

Universidade do Minho

Conselho de Cursos de Engenharia

Licenciatura em Engenharia de Sistemas e Informática



Disciplina de Opção III

Ano Lectivo de 2006/2007

PDA's no Levantamento de Informação em Arquivos Históricos

Flávio Miguel Xavier Ferreira (38187)

Hugo José Pereira Pacheco (38204)

José Filipe Henriques Vilas Boas (38224)

Supervisão: Pedro Rangel Henriques

22 de Janeiro de 2007

Data de Recepção	
Responsável	
Avaliação	
Observações	

Dedicatória

Dedicamos este trabalho e todo o esforço nele contido ao nosso Mui Nobre Curso e a todos os que dele fazem parte, em especial o homem que muito temos em consideração Pedro Rangel Henriques.

Resumo

Os estudos demográficos consistem no estudo das populações de um dado período histórico através da análise estatística dos registos que documentam as tradições e costumes dessa época.

Este projecto enquadra-se num trabalho de desenvolvimento de software de suporte à recolha e ao processamento de informação de registos demográficos em dispositivos móveis, nomeadamente PDAs, no âmbito do projecto SEED.

O método tradicional de recolha aplicado na demografia baseia-se no preenchimento de formulários manuscritos, pelo que esta é a forma mais nativa de recolha de informação para os historiadores, visando este projecto aplicar o mesmo conceito ao dispositivo em estudo.

No entanto, um registo demográfico pode ser visto como uma partícula de informação útil em diversos campos das Ciências Sociais, para o estudo de língua, tradições e costumes dos nossos antepassados. Desta forma, são exploradas igualmente técnicas de transcrição completa dos registos para formato digital, com a inclusão de diversos automatismos no sentido de viabilizar e agilizar esta solução.

Todo o processo de recolha dos registos tem em vista o armazenamento da informação em bases de dados distribuídas, que possuem mecanismos próprios de sincronização com uma base de dados central.

Área de Aplicação: Recolha e análise de registos demográficos

Palavras-Chave: Registos Demográficos, Transcrição de Documentos, PDA, Pocket PC, C#, Compact .Net Framework, .Net Framework, XML, XML Schema, SQL, SQLite.

Conteúdo

1	Introdução	5
1.1	Apresentação do Caso de Estudo	5
1.2	Motivação e Objectivos	5
1.3	Estrutura do Relatório	6
2	Contextualização do problema	7
2.1	Módulo de Aquisição de Dados	9
2.2	Módulo de Conversão e Fusão de BDP's - ConFusBDP	9
2.3	XML na Demografia Histórica: Anotação de Registos Paroquiais	9
3	Ideia a Desenvolver	11
3.1	Mobilidade da Base de Dados	12
3.2	Aquisição de Dados	13
3.2.1	Recolha de Dados Baseada no Preenchimento de Formulários	13
3.2.2	Transcrição Completa de Registos com Anotações	14
3.3	Navegação pela Informação dos Indivíduos e Famílias	16
4	Implementação	17
4.1	Tecnologias	18
4.1.1	Base de Dados	18
4.1.2	GraphViz - WebDot	20
4.1.3	GPS	21
4.1.4	Reconhecimento de Voz	21
4.1.5	XML	21
4.1.6	Plataforma .Net	22
4.2	Módulo de Especialização e Conversão para Pocket BDP	23
4.3	Camada de Base de Dados	24
4.4	Módulo de Recolha de Dados Baseada no Preenchimento de Formulários	25
4.5	Módulo de Navegação pela Informação dos Indivíduos e Famílias	27
4.6	Módulo de Transcrição de Registos para Documentos XML	29
4.6.1	XML Schema	30
5	Análise de Resultados	35
5.1	Dificuldades de Implementação	35
6	Conclusão	38
6.1	Trabalho Futuro	39

<i>CONTEÚDO</i>	2
A XML Schemas Utilizados	45

Lista de Figuras

2.1	Exemplo de um registo demográfico	7
2.2	Arquitectura do sistema SEED	8
3.1	Arquitectura global do sistema a desenvolver	11
3.2	Esquema possível para o módulo de navegação pela informação dos indivíduos e famílias	16
4.1	Estrutura e interacção entre os módulos desenvolvidos para o PDA	17
4.2	Menu inicial da aplicação de recolha de informação	18
4.3	Arquitectura do sistema geral e interacção com os outros módulos do SEED	19
4.4	Exemplo da conversão de uma BDC para uma Pocket BDP	23
4.5	Exemplo da conversão de uma Pocket BDP para uma BDP	24
4.6	Interacção entre os formulários do programa: Janela Principal	26
4.7	Interacção entre os formulários do programa: Baptismo	26
4.8	Interacção entre os formulários do programa: Casamento	27
4.9	Interacção entre os formulários do programa: Óbito	27
4.10	Exemplo da pesquisa de indivíduos semelhantes num registo de baptizado	28
4.11	Exemplo da recolha de um registo de casamento	29
4.12	Interacção possível na navegação pela informação em indivíduos e famílias	30
4.13	Exemplo da consulta da informação de um indivíduo	31
4.14	Exemplo de alteração da informação de um indivíduo durante a consulta desta	32
4.15	Exemplo da anotação de um registo de baptizado	34
4.16	Exemplo do formulário de normalização de uma localidade	34
A.1	Schema para registos de baptizado	45
A.2	Schema para registos de casamento	46
A.3	Schema para registos de óbito	47

Lista de Tabelas

3.1	Registo de baptismo	14
4.1	Registo de baptismo com anotações	33

Capítulo 1

Introdução

1.1 Apresentação do Caso de Estudo

Este projecto contribui para o projecto de desenvolvimento SEED, no contexto de uma parceria entre o Departamento de Informática e o NEPS (Núcleo de Estudos de População e Sociedade).

O NEPS usa o Método de Reconstituição de Paróquias (MRP), criado por Norberta Amorim, para realizar estudos sobre os hábitos e evolução da População e Sociedade desde 1640 aos nossos dias. O SEED, uma aplicação informática desenvolvida em torno duma base de dados paroquial, vem oferecer o suporte informático ao MRP para recolha, armazenamento e análise dos dados.

1.2 Motivação e Objectivos

Uma vez que a área de aplicação deste projecto são as Ciências Sociais, a principal motivação é desenvolver aplicações intuitivas para pessoas não directamente relacionadas com o campo da informática, o que exige uma adaptação dos conceitos e soluções desenvolvidos. O objectivo deste projecto é desenvolver as interfaces e o software necessários para tornar um dispositivo móvel, nomeadamente um PDA, num instrumento realmente útil em Ciências Sociais para uso nos "trabalhos de campo" (arquivos, conservatórias, bibliotecas, etc.), aquando da deslocação dos historiadores para recolha de dados junto das suas fontes documentais, concretamente para levantamento de registos paroquiais de baptizado, casamento e óbito.

Em termos de sistemas informáticos, a motivação da nossa parte é aumentar a experiência na utilização de tecnologias .Net e da recente linguagem C# e adquirir metodologias e técnicas próprias da programação direccionada a ambientes embebidos.

Definidas as motivações, o objectivo técnico é demonstrar as potencialidades de um PDANA visualização de grandes quantidades de informação, através de aplicações completas e bem programadas, que se revelem úteis no contexto da História e que transcendam o âmbito académico em que se inserem e ganhem aplicação no mundo real.

1.3 Estrutura do Relatório

Após este capítulo de introdução, segue-se um capítulo de contextualização do problema, no qual se refere de forma mais detalhada o enquadramento deste projecto no seio do projecto SEED e no trabalho de desenvolvimento já realizado.

De seguida apresenta-se a ideia a desenvolver, com a descrição dos dois módulos de aquisição de dados concebidos.

À ideia concebida segue-se a fase de implementação dessa mesma ideia, com um estudo detalhado das tecnologias envolvidas no processo.

Os dois capítulos seguintes definem a análise de resultados expectados e conseguidos e uma conclusão geral do projecto.

Por último, como apêndice, são disponibilizados a bibliografia relevante, uma lista de siglas e acrónimos específicos da área de estudo ou tecnologias utilizadas e uma versão simplificada da estrutura definida para os registos estudados.

Capítulo 2

Contextualização do problema

A demografia histórica é um ramo da ciência histórica que procura reconstruir a estrutura populacional de uma determinada sociedade e analisar e explicar as mudanças nessa estrutura através do tempo.

Os estudos demográficos efectuam um “corte transversal“ da sociedade em estudo, para um dado momento histórico. Para isso, baseiam-se na recolha e análise estatística dos registos demográficos de uma época, documentos que registam os eventos mais importantes da vida de um indivíduo e da sua família.

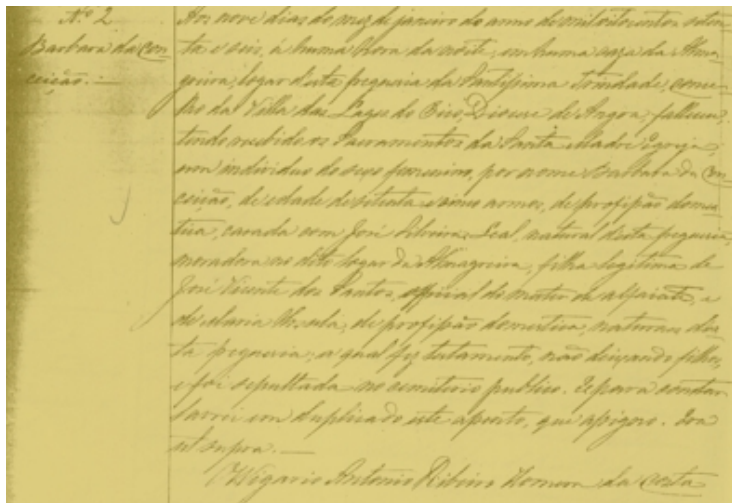


Figura 2.1: Exemplo de um registo demográfico

A recolha dos registos demográficos implica um método de classificação e levantamento de informação coerente. O método referido neste projecto é o Método de Reconstituição de Paróquias (MRP), desenvolvido por Norberta Amorim[16], que propõe a seguinte classificação dos registos:

- registos de baptismo

- registos de casamento
- registos de óbito

Através do cruzamento dos diversos registos é possível recriar as relações familiares entre indivíduos e reconstruir a árvore genealógica de uma família.

O MRP foi inicialmente desenvolvido com o intuito de armazenar os registos na forma de formulários manuscritos.

No entanto, o grande volume de dados e número de registos que caracterizam os estudos demográficos suscitam a necessidade de informatização do repositório de formulários.

Neste contexto surge o projecto SEED, um Sistema de Informação para o Estudo da Evolução Demográfica, composto por uma arquitectura em 4 camadas, descritas de seguida:

Camada de Aquisição de Dados possui as interfaces para a recolha de informação a partir dos registos demográficos;

Camada de Conversão e Fusão de BDPs responsável por fundir os dados de uma paróquia na BDC. Envolve a resolução de conflitos e de colisões;

Camada de Análises permite a realização de diversos tipos de análises sobre a informação dos RPs armazenados;

Camada de Gestão do Conhecimento inclui funções de armazenamento, verificação, validação e divulgação da informação.

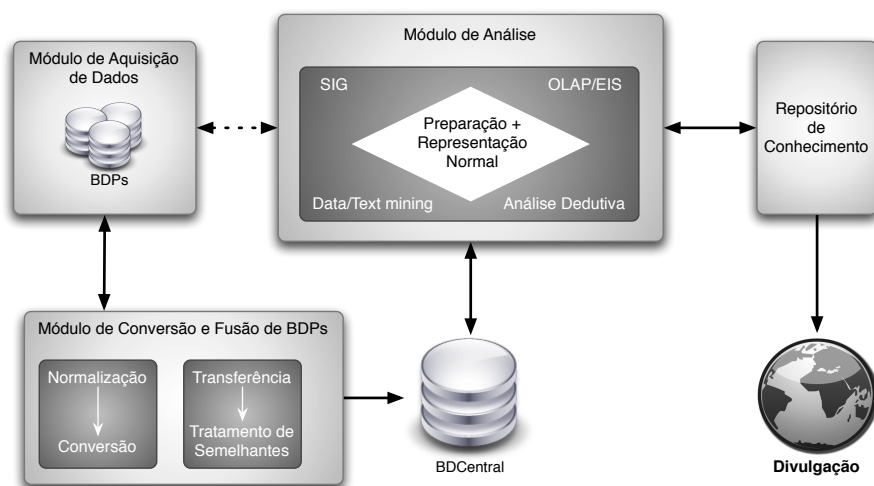


Figura 2.2: Arquitectura do sistema SEED

Este capítulo refere algum trabalho relevante para o nosso projecto já efectuado no seio do projecto SEED, quer ao nível dos estudos de anotação para transcrição de registos e compactação de informação visual efectuados, quer ao nível dos dois módulos desenvolvidos.

2.1 Módulo de Aquisição de Dados

O estudo da demografia baseia-se no comportamento das populações ao longo dos anos. Estes comportamentos são avaliados consoante a natalidade, nupcialidade, e mortalidade das populações. Actualmente, existem grandes arquivos de dados que possuem toda esta informação acerca das populações, mas antigamente estes registos estavam ao cargo da Igreja e armazenados localmente nas paróquias, sem qualquer tipo de ligação entre os diversos registos das diversas paróquias, o que dificulta os estudos dos historiadores.

A aplicação SRP (Sistema de Reconstituição de Paróquias)[17] tem como principal objectivo auxiliar os historiadores na recolha e consolidação desta mesma informação. Através desta aplicação, os demógrafos poderão agrupar os diversos registos numa base de dados central, de forma a concretizarem os seus estudos da evolução demográfica.

Esta aplicação foi baseada nos conhecimentos adquiridos ao longo dos anos pelos historiadores e em anteriores técnicas de levantamento de informação desenvolvidas pelos mesmos, nomeadamente o Método de Reconstituição de Paróquias (MRP).

2.2 Módulo de Conversão e Fusão de BDP's - ConFusBDP

Esta aplicação é responsável pela fusão das diversas BDPs numa BD Central, o que constitui um processo essencial de normalização e tratamento da informação.

O conceito de Base de Dados Paroquial surgiu da necessidade de deslocação dos historiadores para zonas remotas e pouco desenvolvidas, e da necessidade de recolha de informação nessas localidades. Desta forma, toda a informação recolhida é armazenada temporariamente em bases de dados locais, que são futuramente fundidas com a base de dados central através do módulo ConFusBDP. Este módulo é responsável por efectuar a junção das duas bases de dados e resolver todo o tipo de conflitos que possa existir na informação de ambas. Neste processo de normalização de dados é ainda realizada a fusão de BDPs, com detecção e tratamento de indivíduos semelhantes.

Uma outra funcionalidade menos importante deste módulo é permitir a conversão de BDPs de versões anteriores.

2.3 XML na Demografia Histórica: Anotação de Registos Paroquiais

Um dos artigos publicados[20] refere a aplicação do XML ao domínio da Demografia, nomeadamente na anotação dos registos paroquiais estudados pelos historiadores demográficos. Este artigo estuda as similaridades entre os três tipos de registos e propõe uma estrutura de representação de um registo único, genérico aos três tipos de registos existentes. Esta estrutura é descrita numa instância XML Schema, guardada no formato XML, à qual os registos anotados deverão obedecer.

A anotação dos registos originais permitirá armazenar uma cópia integral de cada registo na base de dados, em formato electrónico. Esta existência dos registos transcritos na base de dados adiciona novas funcionalidades tais como:

- a reconstrução das fontes históricas originais, reproduzindo-as em diferentes meios, facilitando a sua disseminação;
- a extracção dos dados necessários ao carregamento das referidas Bases de Dados Paroquiais;
- a análise dos documentos por diversas áreas de estudo, como por exemplo, a linguística.

Este artigo evidencia ainda o processo de edição anotada dos documentos e exemplifica duas transformações possíveis dos mesmos: a visualização em HTML; e a geração automática de SQL para alimentação das Bases de Dados.

Capítulo 3

Ideia a Desenvolver

O objectivo deste projecto é explorar formas de recolha de registos demográficos para dispositivos móveis, mais precisamente, PDAs, com apoio no vasto trabalho de desenvolvimento já feito nesta área.

Isto envolve um processo de adaptação da BDC, com toda a informação existente, a um formato e estrutura que sejam adequados à plataforma Pocket PC e ao modo de operação da aplicação desenvolvida. Dotado de uma base de dados apropriada, o sistema deve permitir a aquisição dos registos para esta mesma base de dados. A recolha de informação é feita registo a registo, com selecção prévia do tipo de registo e técnica de recolha desejada.

Finalizado o processo de recolha de dados, a base de dados resultante poderá ser fundida com a BDC original, o que deverá ser feito por um módulo externo ao nosso sistema (2.2), que correrá no sistema central.

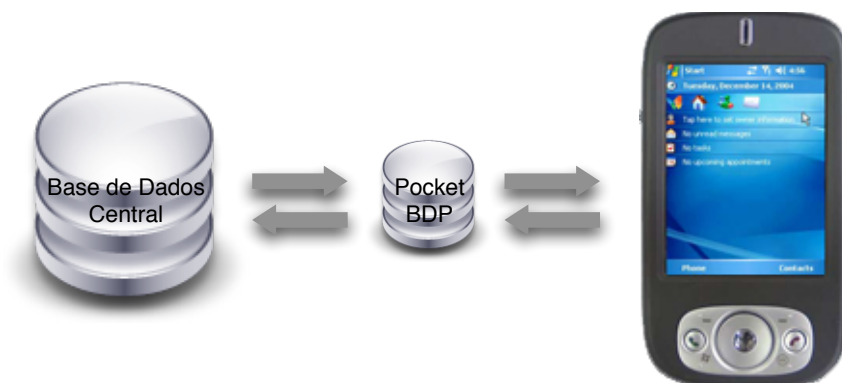


Figura 3.1: Arquitectura global do sistema a desenvolver

Desta forma, além dos diversos problemas tecnológicos, o maior desafio será a adaptação do processo de recolha de informação a um ambiente móvel, caracterizado por baixos recursos de processamento e armazenamento, tal como um espaço visual muito reduzido.

Toda a ideia desenvolvida considera a não existência de uma ligação exterior à internet, assumindo que os locais tradicionais de recolha de dados demográficos

não permitem tal capacidade, pelo menos a um custo reduzido.

3.1 Mobilidade da Base de Dados

O SRP contém duas diferentes visões de armazenamento de informação:

Base de Dados Paroquiais (BDP) são bases de dados contendo a informação de uma paróquia que, neste caso, se pretendem móveis;

Base de Dados Central (BDC) é uma base de dados fixa, localizada, com a mesma estrutura das BDP's, que agrupa a informação de todas as paróquias, ou seja, toda a informação relativa a famílias e indivíduos. As BDPs são posteriormente convertidas e fundidas com a BDC.

A solução ideal seria transportar a BDC para o “campo de trabalho”, pois seria bastante útil na resolução de conflitos e na consulta de informação de outras paróquias. No entanto, um dispositivo móvel dispõe de recursos limitados, principalmente em termos de memória e de processamento, pelo que seria impensável transportar a BDC na sua totalidade, dados os seus milhares de registos, e mesmo que fosse possível armazenar a BDC no dispositivo, o tempo de processamento de operações sobre esta tornaria a aplicação não utilizável.

Uma outra abordagem seria aceder à BDC através de um serviço *online*, não guardando qualquer informação no PDA. Isto solucionaria os problemas relacionados com o poder computacional do dispositivo, pois este seria transferido para o servidor do serviço. Mas, como discutido anteriormente, o acesso à rede nestes locais não é normalmente viável por motivos financeiros ou pela inexistência de pontos de acesso. Logo, esta solução, só por si, não é viável.

A solução encontrada foi, então, um misto das duas anteriores: permitimos ao utilizador levar uma vista da BDC para o terreno, a qual baptizamos de Pocket BDP. Uma vista de uma base de dados é um subconjunto desta, com a mesma estrutura, mas que contém apenas uma pequena parte da informação. Neste caso, é permitido ao utilizador escolher os parâmetros da vista, como por exemplo, de que paróquias é a informação que pretende descarregar, a que altura temporal pertence, entre outros. Seria interessante neste ponto melhorar o mecanismo de selecção de informação, introduzindo técnicas inteligentes de filtragem de dados.

O utilizador terá igualmente a possibilidade de utilizar um serviço *online* para obter novas informações, acrescentando-as à Pocket BDP actual. No entanto embora análogo à solução estudada inicialmente, a existência de uma ligação à rede não é crucial neste sistema, mas complementa o funcionamento do mesmo.

Após a recolha dos dados, a Pocket BDP terá de ser sincronizada com a BDC, utilizando para tal as estruturas já existentes sobre BDPs, o que torna necessária uma conversão de Pocket BDP \rightarrow BDP.

A Pocket BDP será uma base de dados portátil, leve, pequena e suficiente rápida para poder ser usada num ambiente embebido. A escolha da tecnologia utilizada constitui um factor chave na definição das suas propriedades, mas será discutida apenas no capítulo seguinte, de implementação.

3.2 Aquisição de Dados

O sistema de aquisição de dados permitirá duas abordagens distintas no processo de recolha de dados:

baseada no preenchimento de formulários, em que a aplicação criada consiste no preenchimento de campos da BDP definidos à imagem do preenchimento dos formulários usados nas aplicações clássicas para carregamento de BD's e, portanto, semelhantes às aplicadas na demografia;

baseada na transcrição completa dos registos em formato de texto anotado, com um sistema de marcação que permite extrair a informação relevante para a base de dados.

Em cada uma das abordagens, cada documento deverá permitir ter uma ou mais imagens a ele associadas, que corresponderão a cópias totais ou parciais do documento original.

3.2.1 Recolha de Dados Baseada no Preenchimento de Formulários

O processo tradicional de recolha de informação para uma base de dados baseia-se na criação de formulários estruturados à imagem dos métodos demográficos tradicionais de preenchimento de formulários em papel, tal como o MRP. Este revela-se o mais objectivo e, geralmente, o mais intuitivo para os utilizadores.

Em termos de computação, este processo é igualmente simples, uma vez que permite o mapeamento quase directo da informação associada à aplicação para a base de dados.

Cada registo possui muita e extensa informação, o que pode comprometer a usabilidade de uma aplicação implementada segundo este conceito. No sentido de minimizar esse efeito, o processo é dependente do tipo de registo. Cada tipo de registo está associado a um evento único na vida de um indivíduo: baptizado, casamento e óbito, pelo que são evidentes as diferenças entre o tipo de informação específica do evento (o casamento de um indivíduo pode não ser único). O sistema de janelas deverá reflectir esta característica, adaptando o seu esquema ao tipo de registo actual. Devido às diversas limitações de um PDA, é também necessário ajustar a informação dos diversos eventos de maneira a minimizar estas mesmas diferenças, assumindo o objectivo de tornar o sistema de janelas menos pesado.

Para campos nos quais existe uma lista finita e predefinida de opções, o sistema de janelas apenas tem que permitir ao utilizador escolher uma opção. No entanto, para campos sem valores fixos, como por exemplo o nome de um indivíduo, interessa fornecer ajudas contextuais tais como verificar se o novo valor do campo existe na base de dados e listar valores semelhantes existentes. O utilizador poderá, então, escolher entre definir um novo valor (são permitidos valores repetidos, por exemplo, dois indivíduos podem ter o mesmo nome) ou seleccionar um valor previamente existente.

Um exemplo da importância desta funcionalidade é o facto de, ao referenciarmos o mesmo indivíduo em dois eventos diferentes (digamos, o nome dos noivos num casamento e pais no baptizado de um filho), não queremos criar dois indivíduos distintos, mas sim referenciar o mesmo indivíduo nos dois casos.

3.2.2 Transcrição Completa de Registos com Anotações

A recolha de dados baseada no preenchimento de formulários resolve por defeito muitas das ambiguidades na conversão da informação de documentos para uma base de dados relacional. No entanto, a especificidade que a caracteriza previne este tipo de abordagem de ter conhecimento da estrutura do documento, uma vez que cabe ao utilizador filtrar a informação relevante.

Se olharmos para o contexto dos registos demográficos, estes são geralmente documentos antigos, de difícil acesso, deteriorados e com caligrafia e linguagem diferentes, o que dificulta a sua leitura. Por esse motivo, é comum a existência de projectos de transcrição completa de arquivos históricos para formato digital.

Nesta secção, iremos seguir uma abordagem semelhante de transcrição completa dos registos paroquiais, mas suportando anotações para manter o mesmo nível de especificidade que o método tradicional. Em tal método, o utilizador deverá continuar a assumir total poder de decisão sobre a criação das anotações, o que nos permite definir uma estrutura não sequencial, independente da ordem dos campos. Desta forma, o produto final para cada registo será uma transcrição completa desse mesmo registo com as anotações definidas pelo utilizador, em que o texto do documento e as anotações são completamente distinguíveis e separáveis.

Ao mesmo tempo, a transcrição completa dos registos abre portas para a futura reutilização dos mesmos nos mais diversos estudos em outras áreas, tais como a linguística e a grafologia.

Aos dois dias do mez de janeiro do anno de mil oitocentos e oitenta e dois, n' esta Igreja paroquial Matriz da Santissima Trindade, conselho da Villa das Lages do Pico, Diocese de Angra, baptisei solenemente um individuo do sexo masculino, a quem dei o nome de Joao, que nasceu n' esta freguesia as dez horas da noite de desoito do mez de novembro do ano anterior de mil oitocentos oitenta e um, filho legitimo de Manuel Francisco Alvez, agricultor, e de Maria Luiza, de profissao domestica, naturais d' esta freguesia, recebidos na mesma e paroquianos d' ella, moradores do Soldao, lugar d' esta freguesia, neto paterno de Manuel Francisco e de Maria de Jesus, e materno de Jose Leal Camanho e de Maria Luiza. Foi padrinho Manuel Silveira de Brum, casado, agricultor, e madrinha Carolina de Sao Jose. solteira, de profissao domestica; os quais todos sei serem os proprios, E para constar lavrei em duplicado este apento, que depois de ser lido e conferido perante os padrinhos, comigo o nao apignaram por nao saberem escrever. Era ut supra. O vigario Antonio Ribeiro Homem da Costa

Tabela 3.1: Registo de baptismo

Na perspectiva do utilizador, quanto mais informação por registo for necessária, mais monótona será a recolha dos dados, uma vez que será um procedimento várias vezes repetido, dado o elevado número de registos.

Sendo claro que a transcrição completa de um registo será um processo extremamente dispendioso, agravado pelo facto de a escrita do texto ser feita num PDA, a sua monotonicidade deve ser minimizada a todo o custo, através da criação de um interface intuitivo, compacto, extremamente visual e, acima de tudo, com funcionalidades e uma estrutura que evitem a necessidade de informação repetida ou redundante. Note-se que é impossível evitar repetições no texto dos documentos em si, uma vez que o objectivo desta solução é a sua transcrição

na totalidade. No entanto, as técnicas de aprendizagem e reconhecimento de linguagem do utilizador incorporadas no teclado do dispositivo prometem uma melhor experiência na escrita de texto.

A solução escolhida foi a criação de um editor de texto com suporte para a criação/remoção de anotações. Este editor deverá utilizar as potencialidades visuais das formatações de texto para identificar as anotações no texto de um registo, de forma a aproveitar o pouco espaço visual desta classe de dispositivos. Diferentes tipos de anotações deverão estar associados a diferentes formatações de texto.

Como referido na secção anterior, existem três tipos de registos demográficos, um para cada tipo de evento, com estruturas semelhantes mas diferenciadas.

Tal como no sistema de janelas criado, o editor deve ser sensível ao tipo de registo em causa, gerando diferentes contextos de anotações possíveis. A existência de contextos de anotações é um factor importante para a criação de um ambiente mais intuitivo.

Na definição de anotações em que o valor pertença a uma lista finita de opções possíveis, faz sentido incluir um sistema de inferência que tente reconhecer um valor possível para o campo no texto da anotação. Esta funcionalidade engloba aplicar todos os tipos de variações expectáveis sobre a palavra que descreve cada valor (por exemplo, suportar o valor com ou sem acentos, com ou sem maiúsculas) e abreviações conhecidas. Um conceito a explorar é a utilidade de permitir ao utilizador definir e associar novas abreviaturas a valores de anotações possíveis.

Na secção anterior, já referimos importância da existência de ajudas contextuais aquando do preenchimento de certos campos, pelo que desejaríamos aplicar o mesmo conceito a esta abordagem. Relembrando o exemplo da definição do nome de um indivíduo, ao definir esta anotação, a solução mais correcta será criar um menu de contexto com uma listagem de nomes semelhantes na base de dados. No entanto, neste caso o tempo que esta operação demora é crucial, mais ainda que na abordagem anterior, uma vez que esta deverá ser uma funcionalidade de edição do texto do registo em tempo real.

Embora, segundo esta abordagem, o mapeamento de um registo transcrito para a base de dados seja um processo automatizado e silencioso para o utilizador, a possibilidade de conflitos entre os dados do registo em causa e da base de dados é frequente e, ainda que tente orientar o utilizador nesse sentido, o sistema de transcrição não garante que todas as anotações possíveis sejam definidas, por erro do utilizador.

Num cenário de integridade máxima, faz sentido dotar o utilizador da possibilidade de mapear a informação associada às anotações do registo transcrito para o sistema de janelas baseado no preenchimento de formulários, descrito na secção anterior, em vez de a tentar mapear directamente para a base de dados.

Desta forma, o utilizador poderá preencher campos não definidos no registo transcrito e usufruir das vantagens da interface tradicional, tais como: verificar situações de conflito entre a base de dados local do dispositivo e o novo registo a inserir; e garantir que a informação está a ser mapeada para a base de dados tal como desejado, uma vez que este mapeamento é mais directo e menos sujeito a erros.

No processo normal, o sistema deverá apenas validar o registo transcrito segundo a estrutura definida para o tipo de registo em causa e tentar mapeá-lo para a base de dados local de seguida.

3.3 Navegação pela Informação dos Indivíduos e Famílias

Além de mostrar os valores existentes para um dado campo, o sistema deverá permitir ao utilizador navegar na informação dos indivíduos e família a partir desses mesmos valores, através de um sistema de navegação intrinsecamente associado ao sistema de recolha de dados. Para os campos que se revelem relevantes, tais como o nome de um indivíduo, os formulários de recolha deverão permitir comparar o valor actual de um campo com os valores existentes na base de dados para esse mesmo campo e, caso o utilizador o escolha, permitir uma navegação pela informação dos indivíduos e famílias iniciada num dos valores existentes.

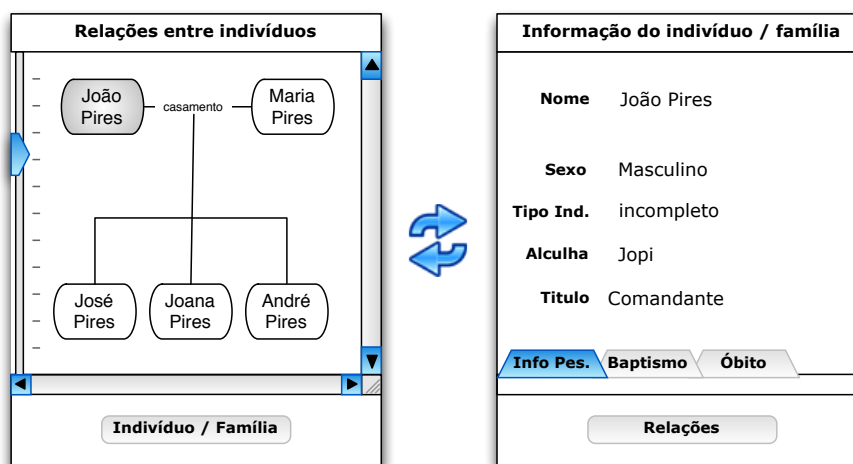


Figura 3.2: Esquema possível para o módulo de navegação pela informação dos indivíduos e famílias

Este sistema de navegação será responsável por permitir ao utilizador ver a informação existente relativa a esse indivíduo e alterar parte dela caso desejado, ou mesmo procurar por familiares desse indivíduo, juntando, em teoria, todos os campos da base de dados numa rede de relações entre sujeitos e, portanto, orientado ao indivíduo e à família.

A visualização da informação deverá ser feita através de um rede de relações, que possua a capacidade de seleccionar indivíduos e relações de casamento. Cada nodo da rede deverá representar um indivíduo diferente, cuja informação pode ser consultada por selecção do próprio. O sistema deve otimizar a representação da rede conforme o indivíduo seleccionado, de modo a privilegiar o grafo de relações desse indivíduo.

Ao nível da implementação, esta funcionalidade pode ser implementada como um módulo completamente independente que permita apenas navegar e alterar valores na base de dados. No entanto, este módulo deverá ser invocado nativamente no módulo de recolha de dados.

Capítulo 4

Implementação

A implementação do projecto procura exprimir as ideias propostas no capítulo anterior. As ideias descritas constroem um sistema composto por quatro aplicações ou módulos. Estes interagem com os módulos já existentes (4.3) e entre si (4.1).

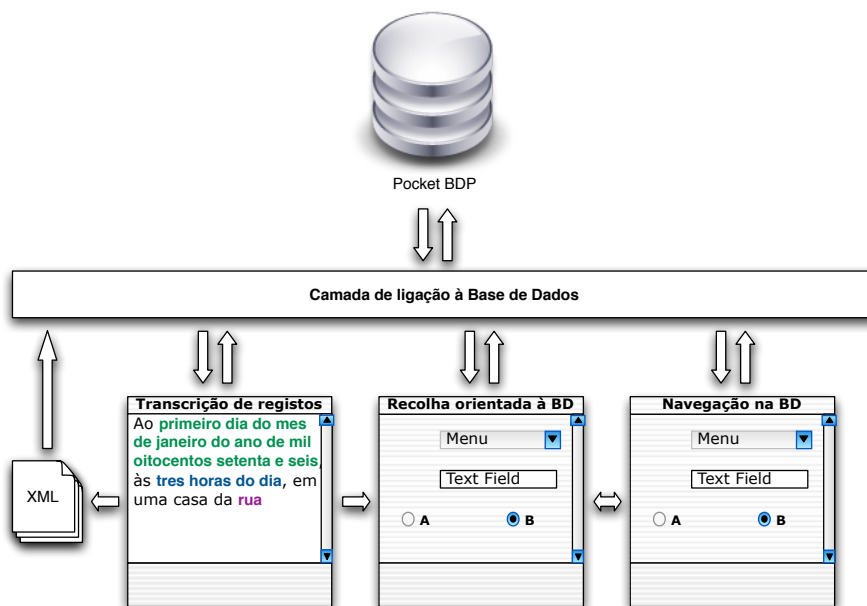


Figura 4.1: Estrutura e interação entre os módulos desenvolvidos para o PDA

O primeiro módulo constitui uma ferramenta de especialização e conversão das bases de dados usadas, que obtém uma vista da BDC e a converte para o formato Pocket BDP.

Foram definidos dois módulos independentes para a aquisição de dados, segundo perspectivas diferentes, mas suportados por uma camada comum que lhes permite interagir de forma estruturada com a base de dados.

Existe ainda um módulo de navegação pela informação dos indivíduos e

famílias, local ao dispositivo, e que permite explorar e editar as informações contidas nesta.

Embora todos os módulos possam ser gerados de forma independente, será também criada uma interface para o utilizador, com um sistema de menus que o oriente na selecção de uma abordagem a utilizar e invoque nativamente os módulos adequados.

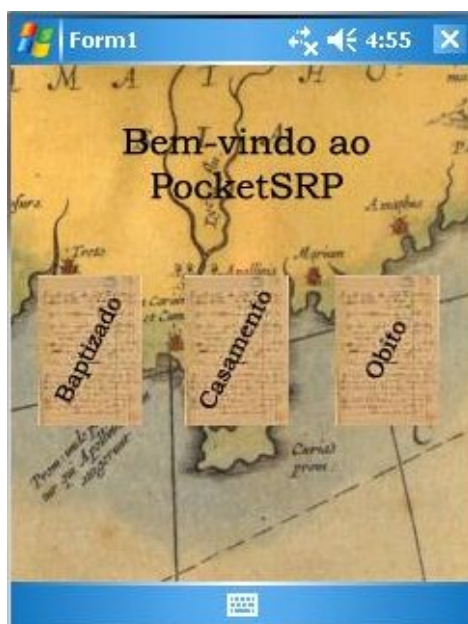


Figura 4.2: Menu inicial da aplicação de recolha de informação

Inicialmente, iremos descrever as tecnologias utilizadas e explicar que razões as tornaram as mais apropriadas para expressar a ideia a desenvolver.

Na descrição de cada módulo apresentam-se as decisões tomadas e restrições/problemas encontrados na implementação ou impostos pelas técnicas e tecnologias utilizadas.

4.1 Tecnologias

Nesta secção, preocupamo-nos com as tecnologias a utilizar na implementação da ideia proposta. Algumas das tecnologias utilizadas nos módulos existentes não possuem suporte na plataforma Pocket PC, e a utilização de tecnologias adequadas exige a criação de módulos de conversão.

Nem todas as tecnologias enunciadas foram utilizadas, quer por dificuldades na sua utilização, quer por falta de suporte adequado na plataforma Pocket PC.

4.1.1 Base de Dados

Para o desenvolvimento do projecto, analisamos várias soluções para o motor de base de dados, que deve servir não só de suporte à aplicação móvel, mas

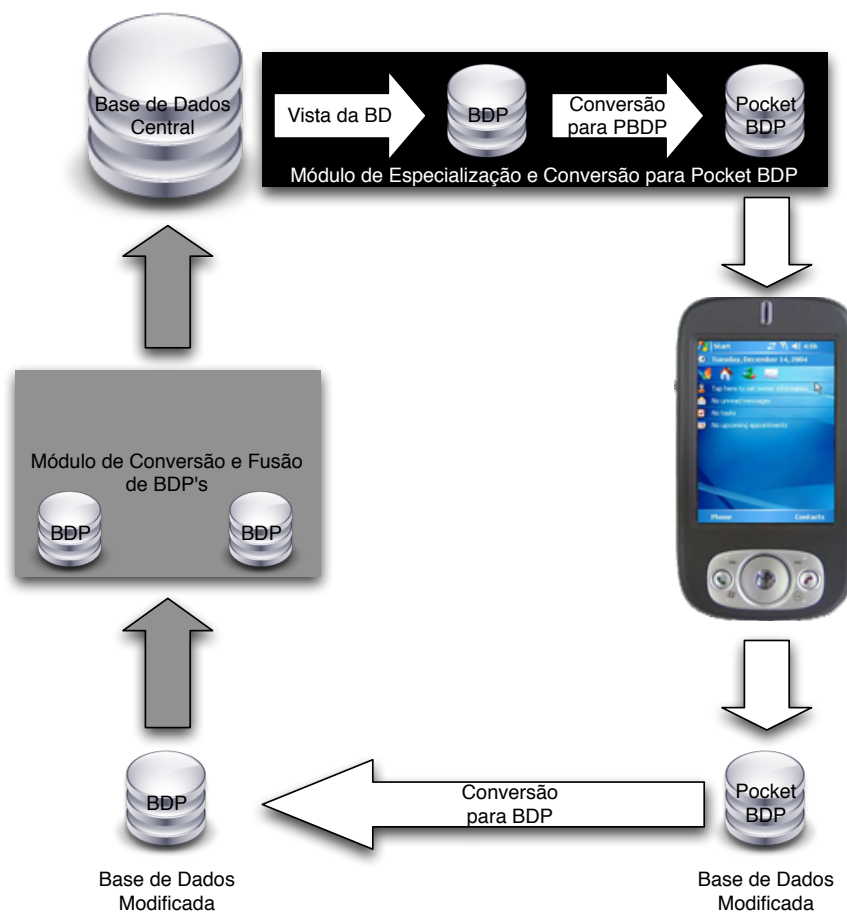


Figura 4.3: Arquitectura do sistema geral e interação com os outros módulos do SEED

também possuir grande compatibilidade com as bases de dados já existentes em formato Access/JET, pontos de partida e chegada obrigatórios.

Das soluções analisadas apresentamos, as três que foram tomadas como possíveis, ou mesmo em parte utilizadas durante uma certa fase de desenvolvimento. Todas elas têm como vantagem o facto de serem “bases de dados de ficheiro”, ou seja, não se basearem no modelo cliente-servidor, não é apropriado para este caso, dado que a aplicação necessita de ser móvel e não estar dependente de um serviço registado numa determinada máquina.

De seguida, apresenta-se as três opções estudadas, tendo a nossa escolha recaído sobre o sistema SQLite e desenvolver em conjunto um módulo de especialização e conversão de Access para SQLite (4.3).

Microsoft Access / Microsoft JET Database Engine

O Microsoft Office Access é um sistema relacional de administração de base de dados, que combina o Microsoft Jet Database Engine com uma interface para

o utilizador. Uma vez que os módulos existentes foram desenvolvidos utilizando este sistema, a nossa intenção inicial era a de utilizar esta tecnologia, de maneira a evitar uma conversão desnecessária dos dados já existentes. No entanto, verificamos que esta não seria de facto a melhor solução, pois o Access contém inúmeras desvantagens e defeitos (como, por exemplo, não ser facilmente escalável[15]), mas sobretudo pela falta de suporte para Access na plataforma Compact .Net Framework, na qual foi descontinuado pela própria Microsoft[4]

SQL Server Compact Edition

Dado que o Access não era apropriado para a nossa aplicação, viramos a nossa atenção para o SQL Server Compact Edition (também conhecido por SQL Server Mobile ou SQL Server Everywhere Edition). Por estar divulgado por parte da Microsoft o seu bom funcionamento e integração com a plataforma Compact .Net Framework[5], e também a facilidade de desenvolver as aplicações de conversão (Access ↔ SQL Server Compact), esta foi a nossa escolha inicial. Mas com a sua aplicação, o SQL Server CE revelou-se rapidamente um motor de base de dados demasiado lento nas queries de procura que nós necessitávamos. Em adição, demonstrou também uma má representação interna dos tipos de dados, gastando demasiado espaço em disco, algo não aceitável numa plataforma limitada.

Logo, por todos estes motivos, partimos para outro sistema de base de dados, o SQLite.

SQLite

Depois dos problemas demonstrados pelos dois sistemas anteriores, optamos por SQLite, mais propriamente a System.Data.SQLite que é uma implementação completa em ADO.NET 2.0 do SQLite, que suporta tanto a plataforma .Net Framework como a Compact .Net Framework[11]. O SQLite é um sistema de gestão de bases de dados relacionais embutido numa biblioteca[9], que enuncia e demonstra as seguintes vantagens:

- implementa a maior parte do standard sql92,
- a totalidade da base de dados é guardada num ficheiro único multiplataforma,
- é open-source e totalmente livre,
- revela-se bastante rápido a realizar as queries.

Por estas razões, e principalmente devido à melhor velocidade apresentada, optamos pela utilização do SQLite.

4.1.2 GraphViz - WebDot

A biblioteca GraphViz permite gerar, fácil e rapidamente, grafos a partir de uma linguagem descritiva própria. A sua extensão WebDot permite associar acções à selecção de nodos ou ligações, motivo que tornam esta ferramenta uma boa solução para a visualização dos grafos de relações entre sujeitos no módulo de navegação pela base de dados.

No entanto, a inexistência de uma distribuição própria para a plataforma Pocket PC e a filosofia de servidor web necessária pela extensão WebDot inviabilizam a sua utilização no contexto deste projecto.

4.1.3 GPS

Embora, inicialmente, um dos objectivos deste projecto fosse permitir utilizar o hardware GPS do dispositivo para adquirir coordenadas geográficas e desenvolver funcionalidades baseadas nessa informação, nenhuma funcionalidade relevante foi descoberta.

Por este motivo, além da procura inicial de soluções viáveis à utilização desta tecnologia, nenhum esforço extraordinário foi despendido na exploração das potencialidades e aplicações do *Global Positioning System*.

4.1.4 Reconhecimento de Voz

Dado o tamanho consideravelmente grande de um registo demográfico, o reconhecimento de voz foi pensado como um possível meio de agilizar e simplificar o processo de transcrição do registo.

No entanto, a solução oferecida pelo *Microsoft Speech Server* implica que a aplicação seja desenvolvida como uma página web à qual o utilizador se liga, o que não se revela viável dada a impossibilidade de utilizar serviços web, uma vez que no cenário de aplicação do projecto a conexão à internet é inexistente.

4.1.5 XML

A XML (Extensible Markup Language) ([1]) é uma linguagem de anotações de estrutura hierárquica, logo, baseada numa árvore. É apresentada pelo *World Wide Web Consortium* como o formato universal para dados e documentos estruturados e assume um papel cada vez mais importante na difusão destes na Web e em todos os meios ([2]). Ao contrário de outros sistemas de anotação, a norma XML não estabelece um conjunto de anotações fixo para um caso específico ou para todos os casos; é uma metalinguagem que indica como definir as anotações próprias de cada família de documentos.

Na sua raiz, toda a informação é vista como texto, intercalada com anotações que definem elementos com uma posição implícita na hierarquia e atributos inerentes a esses elementos. Por esta razão, é a linguagem apropriada para estruturar e armazenar os valores resultantes da transcrição dos registos com anotações, em que as anotações são definidas através do mapeamento de cada campo da recolha de dados, baseada no preenchimento de formulários, para uma associação de um elemento com o texto pertencente ao campo em causa e atributos necessários.

Neste projecto, a utilização da XML não visa de forma alguma substituir a conversão de um documento para a base de dados, mas é vista como um processo intermédio de enriquecimento do texto original dos registos através da criação de anotações. Os registos anotados, enquanto versão digital dos registos transcritos, deverão possuir uma estrutura que suporte a sua conversão para a base de dados.

XML Schema

No entanto, embora permita definir tipos de dados como a composição de tipos de dados mais básicos, a XML não possui um meio de definir a estrutura, conteúdo e semântica de um documento, o que é conseguido na especificação XML Schema[13]. Uma instância da XML Schema define um conjunto de regras a que um documento XML deve obedecer de forma a ser considerado válido para esse esquema[14].

Neste caso específico, a XML Schema é importante na definição da estrutura de um registo paroquial em termos de tipos de dados e relações entre elementos e atributos. Esta estruturação de um registo é fundamental para a aplicação de mecanismos de validação do mesmo, aquando da sua conversão para a base de dados ou do seu armazenamento ou leitura.

4.1.6 Plataforma .Net

A .Net framework é um conjunto de ferramentas, bibliotecas e linguagens criados ou adaptados pela Microsoft para suporte ao desenvolvimento de aplicações para as diversas arquitecturas que suportem todo o tipo de versões do Microsoft Windows. Foi desenvolvida com o intuito de integrar numa só ferramenta as múltiplas linguagens de desenvolvimento para a plataforma Windows.

Sendo a aposta da Microsoft para fazer frente à grande popularidade da máquina virtual Java, possui igualmente uma máquina virtual independente da linguagem .Net utilizada, existente em qualquer plataforma Windows, que executa byte-code .Net proveniente de uma ou mais linguagens .Net.

C#

O C# é uma linguagem de programação orientada a objectos criada pela Microsoft e que faz parte da sua plataforma .Net. Ela é baseada na linguagem C++ e Java. Escolhemos esta linguagem para implementação do nosso projecto visto ser a linguagem orientada aos objectos mais em voga neste momento da plataforma .Net. É uma linguagem ainda recente, que ainda está em crescimento e que é a aposta forte da Microsoft no que diz respeito a este paradigma.

Compact .Net Framework

O Compact .Net Framework é uma versão do .Net Framework, que foi desenhada para funcionar em dispositivos móveis. O Compact .Net Framework contém um subconjunto das bibliotecas do .Net Framework completo, bem como algumas bibliotecas desenhadas especificamente, para dispositivos móveis.

Este desenho do Compact .Net Framework, com uma API bastante limitada, trouxesse nos bastantes dificuldades, na implementação do projecto.

Visual Studio

O Visual Studio é um ambiente de desenvolvimento produzido pela Microsoft, vocacionado para desenvolver aplicações para a plataforma .NET, que contém um desenhador de interfaces gráficas bastante potente e prático. Este IDE possui múltiplas funcionalidades de integração com dispositivos móveis, de modo a tornar a programação para estes dispositivos mais simples e intuitiva.

4.2 Módulo de Especialização e Conversão para Pocket BDP

Este módulo procura ser uma aplicação que irá permitir a conversão entre os vários tipos de formato de bases de dados. Mais propriamente, ele divide-se em duas componentes:

- especialização e conversão da BDC para o formato Pocket BDP;
- conversão do formato Pocket BDP para o formato BDP.

Estas duas funcionalidades encontram-se combinadas no mesmo interface para o utilizador, o qual possui a opção de alternar a operação a concretizar.

A primeira componente permitirá ao utilizador escolher a BDC de origem, e criar uma Pocket BDP de destino, que posteriormente será enviada ao PDA, para ser usada nos outros módulos desenvolvidos. Para além desta funcionalidade básica, esta componente fornece também, ao utilizador, uma série de formulários que lhe permitem escolher as opções de especialização para a criação da vista da BDC. Na implementação actual, é permitido ao utilizador seleccionar os dados segundo a paróquia e o intervalo de tempo desejados, mas prevê-se a inclusão de mais técnicas de especialização da informação.

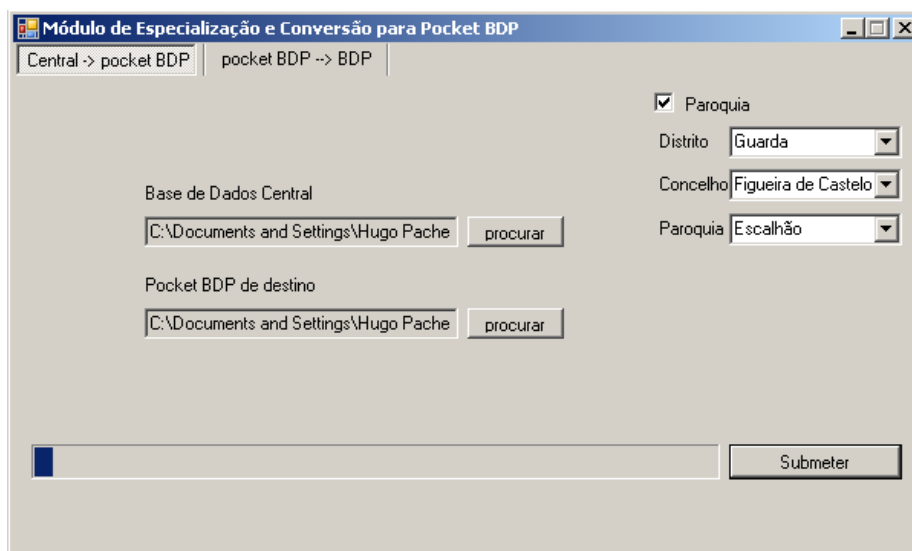


Figura 4.4: Exemplo da conversão de uma BDC para uma Pocket BDP

Ainda nesta componente, é também permitido receber uma BDP, em vez da BDC, sendo nesse caso apenas aconselhado fazer uma conversão directa e não uma vista, devido ao carácter já especializado da BDP.

A segunda componente permite converter a Pocket BDP, resultante da recolha de dados no dispositivo móvel para uma BDP, que, por sua vez, irá ser integrada na BDC. Esta componente, além de permitir o mapeamento entre as bases de dados, acrescenta as tabelas de registos anotados e de imagens digitalizadas à BDP, um assunto que discutiremos mais a fundo na secção seguinte.

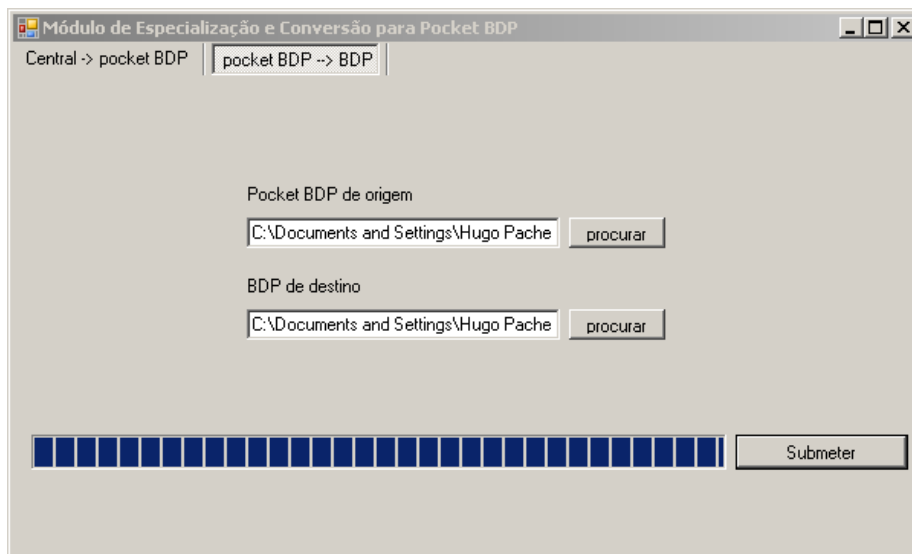


Figura 4.5: Exemplo da conversão de uma Pocket BDP para uma BDP

4.3 Camada de Base de Dados

Esta camada foi desenvolvida com o intuito de proporcionar uma biblioteca que permita um acesso à base de dados transparente por parte dos dois módulos de recolha de dados e do módulo de navegação. Esta deve ser capaz de esconder a complexidade dos acessos à base de dados e permitir, simultaneamente, um mapeamento de um modelo relacional para um modelo por objectos.

A conversão assumida é um compromisso entre os dois paradigmas (relacional e OO) de modo a: manter características relacionais no modelo por objectos, compatibilidade que permite carregar apenas porções de informação necessárias; mas suficientemente abstracto para garantir que não é preciso ter conhecimentos sobre pormenores de implementação da base de dados.

O modelo de base de dados assume grande importância neste processo de conversão. Por motivos previamente descritos, o modelo utilizado foi o SQLite.

A estrutura da Pocket BDP é em tudo semelhante à da BDP, tendo sido apenas mapeados alguns tipos de dados incompatíveis entre Access e SQLite e acrescentadas duas tabelas das quais falaremos mais tarde. Não vamos explicitar a fundo a estrutura da BD, uma vez que a documentação do módulo SRP possui informação detalhada sobre este assunto[17][18].

Uma das dificuldades encontradas foi a forma como a base de dados estava estruturada, e o facto de estar originalmente em formato Access, sendo os principais problemas::

- Inexistência de chaves estrangeiras no desenho da base de dados original, existindo casos em que é impossível a sua introdução, tais como um atributo referir dois atributos em diferentes tabelas.
- Indexação de tabelas por campos de texto quando, na realidade, não existe necessidade para tal, podendo ser substituído por um índice numérico.

- Em alguns (poucos) casos, existem campos redundantes que são irrelevantes para a unicidade da chave primária. Isto deve-se, provavelmente, ao facto de a base de dados ter sido desenvolvida ao longo de um grande período de tempo.
- Agrupamento, numa só tabela “InfoComplementar”, de dados descritivos dos códigos de várias tabelas distintas.
- Todos os valores booleanos estão marcados como não nulos, esta é uma funcionalidade do Access.

Por não querermos mudar radicalmente a estrutura da base de dados, por uma questão de compatibilidade, e uma vez que não é do âmbito do nosso trabalho, optamos por a manter, apesar dos problemas acima descritos. Voltaremos a este tema na secção dedicada ao trabalho futuro (6.1).

Devido à nova abordagem de obter e guardar os documentos originais, tanto por digitalização dos registos originais, como por transcrição integral dos documentos em XML, foi necessário criar duas novas tabelas para acomodar estes ficheiros. Foram então criadas as tabelas Imagens e RegistosXML. O ficheiro digitalizado é guardado dentro da base de dados, no formato blob, pois é irrealista guardar apenas a localização do ficheiro, dado que a base de dados é portátil e muda constantemente de plataforma. Os documentos XML são guardados como texto simples.

A biblioteca consiste então em três componentes fundamentais:

- Uma responsável por fazer a ponte entre a hierarquia de classes e a base de dados bem como providenciar aos outros módulos os métodos necessários ao seu funcionamento (inserção de um novo indivíduo, por exemplo).
- A outra componente, a hierarquia de classes propriamente dita, contém classes fundamentais, como Individuo e Família, e de apoio, como Assistente, Padre, EstadoCivil, etc.

A hierarquia de classes foi desenvolvida à imagem da base de dados, embora algumas das características do paradigma OO pretendidas exigam um processo de transformação. Uma das transformações efectuadas foi a “desindexação”, isto é, uma tabela A que contém uma lista de elementos dependentes de um dado atributo de uma tabela B é transformada numa classe B que contém uma lista de elementos de A . Por exemplo a tabela “Profissões” vai aparecer como uma lista de Profissões na classe Individuo na hierarquia de classes.

4.4 Módulo de Recolha de Dados Baseada no Preenchimento de Formulários

O design visual constitui o maior problema desta implementação, pois os registos possuem informação muito extensa, o que dificulta a sua representação num dispositivo móvel com pequenas dimensões de ecrã. Para combater este problema, e para aumentar a usabilidade do programa, este sistema foi orientado aos três principais eventos (dos quais existem registos) da vida de um indivíduo. Assim sendo, o utilizador terá a hipótese de seleccionar o tipo de registo desejado

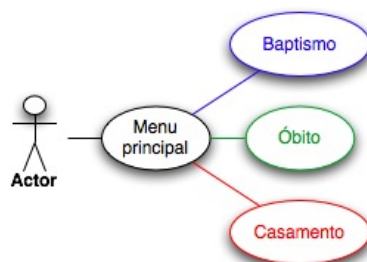


Figura 4.6: Interação entre os formulários do programa: Janela Principal

através de um menu, Ao escolher uma destas opções, a informação visual de recolha de dados apresentada no ecrã será bastante menor, visto que se consegue eliminar toda a informação referente aos dois tipos de eventos não escolhidos.

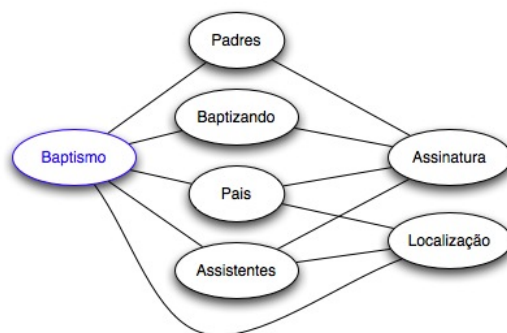


Figura 4.7: Interação entre os formulários do programa: Baptismo

Devido à falta de recursos dos dispositivos móveis, este sistema foi elaborado de forma a reutilizar janelas que recolham o mesmo tipo de informação nos diversos eventos. Com isto, uma janela que contenha um conjunto de informação que esteja presente em mais do que um evento será repetida para esses mesmos eventos, conseguindo-se assim minimizar os gastos de memória.

A informação inserida por um utilizador será passada para a base de dados apenas quando este der como concluída a sua tarefa, de maneira a que o acesso à base de dados ocorra apenas uma vez, no final da inserção de informação. É óbvio que em determinados casos como, por exemplo, a criação de uma nova profissão ou localidade, é necessária a inserção imediata na base de dados, para o caso de serem necessárias novamente antes da conclusão do processo global.

Para suporte e auxílio do historiador quanto à escolha de um indivíduo, é-lhe apresentado o conjunto de indivíduos cujo nome se identifica com a pesquisa do historiador. Assim, o utilizador poderá resolver alguns conflitos prontamente aquando da escolha de um indivíduo. Devido às grandes dimensões da base de dados, este processo torna-se um pouco moroso, visto que a criação desta lista de indivíduos e visualização da mesma se torna muito pesada para um dispositivo deste calibre. Por razões amplamente aceitáveis, a visualização da informação

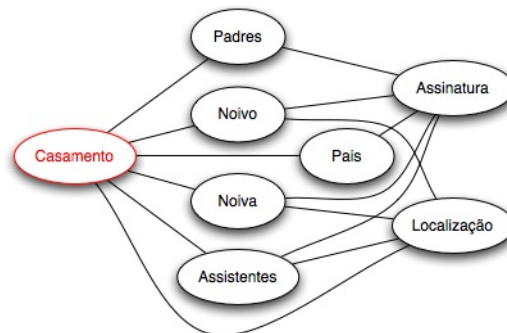


Figura 4.8: Interação entre os formulários do programa: Casamento

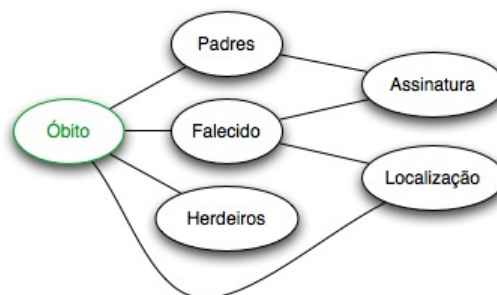


Figura 4.9: Interação entre os formulários do programa: Óbito

consome muito mais tempo do que a execução da query de selecção sobre a base de dados, dados os poucos recursos de memória gráfica do dispositivo.

Segundo o mesmo objectivo, este módulo está igualmente ligado com o Módulo de navegação pela informação dos indivíduos e famílias. Aquando da inserção da informação referente a um evento, poderão ser lançadas várias dúvidas ao historiador, devido à grande criatividade nos nomes existente na altura. O historiador terá sempre a possibilidade de consultar a informação de indivíduos na altura do registo do evento, de forma a esclarecer a sua opção.

4.5 Módulo de Navegação pela Informação dos Indivíduos e Famílias

Idealmente, este módulo deveria consistir em duas representações: um grafo com as relações de um indivíduo e uma janela com a informação de um indivíduo.

Na representação por grafo, deverá estar presente um indivíduo e todas as suas relações familiares. A representação deste mesmo indivíduo deverá mudar consoante o seu estado (falecido, vivo, ...) e consistência da informação referente a este. O grafo também deverá permitir a navegação por todos outros indivíduos com relações de parentesco com este indivíduo. No entanto, a complexidade do grafo produzido não poderá ser excessivamente grande, devido ao

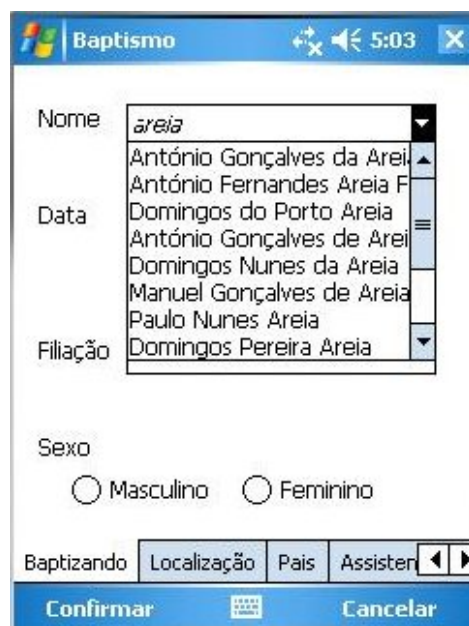


Figura 4.10: Exemplo da pesquisa de indivíduos semelhantes num registo de baptizado

reduzido ecrã que o PDA apresenta.

Esta representação não foi implementada devido à impossibilidade de usar o graphviz num PDA. Seria interessante, como trabalho futuro [6.1](#), a pesquisa de tecnologias de criação de grafos suportadas por um dispositivo com estas características.

Desta forma, a única componente implementada neste módulo foi a representação da informação de um indivíduo através de janelas. Esta representação permite representar toda a informação de um indivíduo, ao contrário da representação por grafo. A partir desta mesma janela, o utilizador pode navegar igualmente pela informação de outros indivíduos relacionados com o primeiro e assim sucessivamente.

Esta representação também permite consultar e alterar a informação de uma família. Aquando da consulta de um indivíduo, podemos automaticamente consultar a família deste e, quando consultamos uma família, podemos automaticamente consultar qualquer membro dessa família. A informação de um indivíduo ou família poderá ser sempre alterada pelo historiador.

Em termos gerais, este módulo acrescenta uma mais valia a este projecto, uma vez que pode ser usado para a consulta e dissolução de dúvidas quanto à identificação de um indivíduo ou família, como também para a correcção de alguma informação ou dado mal inserido ou mal interpretado, além de poder ser reutilizado pelos módulos de aquisição de dados.

The screenshot shows a Windows application window titled "Casamento" with a system tray showing the time as 5:10. The form contains the following elements:

- Data:** A text box containing "22-11-2002 17:08" and a checkbox labeled "Fixa".
- Noivo:** A text box containing "Victor".
- Noiva:** A text box containing "Noiva".
- Casamentos:** A spinner box set to "0".
- Idade:** A spinner box set to "21".
- Estado Civil:** A dropdown menu set to "Solteiro".
- Profissão:** A dropdown menu set to "Agricultor".
- Alcunha:** An empty text box.
- Título:** An empty text box.
- Residência:** A checkbox.
- Assinatura:** A checkbox.
- Navigation:** A row of buttons: "Noivos", "Localização", "Pais", "Assistentes", and left/right arrow buttons.
- Buttons:** "Confirmar" and "Cancelar" buttons at the bottom.

Figura 4.11: Exemplo da recolha de um registo de casamento

4.6 Módulo de Transcrição de Registos para Documentos XML

Um dos principais objectivos deste módulo é ser intuitivo para o utilizador. Desta forma, a interface disponibilizada deverá ser o mais visual possível, isto é, representar o máximo de informação possível através da utilização de botões, tipos de letra e cores diferentes.

Tendo em conta que este módulo inclui um pequeno processador de texto, ele deverá possuir funcionalidades inatas a esta classe de programas, tais como gravar e abrir ficheiros num formato próprio e cortar/copiar/colar selecções de texto. A criação ou remoção de selecções está associada a um menu de contexto com as diferentes anotações existentes para o tipo de registo actual, em que cada anotação utiliza uma representação diferente.

Depois de marcada uma anotação, o utilizador terá que preencher uma lista variável de atributos que constituem a informação inerente a essa anotação. Os atributos representam geralmente uma lista finita de opções, mas podem igualmente conter normalizações do texto da anotação como, por exemplo, datas e idades e nomes de localidades. Essas normalizações são feitas com o recurso a formulários cujos campos garantem a formatação correcta dos dados.

O módulo contém um menu de atributos, que é actualizado com a selecção de parte ou totalidade do texto de uma anotação. Caso mais do que uma anotação seja seleccionada é escolhida a primeira anotação dessa selecção.

Para o processo de migração para a base de dados ser bem sucedido, todos os atributos de cada anotação deverão ser preenchidos. Por essa razão, todas as anotações incompletas são marcadas com o estilo itálico, que é alterado para negrito aquando do preenchimento de todos os atributos necessários.

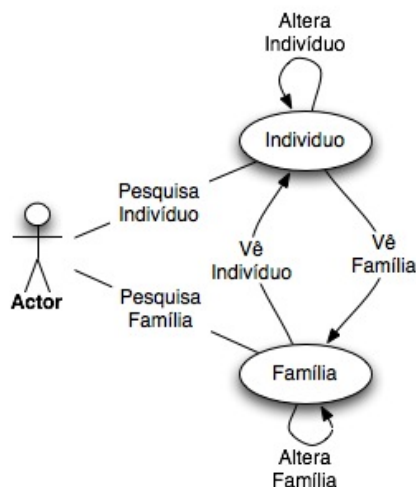


Figura 4.12: Interação possível na navegação pela informação em indivíduos e famílias

Uma outra característica que ajuda a melhorar a interação com o utilizador é a inclusão do maior número de mecanismos de automatização do processo de anotação, de modo a evitar acções redundantes por parte do utilizador, como ter que filtrar opções de atributos inválidas no contexto em que se encontra. Isto envolve a aplicação de métodos de reconhecimento do valor dos atributos de um elemento anotado e a redução do número de atributos possíveis para cada elemento e para cada tipo de registo.

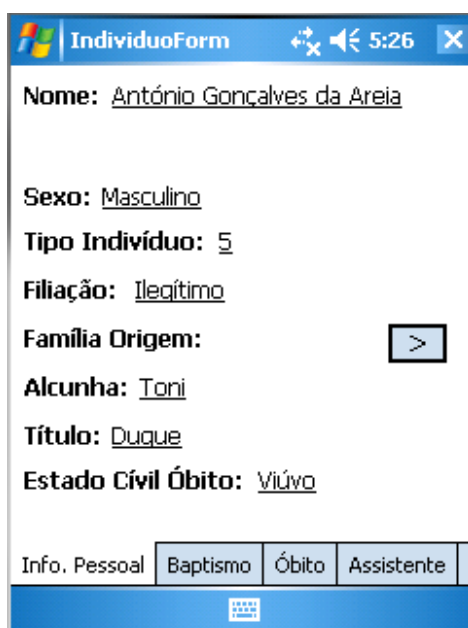
O reconhecimento automático de valores para um atributo é feito com base na associação de expressões regulares a cada valor expectável de modo a suportar variações no texto dessa anotação. Por exemplo, para o atributo *filiação*, são permitidos os valores *legítima*, *ilegítima* e *exposta*. Considerando ambos os sexos do sujeito ou erros de escrita, o valor de cada opção poderia admitir o adjectivo conjugado no masculino, feminino ou sem acentos gramaticais.

A especificação dos valores possíveis para cada anotação está definida em diferentes *schemas* XML, descrita de seguida.

4.6.1 XML Schema

Num projecto anterior sobre a anotação de registos paroquiais segundo o formato XML [20], foi definida uma estrutura geral para os três tipos de registos existentes, sob o princípio que este será mais fácil de aceitar e utilizar pelos utilizadores. Esta abordagem mais permissiva, embora permita definir um sistema genérico de anotação dos documentos, aumenta o leque de possibilidades de anotações e atributos para as mesmas de um registo.

Tendo em conta que um utilizador opta previamente pelo tipo de registo que deseja transcrever, do ponto de vista da automatização do processo de anotação faz todo o sentido definir *schemas* específicos para cada tipo de documento, de forma a restringir o número de opções e forçar o utilizador à estrutura do



IndividuoForm 5:26

Nome: António Gonçalves da Areia

Sexo: Masculino

Tipo Indivíduo: 5

Filiação: Ilegítimo

Família Origem:

Alcunha: Toni

Título: Duque

Estado Civil Óbito: Viúvo

Info. Pessoal | Baptismo | Óbito | Assistente

Figura 4.13: Exemplo da consulta da informação de um indivíduo

documento que pretende.

No entanto, esta estrutura não deve ser demasiado restritiva, considerando que a estrutura de cada registo é passível de sofrer ligeiras alterações consoante a paróquia em causa.

É igualmente importante manter um certo grau de liberdade ao nível da metodologia que cada utilizador poderá usar para a transcrição de um documento. Por exemplo, se uma estrutura para a ordem das anotações fosse definida, um utilizador que pretendesse criar primeiro certo tipo de anotações seria forçado a incluir a sequência de anotações estabelecida no seu método de transcrição.

A estrutura final de cada registo foi criada a partir da especialização do "super-schema" existente para cada registo, apoiada pela estrutura de cada registo ao nível da base de dados, descrita no manual do módulo SRP [17][18], actualmente em utilização pelos historiadores. Uma vez que o *super-schema* do projecto anterior é resultante do estudo sobre registos da paróquia do Pico nos Açores em específico, foram privilegiadas as decisões tomadas no módulo SRP na definição da especificidade de cada registo.

Ao nível da transcrição de documentos, a inclusão de ajudas contextuais relacionadas com a base de dados exige a adição de identificadores específicos da base de dados à estrutura normal dos registos XML. Estes identificadores estão associados à base de dados local e são susceptíveis de mudar com processos de fusão e normalização, pelo que a sua utilização não é a mais adequada.

Consequentemente, não implementando esta funcionalidade é impossível ao utilizador referir indivíduos já existentes, criando sempre novos indivíduos, muito possivelmente clones, no processo de conversão para a base de dados, realçando a importância de um mapeamento dos registos transcritos para os formulários

IndividuoForm 5:27

Nome: António Gonçalves da Areia

Sexo: Masculino Feminino

Tipo Indivíduo: 5

Filiação: Legítimo

Família Originária: Legítimo

Alcunha: Exposto

Título: Duque

Estado Civil Óbito: Viúvo

Info. Pessoal Baptismo Óbito Assistente

Figura 4.14: Exemplo de alteração da informação de um indivíduo durante a consulta desta

de recolha de dados.

Os três diferentes tipos de *schemas* encontram-se descritos de seguida : Note-se que esta representação não possui a declaração de restrições sobre os valores possíveis para cada atributo em diferentes contextos.

De forma a tornar possível a validação de registos incompletos que o utilizador pretenda armazenar temporariamente, foram criados *schemas* específicos que não exigem que os atributos de uma anotação estejam definidos. Consequentemente, os *schemas* criados não são suficientemente restritivos para garantir a integridade do processo de conversão para a base de dados, uma vez que não forçam a existência de nenhum tipo de anotação. No entanto, a especificação dos *schemas* para cada tipo de registo aumenta a probabilidade de sucesso deste processo.

Informação sobre os diferentes tipos de atributos suportados por cada *schema* pode ser consultada no código do próprio *schema* ou no menu de propriedades existentes para cada anotação, ao nível da aplicação. Em anexo encontra-se uma estrutura simplificada dos *schemas* utilizados^A. Todos os *schemas* podem ser encontrados juntamente com a aplicação de transcrição de registos para formato XML.

```

<registoBaptizado>
  <notaMargem><numero>1</numero></notaMargem>
  <texto>
Aos <data deQue="baptismo" norm="1882-01-02">dois dias do mez de janeiro do
ano de mil oitocentos e oitenta e dois</data>, n'esta Igreja parochial
Matriz da Santissima Trindade, <local deQue="baptismo" deQuem="individuo
"><pais>Portugal<distrito>Ilha do Pico<concelho>Lages do Pico<parochia>
Lages do Pico</parochia></concelho></distrito></pais>conselho da Villa
das Lages do Pico, Diocese de Angra</local>, baptisei solenemente um
individuo do <sexo norm="masculino">sexo masculino</sexo>, a quem dei o
nome de <nome deQuem="individuo">Joao</nome>, que nasceu <local deQue="
nascimento" deQuem="individuo"><pais>Portugal<distrito>Ilha do Pico<
concelho>Lages do Pico<parochia>Lages do Pico</parochia></concelho></
distrito></pais>n'esta freguesia</local> as <hora norm="22:00:00">dez
horas da noite</hora> de <data deQue="nascimento" norm="1881-11-18">
dezoito do mez de novembro do ano anterior de mil oitocentos oitenta e
um</data>, <filiacao norm="legitimo">filho legitimo</filiacao> de <nome
deQuem="pai">Manuel Francisco Alvez</nome>, <profissao deQuem="pai">
agricultor</profissao>, e de <nome deQuem="mae">Maria Luiza</nome>, de
profissao<profissao deQuem="mae"> domestica</profissao>, <local deQue="
naturalidade" deQuem="paimae"><pais>Portugal<distrito>Ilha do Pico<
concelho>Lages do Pico<parochia>Lages do Pico<lugar>Soldao</lugar></
parochia></concelho></distrito></pais>naturais d'esta freguesia</local>,
recebidos na mesma e parochianos d'ella, <local deQue="morada" deQuem="
paimae"><pais>Portugal<distrito>Ilha do Pico<concelho>Lages do Pico<
parochia>Lages do Pico</parochia></concelho></distrito></pais>moradores
do Soldao, lugar d'esta freguesia</local>, neto paterno de <nome deQuem
="avoPaterno">Manuel Francisco</nome> e de <nome deQuem="avoPaterna">
Maria de Jesus</nome>, e materno de <nome deQuem="avoMaterno">Jose Leal
Camanho</nome> e de <nome deQuem="avoMaterna">Maria Luiza</nome>. Foi
padrinho <nome deQuem="padrinho">Manuel Silveira de Brum</nome>, <
estadoCivil norm="casado" deQuem="padrinho">casado</estadoCivil>, <
profissao deQuem="padrinho">agricultor</profissao>, e madrinha <nome
deQuem="madrinha">Carolina de Sao Jose</nome>, <estadoCivil norm="
solteiro" deQuem="madrinha">solteira</estadoCivil>, de profissao <
profissao deQuem="madrinha">domestica</profissao>; os quais todos sei
serem os proprios, E para constar lavrei em duplicado este apento, que
depois de ser lido e conferido perante os padrinhos, comigo o nao
apignaram por nao saberem escrever. Era ut supra. O <funcao norm="
vigario" deQuem="autor">vigario</funcao> <nome deQuem="autor">Antonio
Ribeiro Homem da Costa</nome>
  </texto>
</registoBaptizado>

```

Tabela 4.1: Registo de baptismo com anotações

Editor de Registos 5:53

Baptizado nº 0

Aos **dois dias do mez de janeiro do anno de mil oitocentos e dois**, n'esta Igreja paroquial Matriz da Santissima Trindade, conselho da Villa das Lages do Pico, Diocese de Angra, baptisei solenemente um individuo do sexo **masculino**, a quem dei o nome de **Joao**,

Atts

Figura 4.15: Exemplo da anotação de um registo de baptizado

RegionForm 5:07

País Portugal

Zona

Distrito Braga

Concelho Braga

Paróquia

Lugar

Sítio

Gualtar
Guisande
Lamas
Lamações
Lomar
Merelim (São Paio)
Merelim (São Pedro)
Mire de Tibães

Confirmar Cancelar

Figura 4.16: Exemplo do formulário de normalização de uma localidade

Capítulo 5

Análise de Resultados

Neste capítulo de análise de resultados faz sentido comparar em modo geral a ideia inicial com a implementação existente.

A implementação conseguida difere em muito do apresentado na ideia inicial, principalmente no que toca à interacção entre os módulos de aquisição e ao aproveitamento que estes fazem da informação da base de dados, o que não compromete a sua funcionalidade, mas salienta a necessidade futura de fortalecer este aspecto que muito contribui para a interactividade da aplicação.

Dada a inexistência de ajudas contextuais relacionadas com a base de dados no módulo de transcrição de registos, este não permite referir indivíduos existentes na base de dados e, conseqüentemente, não possui nenhuma ligação ao módulo de navegação.

Este facto realça a necessidade de mapear a informação dos registos transcritos para os formulários, o que, se implementado, iria permitir ao utilizador verificar a informação antes de a submeter para a base de dados e seleccionar indivíduos existentes, evitando a criação de clones.

O facto de nenhum destes mecanismos se encontrar na implementação actual compromete em parte o mapeamento correcto dos registos transcritos para a base de dados, o que exigirá uma verificação dos valores inseridos na base de dados aquando da sua fusão com a base de dados central.

Quanto aos módulos de recolha de dados segundo formulários e de navegação pela base de dados, a implementação existente garante praticamente toda as funcionalidades descritas na ideia, excepto a visualização dos grafos de relações entre indivíduos, não implementada por falta de recursos de suporte a esta funcionalidade.

5.1 Dificuldades de Implementação

Ao longo deste trabalho, sobretudo na fase inicial, foram inúmeras as dificuldades apresentadas pela falta de recursos .Net para uma plataforma já limitada que é a plataforma Pocket PC.

De modo geral, as dificuldades relacionam-se com limitações gráficas ou com a falta de bibliotecas de suporte para funcionalidades menos básicas. Um dos

problemas mais gritantes, relacionado com o ambiente .Net, é que o controle de exceções é muito mais fraco na Compact .Net Framework do que na .Net Framework normal, faltando o tipo e descrição das exceções na maioria dos casos. Um exemplo explícito disso é o facto de a classe de validação de XML segundo um XML Schema, desenvolvida para a validação dos registos transcritos, retornar erros de validação completos quando corrida num PC e exceções sem nenhuma informação no PDA.

Relativamente ao editor de registos XML embutido no módulo de transcrição de registos, a maior dificuldade foi a utilização de texto formatado, com cores ou estilos variados. Embora estranho, a Compact .Net Framework não possui uma classe de suporte a RichText (o controlador existente de terceiros é pago e não existe nenhuma informação na Web sobre como utilizar estas funcionalidade), pelo que a criação de um tal controlador implicou o desenvolvimento de raiz, com base em reverse engineering de mensagens de sistema e muito cálculo binário de posições de memória, uma vez que a plataforma .Net é fechada.

Uma outra dificuldade no editor de registos, esta apenas ao nível do problema, foi manter a estrutura XML implícita no editor actualizada com o texto formatado apresentado ao utilizador. Isto envolve a associação de eventos próprios a diferentes acções do teclado do dispositivo, de forma a poder replicar essas mesmas acções.

Uma outra contrariedade que acentua a falta de recursos para esta plataforma é a inexistência de uma ligação exterior no contexto da nossa aplicação, uma vez que para este género de plataformas o paradigma cliente/servidor através da utilização de webservices funciona como uma maneira de poupar processamento ao dispositivo móvel, o que seria extremamente importante na interacção com a base de dados, e de disponibilizar tecnologias apenas disponíveis para PCs, uma vez que o processamento dos dados no servidor é independente do cliente.

Por fim, mas também importante de referir, foi o por vezes difícil mapeamento entre a base de dados e os conceitos demográficos, no que toca às propriedades de cada tipo de registo demográfico e a por vezes não normalização das tabelas da base de dados ou a associação de tipos menos correctos a determinados campos. Um exemplo disso é a definição dos índices das tabelas como texto e não como numéricos, o que dificulta em muito o processo de geração de índices únicos, exigindo queries e cálculos adicionais para a sua transformação.

Por exemplo, umas das dificuldades encontradas a desenvolver a camada de base de dados, foi a necessidade de incrementar os identificadores únicos, usados nos objectos e na indexação de tabelas na base de dados (id do indivíduo, id de família, id de assinatura, etc), dado que estamos a usar um sistema OO e necessitamos de gerar novos índices em conformidade com a base de dados, mas evitando estar constantemente a efectuar ligações à base de dados para ir buscar o próximo "id". Portanto, para este propósito, são carregados os valores dos "ids" ao criar a interface de de conexão à base de dados, que por sua vez carrega esses valores para variáveis de classe, garantindo que todos os objectos criados têm identificadores únicos.

Resumindo, a maioria das dificuldades encontradas estão relacionadas com a política de fontes fechadas da Microsoft, uma vez que perante a falta de meios é nos impossível consultar código de outras aplicações ou das próprias classes de modo a adaptar estas às nossas soluções, aumentando a dificuldade de implementação de funcionalidades por si complicadas dados os escassos recursos

de um dispositivo móvel.

Capítulo 6

Conclusão

Ao longo do relatório, apresenta-se o SEED, um sistema informático de apoio aos estudos demográficos, e descreve-se um projecto que empreendemos no sentido de estudar formas de tornar o uso de PDA's viável para a aquisição de dados no campo, isto é, com arquivos onde as fontes documentais são guardadas.

Concluído o projecto, pensamos ter desenvolvido um conjunto de módulos e uma aplicação que respondem positivamente ao esforço neles depositado, funcionais, com potencial de aplicação na área de estudo e com uma solução que resolve os problemas de esquematização, tanto na recolha como na representação de informação da base de dados.

No entanto, foram inúmeras as dificuldades relacionadas com a plataforma de desenvolvimento, devido à grande falta de recursos, tanto no framework como em fóruns de comunidades, resultantes da política fechada do ambiente .Net e da especificidade da plataforma de desenvolvimento para PDA. Estas dificuldades acentuaram o esforço inicial e tornaram algumas das funcionalidades projectadas impraticáveis no contexto deste projecto, visto envolverem a implementação de raiz de ferramentas de suporte às funcionalidades Pocket PC necessárias ou a compra de licenças das mesmas.

Todas as contrariedades encontradas levam-nos a afirmar que o “statu quo” do desenvolvimento para PDA em geral ainda não está suficientemente “maduro” de modo a garantir funcionalidades não básicas. Deste modo, a tecnologia ainda tem que evoluir nesse sentido e a utilização de um ambiente mais livre e multi-plataforma como, por exemplo, Java, poderia ter as suas vantagens.

Concluindo, como qualquer projecto deste cariz, a solução idealizada não foi desenvolvida na sua totalidade devido à sua componente teórica, mas a solução implementada é largamente suficiente para demonstrar a praticabilidade dos conceitos de transcrição de documentos representação de dados aplicados. Consideramos que as metodologias de trabalho e de estruturação da solução ajustadas e seguras se reflectem numa aplicação bem programada e com código organizado, de modo a garantir condições para um trabalho futuro de melhoria da mesma.

Independentemente dessas considerações e como qualquer produto de mercado, o seu maior sucesso dependerá da reacção dos utilizadores finais às interfaces e funcionalidades disponibilizadas, factor apenas calculável aquando da sua aplicação em campo.

6.1 Trabalho Futuro

Terminado este projecto, muitos campos podem ainda ser explorados, quer por não terem sido explorados nesta implementação, quer pelo período de tempo limitado do projecto.

Enunciamos de seguida algumas funcionalidades ou propriedades que pensamos virem a acrescentar valor à solução desenvolvida:

- Ligar directamente a aplicação de especialização e conversão (4.2), ao dispositivo, evitando o uso do *active-sync* por parte do utilizador.

Esta ligação teria de ser feita sem a utilização de um *web-service*, mas antes via ligações, USB e Bluetooth por exemplo, pelas razões já referidas anteriormente (3.1).

- Melhorar fortemente as opções de selecção da vista da BD a levar para o campo, ou mesmo acrescentar sugestões inteligentes, tais como indicar as paróquias que costumam ter mais interacções com a paróquia que vai alvo do estudo.
- Adicionar funcionalidades de pesquisa e de visualização gráfica das relações entre sujeitos no módulo de navegação pela informação dos indivíduos e famílias. Isto como já foi referenciado(3.3) iria melhorar a navegação pela informação.
- Explorar possíveis utilizações ao nível da localização geográfica fornecida pelo hardware GPS do dispositivo.
Neste campo estarão funcionalidades como associar coordenadas geográficas a localidades conhecidas e pesquisar/indexar informação da base de dados conforme as coordenadas geográficas actuais.
- No sentido de agilizar e adaptar o processo de escrita, faz sentido explorar melhor a utilização de expressões regulares na automatização e simplificação do processo de marcação de anotações, adicionando-lhe um sistema de gestão de abreviaturas mantido pelo utilizador.
- Explorar a ligação à rede. Este ponto está descrito ao pormenor em baixo.

Como o projecto desenvolvido visa a transcrição de registos no campo de trabalho dos historiadores, não foi explorada a utilização de uma ligação à Web.

No entanto, sendo o dispositivo um dispositivo móvel, com inúmeras possibilidades de este se conectar ao exterior pela rede de telemóvel ou por uma rede local, podemos considerar o melhoramento das funcionalidades do sistema através da utilização de tecnologias Web ou da criação de novas funcionalidades dependentes de uma ligação à Internet:

- Criar um sistema de navegação online pela informação dos indivíduos e famílias. Este sistema, além de clonar o interface gráfico de navegação para formato web, poderá ainda gerar visualizações dos documentos XML. A disponibilização de registos em formato XML já foi previamente explorada no projecto de Rafael Félix, através da aplicação de um *stylesheet* XSL aos documentos[20]. Este sistema é independente do dispositivo móvel, uma vez que a sua utilização faz mais sentido fora do local de recolha dos dados.

- Fazer a sincronização no campo, caso exista o acesso à rede, submetendo assim quaisquer mudanças que tenham ocorrido. Esta solução implica que os dados passem por um módulo de fusão e conversão semelhante ao já desenvolvido (2.2), existente no servidor, que seria encarregue de os colocar na BD central.

O projecto SEED é um projecto com história e já com uma utilidade comprovada. No entanto, novas tecnologias e novos requisitos exigem que o trabalho até hoje realizado no projecto seja reintegrado numa solução escalável, verdadeiramente modular, extensível e duradoura. Para tal, elaboramos um possível plano de acções a tomar, que esperamos alimentar trabalhos futuros neste contexto:

- melhorar a estrutura e tipos de dados utilizados na base de dados,
- mudar a plataforma em que se baseia a base de dados actual (Access), que já provou ter muitas deficiências e não ser facilmente escalável,
- criar uma "verdadeira" base de dados central, que utiliza o modelo cliente-servidor,
- associar à BDC inúmeros serviços, desde obtenção de vistas específicas, documentos XML, acesso online a informação, geração automática de páginas WEB, etc,
- implementar várias SIG, permitindo fazer estudos geográficos e territoriais sobre os mais diversos aspectos da demografia histórica, baseados no dados recolhidos.

Para levar este projecto adiante, julgamos que será imprescindível estudar detalhadamente todos os aspectos da conversão e efectuar uma especificação formal da solução, pois a conversão terá que ser completa, segura e sem falhas.

O acesso a BD em tempo real implica para além da mudanças na estrutura da base de dados, a sua conversão para novas soluções tecnológicas. Dessa forma a BDC, poderia ser implementada em PostgreSQL, por esta ser uma alternativa livre e de elevada qualidade, (outros sistemas de poderiam ser utilizados, tais como Oracle). As BDPs utilizariam a estrutura da nova BDC, bem como SQLite na sua implementação que comprovou a sua utilidade como base de dados, leve, portátil e rápida.

Desenvolver-se-ia um novo módulo de fusão e conversão, que permitiria trabalhar com a nova BDC, fundindo-a com as BDP e resolvendo os conflitos criados interactivamente. Esta iria funcionar como um serviço externo disponibilizado por a BDC, através de uma aplicação tradicional, mas também de um serviço de rede.

O trabalho desenvolvido anteriormente não seria perdido, mas sim, aproveitada toda a investigação e conhecimento obtido, sendo este ultimo módulo visto como uma actualização do conFusBDP. O mesmo se aplicaria à revisão do módulo de recolha de dados.

Seriam também desenvolvidas aplicações de migração de bases de dados, entre os formatos antigos e novos, permitindo uma conversão de todo o material de forma faseada.

Pensamos que esta abordagem seria útil para este projecto a longo prazo, dotando-o com capacidades de crescimento, providenciando mais e melhores serviços, bem como solidificar os existentes.

Bibliografia

- [1] eXtensible Markup Language.
<http://en.wikipedia.org/wiki/XML>.
- [2] Extensible Markup Language (XML).
<http://www.w3.org/XML/>.
- [3] Microsoft Developer Network.
<http://msdn2.microsoft.com/>.
- [4] Microsoft Jet Database Engine.
http://en.wikipedia.org/wiki/Microsoft_Jet_Database_Engine.
- [5] Microsoft SQL Server: SQL Server 2005 Compact Edition.
<http://www.microsoft.com/sql/editions/compact/default.aspx>.
- [6] OpenNETCF.org.
<http://www.opennetcf.org/home.ocf>.
- [7] Rich Text Format (RTF) Specification.
[http://msdn2.microsoft.com/en-us/library/aa140277\(office.10\).aspx](http://msdn2.microsoft.com/en-us/library/aa140277(office.10).aspx).
- [8] SQL Mobile.
http://en.wikipedia.org/wiki/SQL_Mobile.
- [9] SQLite.
<http://en.wikipedia.org/wiki/SQLite>.
- [10] SQLite home page.
<http://www.sqlite.org/>.
- [11] System.Data.SQLite.
<http://sqlite.phxsoftware.com/>.
- [12] The Code Project.
<http://www.codeproject.com/>.
- [13] XML Schema.
<http://www.w3.org/XML/Schema>.
- [14] XML Schema (W3C).
http://en.wikipedia.org/wiki/XML_Schema.
- [15] Microsoft Access, 2006.
http://en.wikipedia.org/wiki/Microsoft_Access.
- [16] Maria Norberta B. Amorim.
Uma Metodologia de Reconstituição de Paróquias.
Technical report, Universidade do Minho, 1991.

- [17] Fernanda Faria.
Seed - módulo de aquisição de dados - relatório de desenvolvimento.
Technical report, NEPS / Departamento de Informática Universidade do Minho, 2004.
- [18] Fernanda Faria.
SEED - Módulo de Aquisição de Dados - Manual do Utilizador, 2005.
- [19] João Antero Gonçalves Ferreira.
Sistema de aquisição de dados para a reconstituição de paróquias : a re-
produção social em s. João das Caldas (1600-1910).
Master's thesis, Instituto de Ciências Sociais - Universidade do Minho,
2001.
- [20] Rafael Félix.
Sistema de digitalização e anotação de documentos.
2002.

Lista de Siglas e Acrónimos

ADO.NET	<i>ActiveX Data Objects .NET</i>
BD	Base de Dados
BDC	Base de Dados Central
BDP	Base de Dados Paroquial
CE	<i>Compact Edition</i>
ConFusBDP	Módulo de conversão e fusão de bases de Dados Paroquiais
GPS	<i>Global Positioning System</i>
GUI	<i>Graphical User Interface</i>
HTML	<i>HyperText Markup Language</i>
IDE	<i>Integrated Development Environment</i>
JET	<i>Joint Engine Technology</i>
MRP	Método de Reconstituição de Paróquias
NEPS	Núcleo de Estudos da População e Sociedade
OO	Orientado aos Objectos
PC	<i>Personal Computer</i>
PDA	<i>Personal Digital Assistant</i>
Pocket BDP	Base de Dados Paroquial para Pocket PC
RP	Registo Paroquial
SEED	Sistema de Informação para o Estudo da Evolução Demográfica
SIG	Sistema de Informação Geográfica
SQL	<i>Structured Query Language</i>
SRP	Sistema de Reconstituição de Paróquias
USB	<i>Universal Serial Bus</i>
XML	<i>Extensible Markup Language</i>

XSD *Extensible Markup Language Schema Definition*

XSL *Extensible Stylesheet Language*

Apêndice A

XML Schemas Utilizados

Segue em seguida a estrutura simplificada dos três tipos de XML Schemas utilizados, um por cada registo de baptizado, casamento e óbito.

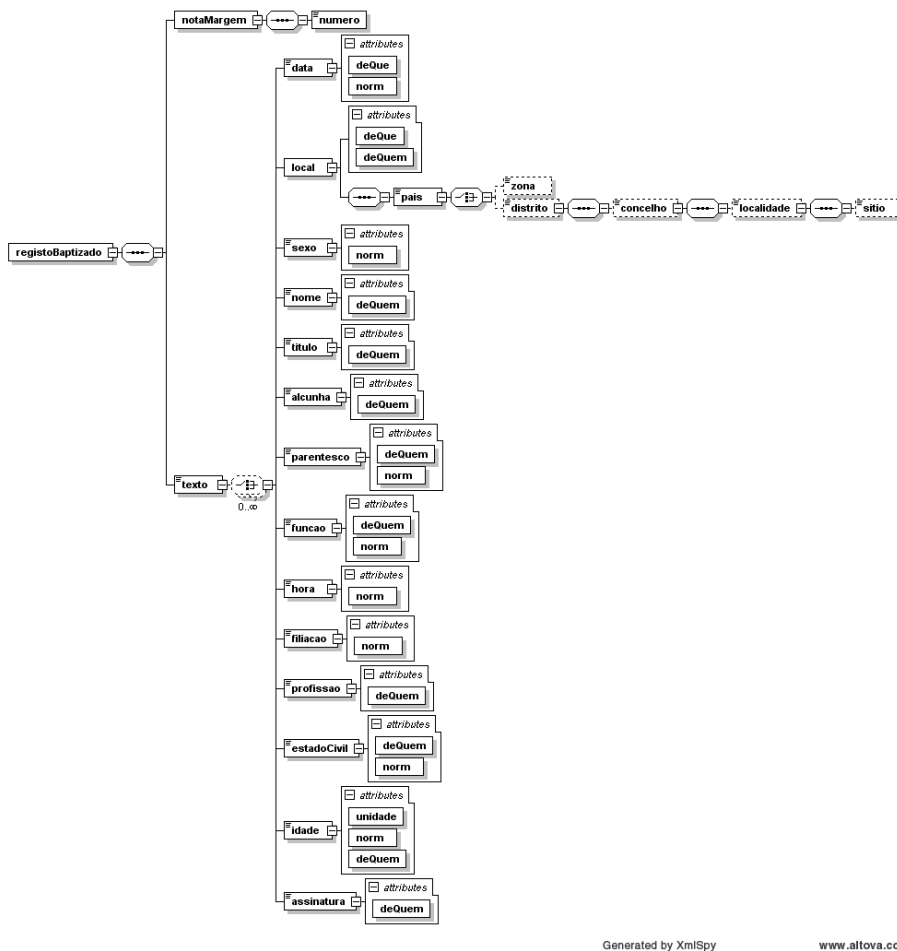
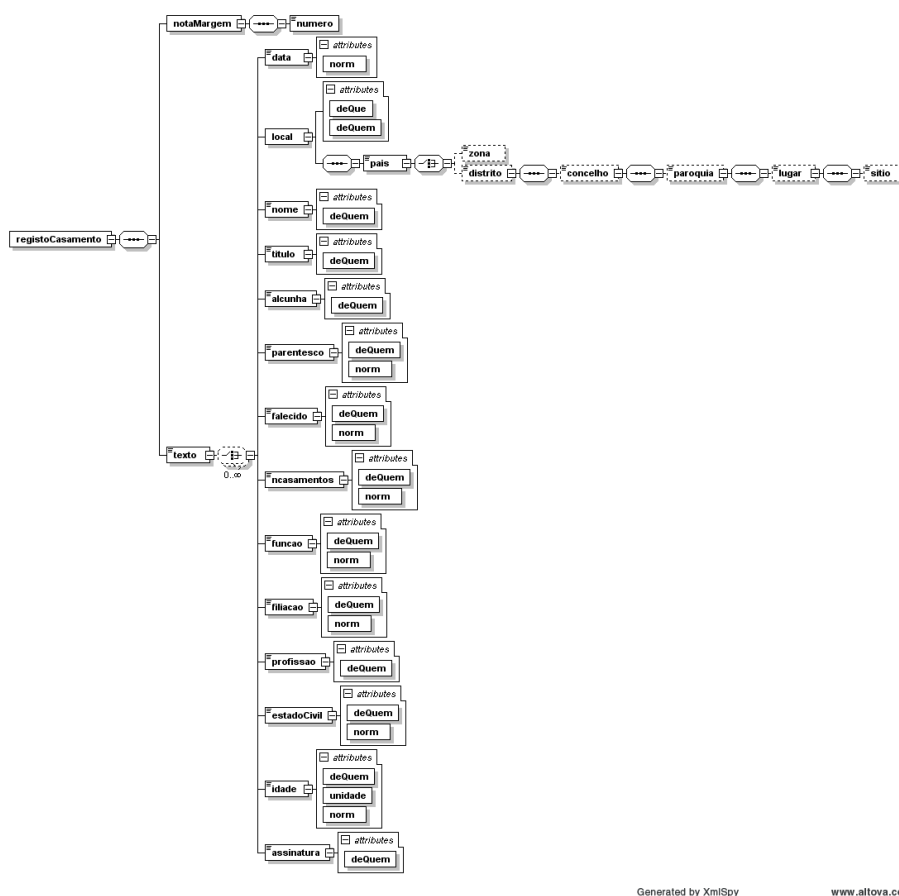


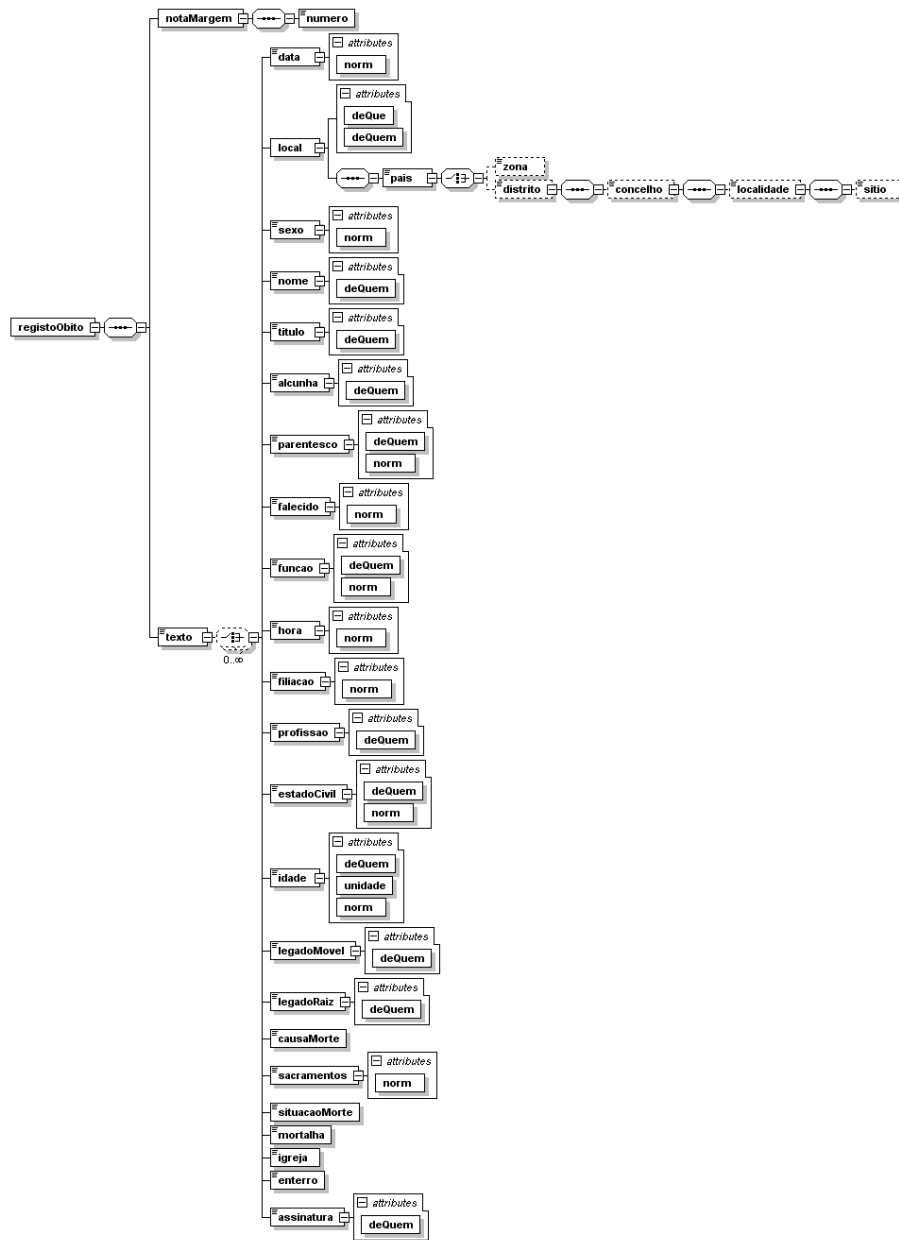
Figura A.1: Schema para registos de baptizado



Generated by XmiSpy

www.altova.com

Figura A.2: Schema para registros de casamento



Generated by XmlSpy

www.altova.com

Figura A.3: Schema para registos de óbito