tipo. Por sua vez, as unidades de instalação podem encapsular documentos compostos e documentos simples. Os vários cenários possíveis são facilmente compreendidos observando a Figura 11.

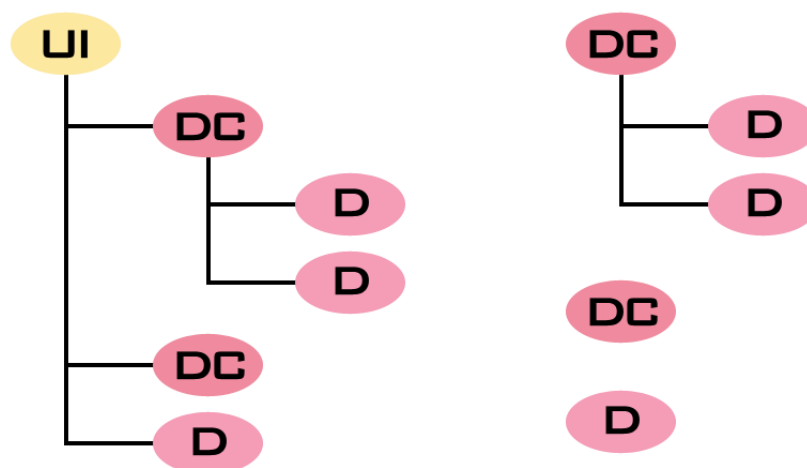


Figura 11. Níveis de registos descritivos e seus possíveis encapsulamentos

No momento de associação, qualquer um destes tipos pode ser inserido como elemento raiz na associação. Neste contexto, por raiz entende-se um registo descritivo cujo nível é um dos três seleccionados, mas o nível imediatamente superior ao seu no fundo em que está inserido não é nenhum dos outros dois. Por exemplo, as UI serão sempre raízes, mas um DC só o poderá ser se o nível imediatamente superior for uma série, secção, ou outro, mas nunca uma UI.

Observando os cenários presentes na Figura 11, quando é seleccionada uma Unidade de Instalação com documentos encapsulados, esta será armazenada juntamente com esses documentos (se os tiver), e o mesmo se sucede com os documentos compostos.

Existem ainda as possibilidades de associar apenas o registo seleccionado, excluindo os registos que encapsula, podendo associá-los mais tarde se assim for desejado; pode ainda associar um tipo de registo que não um dos três descritos anteriormente.

Nesse caso, e como representam uma posição mais elevada na hierarquia dos fundos, podem encapsular um ou vários dos tipos permitidos. Aí serão procurados os elementos raiz (encontrados ao percorrer todos os registos encapsulados pelo registo inicial) e sequencialmente serão inseridos juntamente com os seus subníveis. Obviamente que no caso em que o registo descritivo não possua nenhum dos três tipos seleccionados, nenhum registo descritivo será associado ao depósito.




Informação geral	
Código de referência	PT/TT/NMARQUES_2-SF001/001/00003
Título	Documento composto 2
Fundo	Fundo de testes
Nível de descrição	Documento composto
Controlo	
Criador	nmarques
Data de criação	08-10-2012 22:36:22
Alterado por	nmarques
Última modificação	08-10-2012 22:37:00
Extensão	
Origem de informação	Inferido
Metros lineares usados	5
  	

Figura 12. Detalhes de registo descrito antes de associação a depósito

Na associação de registos pode surgir o caso em que não haja espaço suficiente no nível selecionado para armazenar todos os elementos envolventes. Aqui surgem duas situações particulares, que envolvem a procura de novos níveis e a fragmentação de registos.

Os itens raiz (e respectivos subitens) são tratados pela aplicação de forma sequencial (um de cada vez) e para cada um deles é calculado o valor de metros lineares a si associado (com base nos valores desta extensão para todos os subitens). Esse valor é depois comparado com o espaço físico disponível no nível selecionado pelo utilizador. Se for menor, então será inserido neste nível e passa-se à raiz seguinte. Se não, é procurado um outro nível onde caiba, preferencialmente o mais próximo do inicialmente escolhido. No caso de não existir nenhum nível onde caiba o item, este não poderá ser inserido.

Uma alternativa também utilizada neste projeto consiste na fragmentação dos registos descritivos e será descrito de seguida, na secção 3.3.2.7.

3.3.2.7. Fragmentação de registos descritivos

Quando são associados registos descritivos aos depósitos, é desejável que todos os objetos representados por esses registos caibam no nível selecionado, mas por vezes essa situação não é possível. Surge portanto, a necessidade de “fragmentar” os registos descritivos, por forma a tentar aproveitar o espaço livre nos vários níveis de depósito.

Esta operação consiste em separar um registo descritivo que contenha registos encapsulados, e colocá-los nos níveis mais próximos que possuam espaço para os armazenar.

Dos três níveis de registos descritivos, a fragmentação só é possível com os documentos compostos. Devido à natureza das unidades de instalação (são livros, caixas, pastas), estas não podem ser divididas em partes mais pequenas; por sua vez os documentos simples são indivisíveis. A fragmentação dos documentos compostos também só poderá ocorrer nos casos em que estes não estejam encapsulados numa unidade de instalação, ou seja, têm que ser obrigatoriamente elementos raiz, mas têm que ter obrigatoriamente documentos simples encapsulados.

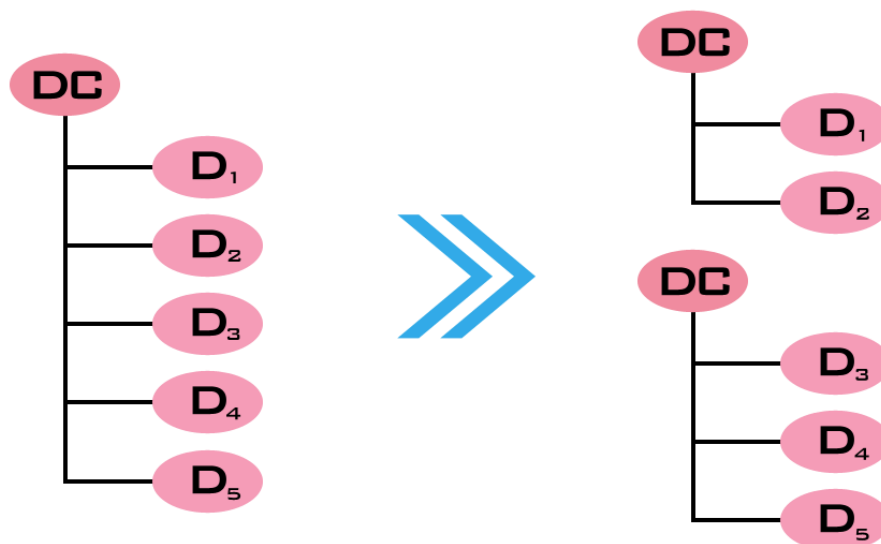


Figura 13. Exemplo de fragmentação de registos descritivos

A Figura 13 mostra um dos vários exemplos da fragmentação de um documento composto (a fragmentação poderia ser maior, podendo neste caso específico resultar numa fragmentação máxima de até 5 blocos), sendo que uma parte irá ficar num nível de depósito e a outra parte irá ficar noutra. A

inclusão do documento composto DC em cada uma das árvores indica a origem dos documentos simples. O documento composto será assim replicado e a informação de maior relevância do mesmo que deverá ser alterada (em cada uma das réplicas) são o valor de metros lineares consumidos, que consistem agora na soma dos metros lineares dos documentos simples associados a cada réplica, e nos seus níveis físicos (que serão diferentes).

Depósito	Nível	Fundo	Registo	Tipo	Fragmentado?
✓ Depósito #001	Prateleira #002	Fundo de testes	Documento composto 2	DC	Sim
✓ Depósito #001	Prateleira #002	Fundo de testes	Documento simples 2	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento composto 2	DC	Sim
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 1	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 3	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 4	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 5	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 6	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 7	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 8	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 9	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 10	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 11	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 12	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 13	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 14	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 15	D	Não
✓ Depósito #001	Prateleira #003	Fundo de testes	Documento simples 16	D	Não
✓ Depósito #001	Gaveta #001	Fundo de testes	Documento composto 2	DC	Sim
✓ Depósito #001	Gaveta #001	Fundo de testes	Documento simples 17	D	Não
✓ Depósito #001	Gaveta #001	Fundo de testes	Documento simples 18	D	Não

21 registos. Fechar

Figura 14. Relatório de associação de registos descritivos num depósito

Após cada nova associação de registos descritivos a depósitos, a aplicação apresenta um breve relatório sobre o resultado da mesma (sucesso, insucesso, níveis físicos a que os registos ficaram associados, etc.). A Figura 14 mostra o resultado da associação de um único Documento Composto que encapsula 18 documentos simples. Devido ao nível físico inicialmente selecionado não possuir capacidade livre suficiente para armazenar o Documento Composto por inteiro, este foi fragmentado (neste caso em três níveis físicos) e disperso no depósito.

As fragmentações são sempre feitas de modo a que os fragmentos sejam armazenados o mais próximo uns dos outros: são testados os níveis imediatamente a seguir ao inicial, depois os imediatamente antes e só depois (da mesma forma) são procurados os níveis mais distantes.

Os fragmentos são criados tentando maximizar a utilização do espaço disponível: sabendo qual o espaço disponível no nível em que se está a tentar inserir, são testados vários registos descritivos que, em conjunto, caibam no nível físico e que deixem o menor espaço livre possível.

Os contentores são ignorados neste processo, uma vez que não suportam a fragmentação de registos descritivos.

Desta forma são reduzidos o número de fragmentos por registo e o número de níveis cujo espaço livre é incapaz de suportar mais registos.

3.3.2.8. Desfragmentação/otimização de depósitos

Ao longo do ciclo de vida de um depósito, espera-se que lhe sejam associados centenas ou mesmo milhares de documentos, o que na aplicação (no que toca à gestão de fundos) pode resultar num semelhante número de registos descritivos.

Se não houver controlo do ponto de vista organizativo, existe a possibilidade de se misturarem registos de séries ou secções diferentes, ou mesmo até de fundos diferentes, o que não representa de todo, situações ideais.

O objetivo da otimização ou desfragmentação de um dado depósito não é mais que uma reorganização dos seus registos, novamente ordenados por fundos e subsequentes ordens internas (subfundos, séries, secções, etc.) ao mesmo tempo que se tenta minimizar o espaço consumido pelos mesmos.

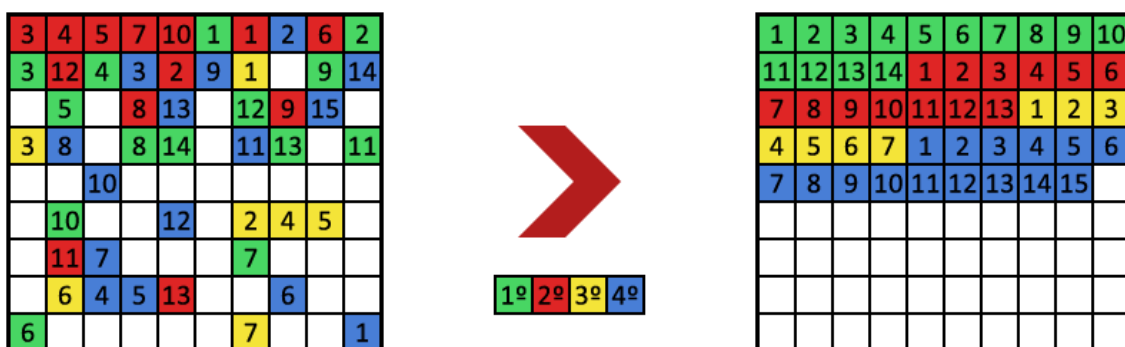


Figura 15. Reorganização de fundos num depósito

No entanto, nenhuma informação será realmente modificada, pois esta reorganização será apenas temporária para que o administrador/gestor do

depósito saiba o que fazer quando decidir melhorar a organização do mesmo. Quando a operação termina, é apresentado uma nova cópia do depósito contendo as novas posições dos registos, bem como um relatório com todos os movimentos e operações necessárias para atingir tal estado, a partir da situação atual do depósito.

O processo de otimização começa com a ordenação dos fundos, ordem esta que deverá ser definida pelo utilizador, tal como pode ser observado na Figura 16. No depósito selecionado são obtidos todos os fundos que possuam associações de registos seus no mesmo. A ordem definida pelo utilizador (ao mover os fundos para cima e para baixo na lista) corresponde à ordem pela qual os fundos serão tratados e armazenados no depósito (de cima para baixo).

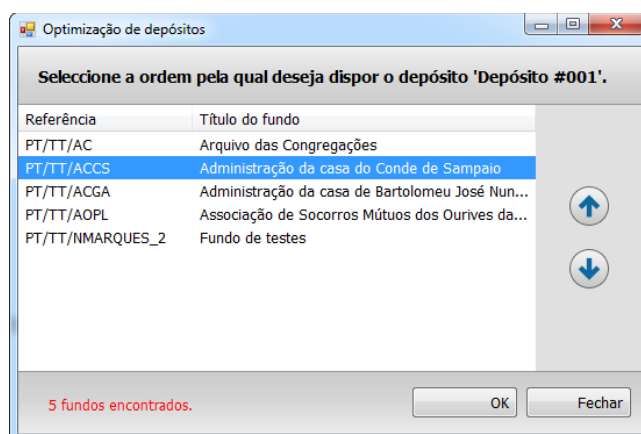


Figura 16. Preparação de otimização de um depósito

O passo seguinte passa por reconstruir as possíveis hierarquias existentes nos fundos e que foram destruídas nas associações feitas pelos utilizadores. Por exemplo, e por vários motivos, na primeira árvore apresentada na Figura 11, cada elemento poderia ter sido inserido individualmente no depósito, apesar de originalmente definirem uma hierarquia/encapsulamento no fundo. Este passo é responsável pela reconstrução de casos deste género, bem como desfazer a fragmentação de registos descritivos (o caso inverso ao descrito na secção 3.3.2.7 e observado na Figura 13).

Os contentores existentes no depósito também terão que ser movidos, para ficarem junto dos restantes elementos do fundo a que estão associados

- **Movimentação de contentores**

Com a movimentação dos objetos dos fundos no depósito, é necessário mover também os contentores existentes. No entanto, existe uma particularidade acerca dos contentores. Uma vez que geralmente ocupam muito espaço no depósito, é necessário que todos os contentores sejam realocados antes dos restantes objetos para garantir que irão caber no depósito.

Um exemplo do que poderá acontecer de errado se for tomada esta decisão é o seguinte: suponha-se que existe um contentor que cabe apenas em um dos vários níveis do depósito, uma vez que todos os outros níveis têm uma capacidade inferior à que o contentor necessita. Se a operação de movimentação de objetos colocar registos nesse nível, no momento de mover o contentor de volta para esse nível, já não caberá lá por estar preenchido com outros objetos e portanto a operação de otimização irá falhar.

Desta forma, os contentores serão organizados pela ordem dos fundos, mas não irão ficar necessariamente junto dos restantes objetos do seu fundo, para prevenir situações como a descrita.

- **Movimentação de objetos**

Após a reconstrução das hierarquias e selecionada a nova ordem dos fundos, é altura de indicar a nova posição dos seus objetos no depósito. Esta operação indica para todos os registos qual o movimento realizado, referindo a sua origem e o seu destino, em termos de níveis do depósito.

- **Fragmentação de documentos compostos**

Por forma a otimizar o espaço consumido, alguns registos serão fragmentados de forma semelhante à descrita na secção 3.3.2.7., garantindo assim que o espaço disponível será aproveitado o máximo possível.

Sempre que se justifique, as operações indicam as mudanças de posição no depósito (nível de origem e de destino) e as mudanças de hierarquias (registo de origem e de destino), tornando mais intuitivo o processo de análise do relatório.

Na Figura 17 é possível observar parte do relatório de otimização. Algumas das ações realizadas num dos fundos deste depósito são a desfragmentação de registos, correção de hierarquias e finalmente a movimentação de objetos. Por forma a tornar mais simples a leitura do relatório, o utilizador poderá ver as operações de um único fundo, bastando selecioná-lo na lista apresentada na zona direita do formulário.

Todas estas operações são efectuadas num novo depósito temporário criado para o efeito. Ou seja, os dados presentes no depósito sobre o qual estão a ser executados testes de performance não serão alterados, apenas lidos. Todas as operações de escrita são feitas sobre este novo depósito.

Quando o utilizador termina a sua análise do relatório final, este depósito poderá ser então eliminado.

4. Performance da aplicação

Ao longo do desenvolvimento desta aplicação surgiram várias funcionalidades (e código associado) que era sabido à partida que poderiam causar problemas de performance, devido à sua complexidade, e à quantidade de dados manipulados durante a sua execução.

Os principais congestionamentos centram-se sobretudo nas operações de associação de registos descritivos aos depósitos, "desfragmentação" ou otimização de depósitos, e carregamentos de nodos para as árvores utilizadas no formulário principal (consultar secção 3.3.2.3 e Figura 8). Parte destes problemas foi resolvido, recorrendo a soluções alternativas, enquanto outros, devido à sua natureza, apenas se conseguiu melhorar ligeiramente a sua performance. De seguida, será feita uma análise caso-a-caso dos problemas encontrados e respectivas soluções utilizadas.

4.1. Carregamento de TreeViews

No processo de definição de depósitos, quanto maior o detalhe, maior será o número de níveis presentes, o que resulta numa árvore com mais nodos. Da mesma forma, quanto maior for o nível de registos descritivos associados a um dado nível, mais registos serão apresentados na árvore de registos descritivos (a cada momento, a árvore apresentada à direita no formulário principal mostra apenas os registos descritivos associados ao nível de depósito selecionado). Portanto, a cada nova seleção de um nível de depósito, é necessário carregar todos os registos descritivos a si associados, o que implica esperar que todos os nodos da árvore sejam carregados.

Para remediar esta situação, e reduzir substancialmente o carregamento de nodos são carregados apenas as raízes da árvore e nas raízes que possuam filhos, é colocado apenas um nodo fictício (apenas para que o utilizador saiba que aquele nodo em particular possui subnodos). Quando o utilizador carrega pela primeira vez no botão que expande essa subárvore, o nodo fictício é removido e são carregados os filhos desse nodo. Para os netos é aplicada a mesma política do nodo fictício. Por formar a melhorar ligeiramente a performance desta operação, poderá também ser usada uma

Thread adicional para carregar os nodos enquanto a aplicação espera, impedindo que esta possa bloquear.

Por exemplo, para uma árvore que possua cerca de 200 nodos, se todos os nodos e subnodos forem carregados de uma só vez, a aplicação pode ficar bloqueada durante vários segundos. Com esta alternativa, não só serão carregados menos nodos de cada vez, como melhora ligeiramente a rapidez da aplicação.

4.2. Otimização de depósitos

A execução dos algoritmos de otimização de depósitos passa por várias fases, tal como foi descrito na secção 3.3.2.8, sendo inicialmente obtidos todos os fundos associados ao depósito, e depois de ordenados, são recriadas as suas hierarquias, finalizando com a inserção de todos os registos novamente numa cópia do depósito.

No entanto, uma parte destas tarefas, e a mais complexa, pode ser executada concorrentemente. A recriação das hierarquias de cada fundo, apenas depende dos registos descritivos que pertencem a esse fundo, não havendo cruzamento de dados entre os vários fundos. Assim, depois de o utilizador definir a ordem dos fundos, são criadas *Threads* adicionais (tantas quanto o número de fundos encontrados) e a recriação das hierarquias é feita de forma concorrente para os vários fundos.

Com a realização de testes, antes e após a criação desta alternativa, foi possível observar melhorias significativas na execução do algoritmo. Na maior parte dos casos (com um número considerável de registos descritivos), a execução sequencial poderia levar entre 2 a 5 minutos. Após a implementação de concorrência no algoritmo, o seu tempo de execução passou de 10 a 40 segundos, o que revela uma melhoria bastante significativa.

Quanto ao restante algoritmo, terá sempre que ser executado de forma sequencial, pois a posição de inserção dos registos descritivos de um fundo nos níveis de depósito depende sempre do resultado da inserção do fundo anterior (devido a possíveis fragmentações, ou falta de espaço de um nível).

4.3. Associação de registos descritivos

Devido à complexidade do algoritmo de associação de registos descritivos, não foi possível melhorar muito a sua performance e consequente tempo de execução. A quantidade de testes e verificações adicionais é tão grande, e a sua execução é obrigatoriamente sequencial, que as melhorias apenas foram conseguidas apenas nas funções mais simples (e ao mesmo tempo, menos relevantes).

Antes de inserir um registo (ou árvores de registos), é necessário antes de tudo obter os níveis raiz a serem inseridos. Isto simplifica o processo, pois permite filtrar os itens a serem inseridos em cada subárvore e assim dividir as tarefas por cada uma delas. Seguidamente deve verificar-se se algum dos registos de cada árvore já existe no depósito. Isto implica uma procura exhaustiva por todos os registos a serem inseridos e verificar se no depósito já existem. Depois são feitos cálculos sobre o total de metros lineares consumidos por esta tentativa de inserção, pois dependendo desse valor, do nível de descrição e do nível físico onde se pretende inserir, os resultados podem variar entre inserir corretamente, de forma fragmentada ou não inserir de todo, sendo que para estes três casos existem mais verificações adicionais.

Ou seja, a complexidade deste algoritmo é bastante elevada, no entanto foi feito um esforço por tentar otimizar as várias funções envolvidas no processo, tendo conseguido algumas pequenas melhorias.

Ainda assim, se se tentar associar um número muito grande de registos descritivos de uma só vez (por exemplo, introduzir todo o fundo de uma assentada), o algoritmo poderá levar alguns minutos a terminar, até porque mesmo com tentativas de melhorar o código das funções, cada registo descritivo tem de ser acedido e tratado individualmente, o que faz com os pedidos aos *Core Services* (e consequentemente à base de dados) sejam elevados, o que acaba por deteriorar um pouco a sua performance.

5. Conclusão

Neste capítulo será feita uma síntese relativa a todo o trabalho desenvolvido e possíveis trabalhos futuros, que incluem melhorias ao estado atual do módulo e à introdução de novas funcionalidades.

5.1. Síntese

A preservação de património arquivístico assume um papel fundamental na sociedade, na medida em que auxilia o estudo do nosso passado (como por exemplo, no estudo da genealogia). Assim, é de extrema importância uma boa catalogação destes documentos e objetos, e as tecnologias atuais são agora mais relevantes, pois permitem a fácil disseminação desta informação, bem como cópias digitais da mesma, garantindo desta forma que a informação nunca mais se perca e que sobretudo esteja acessível em qualquer parte do mundo, sem que seja necessário visitar um arquivo, biblioteca ou museu para a consultar.

Em Portugal, a gestão dos arquivos nacionais e regionais é assegurada pela aplicação DigitArq, que permite a catalogação de fundos e dos seus documentos. No entanto, esta aplicação não permite gerir os depósitos em que estes fundos são armazenados, nem ter uma noção do espaço físico envolvido neste processo. O objetivo central desta dissertação passou por tentar criar uma solução para este problema e simultaneamente auxiliar o processo de otimização de depósitos.

Para tal foi analisado o funcionamento dos arquivos, das normas (nacionais e internacionais) que os regem, a forma como são geralmente estruturados (especificamente, foi feita uma visita ao Arquivo Distrital de Braga, onde foram apresentadas as salas/depósitos principais e descrito o processo de organização dos seus fundos) e finalmente o estudo da aplicação DigitArq da KEEP Solutions.

O resultado deste projeto consiste num módulo de *software* que fornece as ferramentas básicas necessárias para a gestão de depósitos, permitindo a associação de registos descritivos oriundos em fundos a depósitos, obter o espaço consumido pelos mesmos e permitir também sugestões de otimização

dos espaços ao promover alterações na forma como os registos descritivos estão organizados. É expectável que com o passar do tempo esta aplicação atinja um estado mais maduro, ao serem acrescentadas funcionalidades como as sugeridas na secção seguinte, 5.2, incluindo outras que surjam através do *feedback* dos utilizadores principais da aplicação, os arquivistas e os gestores de depósitos, e que sabem de facto o que é necessário que esta ferramenta seja capaz de realizar.

5.2. Trabalho futuro

Uma vez que a realização deste projeto consistia na criação de um módulo experimental ou protótipo para a gestão de depósitos, algumas possíveis funcionalidades não foram abordadas e outras não foram aprofundadas o suficiente para garantir a melhor performance possível à aplicação. Além disso, o objetivo principal passava por provar que era possível criar um módulo deste tipo que servisse às necessidades dos arquivistas e gestores de depósitos, e que fosse capaz de ser integrado nos restantes, que constituem o projeto DigitArq. Esse objetivo foi cumprido.

No entanto, e como foi referido anteriormente, algumas funcionalidades interessantes, e que poderiam melhorar a usabilidade do módulo, não foram implementadas.

Por exemplo, os mecanismos de pesquisa, apesar de existirem, são simples (o utilizador introduz a cota do nível do depósito ou a referência do registo descritivo e estes são carregados na aplicação, se existirem); com a implementação de mecanismos de pesquisa avançada que, nos depósitos, permitisse filtrar por designação, tipo de nível, por espaço disponível, etc., e nos registos descritivos permitisse a filtragem por designação, nível descritivo, fundo, etc., poderiam ajudar os utilizadores a encontrarem mais facilmente o que procuram.

Em termos de usabilidade, a possibilidade de *drag-n-drop* traria mais rapidez na movimentação de nodos associados a níveis físicos ou registos descritivos nas respectivas árvores. No entanto, foram introduzidas funções nas árvores que permitem movimentar os seus nodos para cima e para baixo. A partir do momento em que este módulo esteja integrado com os restantes,

surge a oportunidade de levar a capacidade de *drag-n-drop* a outro nível, permitindo arrastar registos descritivos diretamente do módulo de BackOffice para o de gestão de depósitos, simplificando significativamente o processo de associação de objetos aos depósitos.

A utilização das funções existentes nos *Core Services* resultam em alguns casos na comunicação de demasiados objetos, por vezes desnecessários, o que aumenta consideravelmente o tráfego utilizado pela aplicação e reduz a sua performance ao tentar filtrar esses casos desnecessários. Assim, com base na realização de testes de utilização mais compreensivos, algumas destas funções poderão ter que ser afinadas, por forma a reduzir possíveis congestionamentos durante a execução da aplicação. Em alguns casos poderá ser mesmo necessário escrever *queries* SQL mais complexas e aplicá-las diretamente na base de dados, em vez de passar pelos processo de tradução descritos nas secções 3.3.2.1 e 3.3.2.2, pois as *queries* permitem fazer uma filtragem mais complexa e abrangente, diretamente a partir do servidor de base de dados, o que poderá resultar no aumento da performance da aplicação cliente.

Apesar das possíveis melhorias referidas nesta secção, as mesmas não invalidam de forma nenhuma o trabalho realizado e descrito nesta dissertação, uma vez que se tratam sobretudo de questões relacionadas com a melhoria de performance e usabilidade da aplicação desenvolvida.

6. Referências bibliográficas

- [1] Decreto-Lei nº 93/2007 de 5 de Abril.
http://dgarq.gov.pt/files/2008/09/93_2007.pdf, (Acedido: Junho 2012).
- [2] CONNOLLY, T., AND BEGG, C. *Database systems: a practical approach to design, implementation, and management*. International computer science series. Addison-Wesley, 2005.
- [3] DIREÇÃO GERAL DE ARQUIVOS. Questionário à situação arquivística do Estado: Glossário.
http://dgarq.gov.pt/files/2012/04/Glossário_v2.pdf, (Acedido: Junho 2012).
- [4] COMITÉ DE NORMAS DE DESCRIÇÃO DO CONSELHO INTERNACIONAL DE ARQUIVOS. *ISAD(G): Norma geral internacional de descrição arquivística: Adoptada pelo Comité de Normas de Descrição, pag. 45, Estocolmo, Suécia, 19-22 de Setembro de 1999*. Standards. Conselho Internacional de Arquivos, 2000.
- [5] INSTITUTO PORTUGUÊS DO PATRIMÓNIO CULTURAL. Decreto-Lei nº 149/83 de 5 de Abril.
http://dgarq.gov.pt/files/2008/10/149_831.pdf, (Acedido: Junho 2012).
- [6] ARQUIVO DISTRITAL DO PORTO. DigitArq: Produção, conversão e gestão de conteúdos digitais de arquivo.
http://www.adporto.pt/index.php?option=com_content&task=view&id=17&Itemid=93&limit=1&limitstart=1&lang=pt, (Acedido: Abril 2012).
- [7] FIELDING, R., AND TAYLOR, R. *Principled design of the modern web architecture*. ACM Trans. Inter. Tech., 2, 2002.
- [8] MSDN. Overview of Globalization and Localization.
<http://msdn.microsoft.com/en-us/library/aa292205.aspx>, (Acedido: Agosto 2012).

[9] SOCIETY OF AMERICAN ARCHIVISTS. ENCODED ARCHIVAL DESCRIPTION WORKING GROUP AND LIBRARY OF CONGRESS. NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE. *Encoded Archival Description tag library, version 2002*. EAD technical document. Society of American Archivists, 2002.

[10] INTERNATIONAL COUNCIL ON ARCHIVES. *ISAAR (cpf): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*. Standards / International Council on Archives. International Council on Archives, 2004.

[11] INTERNATIONAL COUNCIL ON ARCHIVES COMMITTEE ON DESCRIPTIVE STANDARDS. *ISAD(G): General International Standard Archival Description: adopted by the Committee on Descriptive Standards, pag. 7, Stockholm, Sweden, 19-22 September 1999*. Standards. International Council on Archives, 2000.

[12] INTERNATIONAL COUNCIL ON ARCHIVES COMMITTEE ON DESCRIPTIVE STANDARDS. *ISAD(G): General International Standard Archival Description: adopted by the Committee on Descriptive Standards, pag. 10-11, Stockholm, Sweden, 19-22 September 1999*. Standards. International Council on Archives, 2000.

[13] KEEP SOLUTIONS. DigitArq: Archival management software.
http://www.keep.pt/?page_id=289&lang=en, (Acedido: Dezembro 2011)

[14] KEEP SOLUTIONS. Gestão de Arquivos Definitivos.
http://www.keep.pt/?page_id=704&lang=en, (Acedido: Dezembro 2011).

[15] ZAPTHINK. What belongs in a Service contract?
<http://www.zapthink.com/2005/08/24/what-belongs-in-a-service-contract/>,
(Acedido: Agosto 2012).

