

# Topic Maps aplicados ao sistema de informação do Museu da Emigração

Giovani Rubert Librelotto<sup>1</sup>,  
José Carlos Ramalho<sup>2</sup>, and Pedro Rangel Henriques<sup>2</sup>

<sup>1</sup> UNIFRA, Centro Universitário Franciscano, Santa Maria - RS, 97010-032, Brasil  
giovani@unifra.br

<sup>2</sup> Universidade do Minho, Departamento de Informática  
4710-057, Braga, Portugal  
{jcr, prh}@di.uminho.pt

**Abstract.** O presente artigo apresenta uma aplicação do *Metamorphosis* ao caso do Museu Virtual da Emigração. O Museu da Emigração contém no seu espólio fontes de informação compostas por documentos XML e por uma base de dados relacional. A função do *Metamorphosis* é propiciar uma visão homogénea destes recursos através da criação de um topic map que represente este universo de discurso. Para isso, usa-se o *Oveia* para a extracção do topic map e o *Ulisses* para a navegação sobre o conhecimento do domínio. O resultado é uma visão integrada do Museu da Emigração, de acordo com a ontologia especificada pelo projectista.

## 1 Introdução

Devido ao facto de as pessoas viverem em constante mudança, o ritmo das suas vidas as faz incorrer em viagens (por puro lazer, por necessidade ou por trabalho). Num passado não muito longínquo, era frequente encontrar em almanaques locais, histórias sobre a vida de determinados indivíduos ou sobre eventos em que os mesmos participavam. Também era usual fazer-se um diário de viagem quando se realizava uma jornada por vários locais.

O *Museu da Emigração*<sup>3</sup> (ME) foi criado para reunir e disponibilizar à comunidade informações diversas acerca dos emigrantes em geral e, em particular, daqueles que emigraram para o Brasil.

Particularmente importante nesse sentido é que o ME é um museu virtual. Para tal, a informação a ser exibida no ME é recolhida a partir de diversas fontes de dados, tais como registos de passaportes, almanaques, jornais, diários de viagem e certidões de nascimento. Esta lista de possíveis fontes de dados não está limitada, pois novas fontes (recheadas de informação relevante para o ME) podem sempre aparecer.

Combinando várias fontes – como por exemplo, as fontes que contém dados sobre os almanaques, diários de viagem e passaportes – pode-se disponibilizar uma quantidade significativa de informação sobre um determinado indivíduo,

<sup>3</sup> Museu da Emigração – <http://www.museu-emigrantes.webside.pt>

reconstruindo-se assim o seu percurso de vida que irá contextualizar e enquadrar o seu processo de emigração.

Pensando agora em termos do Sistema de Informação digital (SIME) que irá suportar o Museu Virtual da Emigração, é importante observar que, pela natureza dos documentos aqui envolvidos e por razões práticas e factuais, as fontes de informação estão armazenadas em suportes heterogéneos: algumas são guardadas em bases de dados relacionais, enquanto outras estão na forma de textos anotados em XML.

Neste artigo pretende-se a integração de *fontes nominativas históricas heterogéneas*, isto é, armazenadas de formas diversas em suportes diferentes através do uso do Metamorphosis, o qual vai representar o conhecimento extraído do SIME na norma ISO 13250 Topic Maps. O objectivo é propiciar à comunidade em geral a extracção do conhecimento a partir das fontes de dados do Museu da Emigração, através da navegação conceptual sobre a informação contida nas seguintes fontes heterogéneas.

Este artigo encontra-se dividido em 4 secções. A Secção 2 descreve as fontes de dados que fazem parte do SIME, apresentando sua estrutura e conteúdo. A Secção 3 define os passos a serem seguidos para obter a interoperabilidade sobre os recursos de informação do Museu da Emigração através do uso do Metamorphosis: o primeiro passo, apresentado na Subsecção 3.2, descreve como o Oveia propicia a geração de um topic map a partir de recursos heterogéneos contendo o conhecimento do universo de discurso do SIME; por fim, a Subsecção 3.3 disponibiliza a visualização do domínio do ME através de páginas HTML geradas pelo Ulisses. A conclusão do caso de estudo em questão se apresenta na Secção 4.

## 2 Sistema de Informação do Museu da Emigração

Normalmente a informação relevante que se pretende extrair de qualquer sistema está armazenada em formatos distintos. Este foi o caso encontrado no âmbito do SIME. Os dados que alimentam esse sistema estão armazenados em documentos XML (para as fontes de dados referentes aos *Almanaques* e aos *Diários de Viagem*), de onde se pode extrair as histórias da vida e de viagens referentes aos emigrantes. Combinando estes documentos XML com uma base de dados que contém a informação relativa aos seus passaportes, fica-se com um conjunto de informações sobre o percurso de vida e histórias com relevo de emigrantes.

As próximas subsecções descrevem a estrutura e exemplificam o conteúdo de cada fonte de dados pertencente ao SIME.

### 2.1 Notas Biográficas contidas em Almanaque

Os *almanaques* são publicações, normalmente periódicas e nacionais (embora por vezes possam ter um cariz mais local), com calendário, informações científicas, tabelas, informações úteis sobre o cotidiano e alguns textos mais amenos como contos, notas biográficas, descrição de eventos sociais, entretenimento, etc.

Contudo, os almanaques não se encontravam escritos de uma forma facilmente manipulável (de um ponto de vista informático). De um modo geral, as histórias dos almanaques não possuem um conteúdo semelhante entre si, podendo abranger um grande leque de descrições sociais, psicológicas, físicas, além de conter informação ambígua e parcial.

Para seu armazenamento digital, definiu-se então uma estrutura XML que permita a extração e catalogação das informações possíveis de ser encontradas nas Notas Biográficas destes almanaques. Abaixo encontra-se a anotação em XML do exemplo de história de vida do emigrante “*José Pereira Leite*”.

```
1 <Notas_Biograficas>
2   <fonte>Almanaque de Fafe</fonte>
3   <titulo>José Pereira Leite</titulo>
4   <tipoMemoria valor="outro"/>
5   <MEMORIA>
6     <PESSOA ID="ID000006">
7       <nome>José Pereira Leite</nome>
8       <profissao>Comércio</profissao>
9     </PESSOA>
10    <HISTORIA_A>
11      <notasAdicionais>
12        <perfil valorPerfil="psicologico">Sincero e Dedicado</perfil>
13      </notasAdicionais>
14      <localNasc>Natural de Chacim, Cabeceiras de Basto</localNasc>
15      <paisEmigrou>Brasil</paisEmigrou>
16      <notasAdicionais>
17        <perfil valorPerfil="profissional">Graças à sua força de vontade e
18        actividade, possui uma boa casa de negócio</perfil>
19        <relSociais>Tem ao seu serviço o Sr. Joaquim de Macedo Ferraz e Souza,
20        conhecidíssimo na vila de Fafe</relSociais>
21      </notasAdicionais>
22      <docAnexo>foto6</docAnexo>
23    </HISTORIA_A>
24  </MEMORIA>
25 </Notas_Biograficas>
```

Esta história anotada em XML possui a mesma informação acerca desse emigrante que a página de um almanaque em papel. A informação está estruturada e explicitamente marcada (para permitir uma interpretação objectiva, não ambígua), tornando viável o seu arquivo electrónico e o processamento deste documento por um parser XML. As notas biográficas retiradas dos Almanaxes de Fafe no âmbito do ME estão juntamente armazenadas em ficheiros de texto que seguem o formato acima exemplificado.

## 2.2 Diários de Viagem

Um *Diário de Viagem* é um documento onde uma pessoa escreve suas jornadas, ou seja, notas de embarque, repouso e o que fez ou não em certos locais. Estas notas são acompanhadas da data do respectivo apontamento. Resumindo, é um diário que apenas se cinge a uma dada viagem efectuada por uma pessoa, contendo informações como a data, o local e uma descrição textual de sítios visitados (monumentos ou paisagens) ou episódios vivenciados.

Os diários de viagem são muito genéricos, sucintos e podem ocorrer grandes intervalos de tempo em que não existe informação (pode não existir uma descrição contínua, metódica e temporal do relato da viagem).

Assim como no caso dos almanaques, foi definido um esquema genérico o suficiente, tornando possível a catalogação da informação contida num diário. Este esquema define a estrutura lógica, em anotação XML, para um diário de viagem, o qual é exemplificado pelo documento abaixo:

```

1 <Notas_Bio>
2   <PESSOA ID="ID000006">
3     <nome>José Pereira Leite</nome>
4     <idade>25</idade>
5   </PESSOA>
6   <HISTORIA_V ID="HIST0100">
7     <nota data="1905-01-07">Início da viagem à Europa: Amanhã.</nota>
8     <nota data="1905-01-08">Sahida hoje do Rio, para a Bahia, ...</nota>
9     <nota data="1905-01-09">...</nota>
10    <nota data="1905-01-10">...</nota>
11  </HISTORIA_V>
12 </Notas_Bio>

```

Em termos do SIME, todos os diários de viagem recolhidos e pertencentes ao âmbito do ME encontram-se anotados em XML, tal qual o exemplo acima.

### 2.3 Registos de Passaportes

Os passaportes contém a informação referente às viagens efectuadas por um emigrante, informando qual foi o destino da viagem (podendo ser múltiplos destinos) e se ele foi acompanhado por alguma outra pessoa (podendo ser múltiplos acompanhantes).

A informação referente aos passaportes está estruturada seguindo o modelo relacional em quatro tabelas: *Passaportes* (contém os dados básicos dos passaportes), *Viagem* (contém os dados referentes às viagens efectuadas por determinado emigrante), *Acompanhantes* (contém os dados sobre os acompanhantes dos emigrantes em cada viagem) e *Pessoa* (contém os dados pessoais de cada emigrante). A Figura 1 apresenta estas quatro tabelas e os relacionamentos entre elas, no diagrama entidade-relacionamento dessa base de dados.

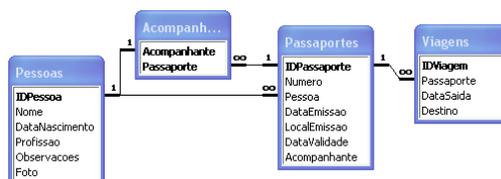


Fig. 1. Diagrama Entidade-Relacionamento da base de dados dos Passaportes.

Os relacionamentos da Figura 1 indicam que uma pessoa pode possuir vários passaportes, assim como ter vários acompanhantes associados a ela. Por sua vez, cada passaporte poderá referenciar vários destinos.

### 3 Passos para a Integração de Fontes Heterogêneas

A partir da definição de quais fontes de informação compõe o SIME – base de dados dos passaportes e documentos anotados XML para os almanaques e os diários de viagem – utiliza-se o *Metamorphosis* para se obter uma visão homogênea destas fontes heterogêneas.

O *Metamorphosis* [3] é um ambiente formado por um conjunto de linguagens de especificação e ferramentas que permitem criar uma interface para integração de informação oriunda de diversas fontes, através do uso de uma ontologia representada em *Topic Maps* [5]. A partir da descrição das fontes heterogêneas de informação e da especificação da ontologia, o *Oveia* [4] (um dos componentes do *Metamorphosis*) extrai automaticamente o respectivo topic map. Depois de guardado, a componente *Ulisses* [3] gera uma interface Web para manipular o topic map extraído.

Neste processo, o topic map não foi validado sintática e semanticamente (face a um conjunto de restrições especificadas numa linguagem apropriada) pelo Processador *XTche* [3] (outro dos componentes no *Metamorphosis*) pelo facto do topic map ter sido criado por uma ferramenta de construção automática, o *Oveia*, reduzindo o número de restrições que possam detectar erros, pois certos casos já são verificados pelo próprio processo de criação. Assim sendo, a validação *XTche* iria apenas confirmar que o *Oveia* realizou seu processo de uma maneira correcta.

O processo envolvendo a geração dessa visão é composto por uma série de etapas. As primeiras etapas requerem um especialista no universo de discurso em questão, pois ele será responsável por definir qual a visão que será oferecida aos utilizadores do sistema, ou seja, aos visitantes do Museu Virtual da Emigração. Esse especialista deve possuir sólidas noções sobre a norma *Topic Maps*, permitindo assim tornar a especificação o mais abrangente possível.

#### 3.1 Por que Topic Maps?

*Topic Maps* [1] é um formalismo para representar conhecimento acerca da estrutura de um conjunto de recursos de informação e para o organizar em tópicos. Esses tópicos possuem ocorrências e associações que representam e definem relacionamentos entre os tópicos. A informação sobre os tópicos pode ser inferida ao examinar as associações e ocorrências ligadas ao tópico. Uma colecção desses tópicos e associações é designada *topic map*. *Topic Maps* pode ser visto como um paradigma que permite organizar, manter e navegar através da informação, permitindo transformá-la em conhecimento.

Tendo em vista que os *Topic Maps* são grafos compostos por tópicos (onde cada tópico representa um tema, identifica recursos e está associado com outros tópicos), o *Ulisses* foi imaginado e desenvolvido para providenciar navegadores estáticos que permitem percorrer estas redes de conceitos. Esses navegadores são compostos de páginas HTML que descrevem os tópicos e usam a ideia de hiper-ligações para implementar as associações e as ligações às ocorrências.

### 3.2 Construção do topic map pelo Oveia

Esta secção tem por objectivo demonstrar, passo a passo, o processo de geração do topic map a partir dos dados extraídos das fontes heterogéneas de informação. O topic map gerado segue a ontologia descrita na especificação XS4TM (*XML Specification for Topic Maps*) [4] que descreve o domínio do ME.

**Especificação das fontes em XSDS.** Esta etapa é responsável por definir a especificação XSDS (*XML Specification for Data Sources*) [4] que determina quais fontes de dados serão integradas e quais os dados que estão em cada fonte. A especificação XSDS divide-se em duas etapas: definição dos recursos físicos a serem interrogados; e selecção das partes de informação contidas nestes recursos necessárias para a construção da rede semântica.

A especificação em XSDS de cada fonte define um elemento `<datasource>`. Abaixo representa-se a especificação das fontes Almanagues e Passaportes:

```
1 <resources>
2   <datasources>
3     <!-- Declaração da Fonte de Dados dos Almanagues -->
4     <datasource extratorDriver="br.uneb.dcet.tmbuilder.drivers.XMLFile" name="XML_Almanaque">
5       <parameter name="pathDocument">C:\\MUSEU\\Almanagues\\Alm.xml</parameter>
6     </datasource>
7     <!-- Declaração da Fonte de Dados dos Passaportes -->
8     <datasource extratorDriver="br.uneb.dcet.tmbuilder.drivers.DataBase" name="BD_Passaporte">
9       <parameter name="connectionURL">jdbc:access://localhost/museu/passaporte</parameter>
10      <parameter name="password"/>
11      <parameter name="user"/>
12      <parameter name="jdbcDriver">sun.jdbc.odbc.JdbcOdbcDriver</parameter>
13    </datasource>
14  </datasources>
15  <datasets/>
16 </resources>
```

Para o acesso aos documentos XML, basta apenas informar o seu caminho na estrutura de directorias do sistema operativo. Aliado a isto, define-se o nome de cada *dataset* que será criado (no atributo *name*) e informa-se o *driver* responsável pela extracção (no atributo *extratorDriver*). Para o acesso na base de dados relacional Microsoft Access® define-se um nome para o *dataset* a ser criado e do driver de extracção. Além disso, é necessário a criação de quatro parâmetros: a URL de conexão, o utilizador, sua palavra-passe e o *driver* JDBC.

Uma vez definidas as fontes de dados, parte-se para a definição de quais elementos (em documentos XML) ou campos (em base de dados) devem ser extraídos. Para isso, o elemento `<dataset>` é utilizado, o qual referencia qual fonte de dados (*datasource*) será a matéria-prima para a extracção.

```
1 <resources>
2   <datasources>
3     ...
4   </datasources>
5   <datasets>
6     <dataset name="Alm-Emigrante" database="XML_Almanaque">
7       <columns>
8         <column identify="true">@ID</column>
9         <column>nome</column>
10        <column>idade</column>
```

```

11 |         <column>profissao</column>
12 |     </columns>
13 |     <statement>//Notas_Biograficas/MEMORIA/PESSOA</statement>
14 | </dataset>
15 | <dataset>
16 |     ...
17 | </dataset>
18 |     ...
19 | </datasets>
20 </resources>

```

O *dataset* acima apresentado efectua uma extracção no documento XML dos Almanques. O conjunto de dados extraídos será referenciado pelo nome de *Alm-Emigrante* e conterà 4 elementos, referidos nos elementos <column>. Cada um desses elementos será encontrado fisicamente através do caminho XPath declarado no elemento <statement>. Para exemplificar, o *dataset* chamado *Alm-Emigrante* formará uma tabela, onde cada linha conterà as seguintes informações extraídas da fonte *XML\_Almanaque*:

- //Notas\_Biograficas/MEMORIA/PESSOA/@ID – será o identificador deste *dataset*, definido pelo atributo *identify*.
- //Notas\_Biograficas/MEMORIA/PESSOA/nome
- //Notas\_Biograficas/MEMORIA/PESSOA/idade
- //Notas\_Biograficas/MEMORIA/PESSOA/profissao

Para o caso da definição do *dataset* extraído da base de dados dos passaportes (contendo o seu identificador, nome, nome do pai e nome de mãe), o seguinte código elemento <dataset> deve ser definido:

```

1 |     <dataset name="Passaporte-Emigrante" database="BD_Passaporte">
2 |         SELECT ID, Nome, NomePai, NomeMae FROM Pessoas
3 |     </dataset>

```

Percebe-se, deste modo, que a linguagem de interrogação às fontes de dados varia de acordo com a própria fonte: quando a extracção é realizada sobre documentos XML, utiliza-se a linguagem XPath; quando a extracção é realizada sobre base de dados, utiliza-se a linguagem SQL.

**Definição da ontologia do universo de discurso.** Esta subsecção apresenta a definição da ontologia para o Museu da Emigração, contendo a representação de todo o conhecimento deste domínio. A definição da ontologia consistiu, inicialmente, em determinar quais conceitos eram considerados importantes no domínio dos Almanques, Diários de Viagem e Passaportes e seguiu a metodologia definida por Gruber [2].

O próximo passo consiste da definição dos relacionamentos entre os conceitos. Então, o especialista identifica quais conceitos estão relacionados e cria uma associação entre ambos.

Para exemplificar, tomamos como exemplo o caso entre os conceitos *Emigrante* e *Histórias de Almanaque*. Ao definir uma associação entre estes dois conceitos, se fez necessário determinar como um conceito se relaciona com o outro, determinando que *um emigrante é referido numa história de almanaque*.

Após a identificação das associações, obteve-se um grafo representando a ontologia, conforme apresentado na Figura 2. Nesta figura, os nodos representam os conceitos, enquanto que as associações são representadas pelos arcos.

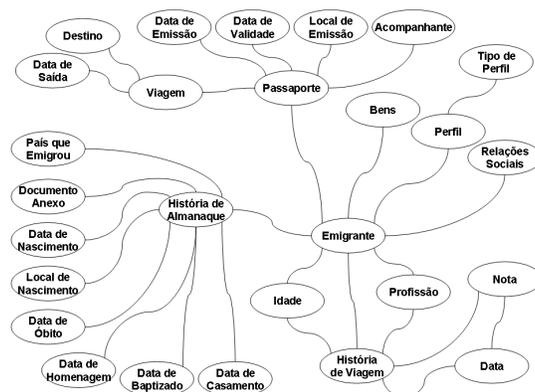


Fig. 2. Ontologia para o Museu da Emigração

Como o conceito *Emigrante* foi encontrado em todas as fontes, acabou por tornar-se um elo de união pois os conceitos *Emigrante* foram fundidos num único. Assim sendo, o conceito resultante mantém todos os relacionamentos que os demais participavam.

**Especificação da Ontologia em XS4TM.** A partir do momento que todos os conceitos e seus relacionamentos foram definidos, foi então especificado em XS4TM, a ontologia apresentada na subsecção anterior.

A primeira etapa da especificação em XS4TM consistiu na codificação dos tipos de tópicos (ou seja, os conceitos). A partir da sua definição, parte-se para a definição dos tipos das associações. Assim como os tipos de tópicos, os tipos das associações são definidos como sub-elementos do elemento <ontologies>.

Ao final deste passo, a ontologia do Museu da Emigração estará definida no elemento <ontologies>. A mesma consiste basicamente da definição dos tipos de tópicos, dos tipos de associações e dos papéis de actuação em associações.

**Instanciação da Ontologia em XS4TM.** Uma vez definida a ontologia, o próximo passo consiste de relacioná-la com os *datasets* especificados em XSDS. Esta relação entre conceitos e fonte de informação é realizada no elemento <instances> da especificação XS4TM.

Em XS4TM, cada elemento //instances/topic e //instances/association possui um atributo chamado dataset, o qual determina de qual *dataset* os dados serão obtidos para a construção do tópico ou associação em questão.

Como exemplo, apresenta-se a seguir a especificação da extracção dos tópicos que possuem o tipo *Emigrante*. Neste exemplo, o tópico *Emigrante* está definido para a fonte de dados *almanaques* (*dataset Alm-Emigrante*, linha 4 a 23). As demais fontes (*diários de viagem* (*dataset DViagem-Emigrante*) e *passaportes* (*dataset Passaporte-Emigrante*) terão uma representação similar.

```

1 | <xs4tm xmlns="http://www.topicmaps.org/xtm/1.0/" xmlns:xlink="http://www.w3.org/1999/xlink">
2 |   <ontologies>
3 |     <!-- Aqui vai a definição da ontologia do SIME -->
4 |   </ontologies>
5 |   <instances>
6 |     <topic dataset="Alm-Emigrante" id="@Alm-Emigrante.@ID">
7 |       <instanceOf>
8 |         <topicRef xlink:href="#Emigrante"/>
9 |       </instanceOf>
10 |       <baseName>
11 |         <baseNameString>@Alm-Emigrante.nome</baseNameString>
12 |       </baseName>
13 |       <occurrence>
14 |         <scope>
15 |           <topicRef xlink:href="#Idade"/>
16 |         </scope>
17 |         <resourceData>@Alm-Emigrante.idade</resourceData>
18 |       </occurrence>
19 |       <occurrence>
20 |         <instanceOf>
21 |           <topicRef xlink:href="#Profissao"/>
22 |         </instanceOf>
23 |         <resourceRef xlink:href="@Alm-Emigrante.profissao"/>
24 |       </occurrence>
25 |     </topic>
26 |   </instances>
27 | </xs4tm>

```

No processo de extracção realizado pelo Oveia, será criado um tópico para cada emigrante encontrado nos *datasets* acima. Num segundo passo, o Oveia faz uma fusão entre os tópicos que possuem o mesmo identificador, de acordo com as regras definidas na norma ISO 13250 [1]. Deste modo, se um mesmo emigrante é encontrado nas três fontes de dados do SIME, um único tópico conterá toda a informação acerca do emigrante.

Esta especificação irá determinar a extracção de todos os tópicos do tipo *Emigrante*, os quais conterão um nome e até quatro ocorrências: idade, profissão (provenientes dos *datasets Alm-Emigrante* e *dataset DViagem-Emigrante*), nome do pai e nome da mãe (originadas do *dataset Passaporte-Emigrante*). Esse processo de associação dos tópicos com os *datasets* deve ser realizado para todos os tipos de tópicos definidos na ontologia desse domínio.

Para a especificação de associações, apresenta-se o exemplo do relacionamento entre *Emigrantes* e *Histórias de Almanaque*. A associação abaixo então relaciona os tais papéis com os tópicos que desempenharão cada função:

```

1 |   <association dataset="Alm-Emigrante">
2 |     <instanceOf>
3 |       <topicRef xlink:href="#Emigrante-historiaAlmanaque"/>
4 |     </instanceOf>
5 |     <member>
6 |       <roleSpec>
7 |         <topicRef xlink:href="#emigrante-referido-historia"/>
8 |       </roleSpec>

```

```

9 |         <topicRef xlink:href="@Alm-Emigrante.@ID"/>
10 |     </member>
11 | </member>
12 |     <roleSpec>
13 |         <topicRef xlink:href="#historia-refere-emigrante"/>
14 |     </roleSpec>
15 |     <topicRef xlink:href="@Alm-HistoriaAlmanaque.pessoa"/>
16 | </member>
17 | </association>

```

Baseada nesta especificação de associações, será criada uma nova associação toda vez que os campos contidos nos *datasets* @Alm-Emigrante.@ID e @Alm-HistoriaAlmanaque.pessoa forem iguais. Na prática, quando isto ocorrer, indica que um emigrante está sendo citado numa história.

A cardinalidade, neste caso, será definida pelas fontes de dados; ou seja, se um mesmo emigrante está referido em várias histórias de almanaques, será criada uma associação para cada referência encontrada nas fontes. Assim, um mesmo emigrante pode ser citado em várias histórias de almanaque.

**Extracção do topic map com o Oveia.** O processo de extracção por intermédio do Oveia foi realizado em 4 etapas. Na Etapa 1 são definidas as informações referentes ao projecto de extracção. Nesta etapa, pode-se escolher um projecto existente para re-processar a extracção (caso o mesmo já tenha sido realizado anteriormente) ou criar um novo projecto.

Após definir o projecto, definem-se as configurações de saída do Oveia, na Etapa 2. Para a especificação do armazenamento do topic map, selecciona-se qual é o tipo da base de dados a ser criada; por exemplo: SQL Server, MySQL, Oracle, etc. A seguir, insere-se os **drivers** de conexão com a base de dados, o protocolo utilizado, o endereço do servidor da base de dados, o utilizador e a palavra-passe necessárias para o estabelecimento da conexão. Para a especificação da criação do topic map num ficheiro XTM[6], basta seleccionar a opção *XML file*.

A Etapa 3 consiste no processo de extracção propriamente dito. Nesta etapa realiza-se a execução da extracção. Após o término do processamento, uma estatística completa referente ao processo é fornecida, onde é informado o tempo de execução de todo o processo, os elementos extraídos e o número de elementos que não foram extraídos, devido a alguma falha.

Ao final de todo o processamento, foi obtido um ficheiro de acordo com a sintaxe XTM com 35588 linhas, contendo a especificação de 1043 tópicos, 25 tipos de tópicos, 1541 associações e 32 tipos de associações, com um tamanho final de 1,14 MB (1.197.322 bytes). Por sua vez, o Oveia armazenou o topic map gerado em tabelas, o qual contem o mesmo número de tópicos e associações que a versão XTM.

### 3.3 Navegação no topic map com Ulisses

A partir do topic map gerado pelo Oveia, o Ulisses se encarrega de produzir o website referente ao domínio do ME.

A integração das fontes é percebida na visualização das informações referentes a um emigrante. A Figura 3 apresenta a página HTML correspondente ao emigrante *Sérgio Carvalho*. Além da ocorrência textual do tipo *Profissão*, o emigrante em questão está envolvido em 27 associações, as quais são de 12 tipos distintos. Essas associações foram definidas a partir dos dados extraídos das 3 fontes. Por exemplo:

- a associação “*Sérgio Carvalho é citado em uma nota com a data 1985-08-13*” é oriunda de um diário de viagem, que citava o emigrante;
- a associação “*Sérgio Carvalho é considerado alto, forte e barrigudo*” tem origem num almanaque;
- a associação “*Sérgio Carvalho tinha o passaporte CM420589*” surgiu dos dados obtidos a partir da base de dados dos passaportes.

Analisando desta maneira, toda a informação referente a este emigrante está apresentada de uma forma homogênea, pois para o utilizador que está a navegar no website pouco importa de qual fonte vieram os dados responsáveis por definir uma associação de um conceito; o importante é obter toda a informação disponível nas fontes de dados sobre o conceito desejado.

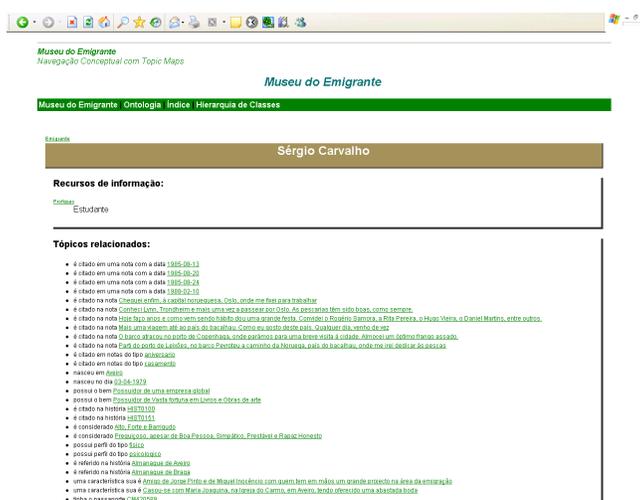


Fig. 3. Página sobre um emigrante

A página referente a um emigrante é o melhor exemplo de uma integração de fontes heterogêneas, pois como se pode observar na Figura 2 (referente à ontologia sobre o domínio do ME), o conceito *Emigrante* está relacionado com histórias de almanaques, passaportes e histórias de viagens, ou seja, todas as fontes de dados podem fornecer informação sobre o mesmo emigrante.

Através dos *links* que descrevem cada associação, pode-se navegar conceitualmente através do grafo formado pelo topic map do ME. Portanto, cada *link*

representa um arco que conecta dois nodos relacionados (as páginas que descrevem os conceitos).

## 4 Conclusão

O caso de estudo em questão permitiu aplicar o *Metamorphosis* na combinação de fontes heterogêneas de dados que alimentam o Sistema de Informação do Museu da Emigração (SIME), de modo a facultar aos possíveis visitantes uma navegação sobre os dados extraídos das fontes de uma forma transparente e homogênea.

Em termos práticos, foram utilizados dois dos módulos do *Metamorphosis* para a obtenção da finalidade pretendida: o *Oveia* extraiu os dados contidos nas fontes heterogêneas de informação (uma base de dados de passaportes e dois documentos XML de almanaques e diários de viagem), construindo um topic map baseado numa especificação *XS4TM*; e o *Ulisses* gerou um website a partir do topic map validado, no qual se pode navegar sobre os recursos do SIME.

A representação do conhecimento do domínio do Museu da Emigração está então descrita num topic map que contém a rede semântica formada pelos conceitos e os seus relacionamentos. Este topic map pode ser intercambiado posteriormente com outras aplicações pois a sintaxe *XTM* é a mais difundida e aceite pela comunidade académica e empresarial para a representação de *Topic Maps*.

O intercâmbio do topic map gerado com outras aplicações pode vir a permitir que o conhecimento acerca deste universo de discurso possa ser integrado com outros topic maps que representem domínios relacionados ao Museu da Emigração, aumentando assim a riqueza semântica da rede resultante.

Igualmente importante, é o facto de se poder gerar uma vista completamente diferente dos dados retirados das mesmas fontes, bastando para isso reformular a ontologia a representar, definindo-a em *XS4TM*.

## References

1. Michel Biezunsky, Martin Bryan, and Steve Newcomb. ISO/IEC 13250 - Topic Maps. ISO/IEC JTC 1/SC34, December 1999. <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>.
2. Thomas R. Gruber. How to Design an Ontology. In *Web page as part of the Ontolingua guided tour*. Accessed in April 2002, 1995.
3. Giovanni Rubert Librelotto. *XML Topic Maps: da Sintaxe à Semântica*. PhD thesis, Departamento de Informática, Escola de Engenharia, Universidade do Minho, 2005.
4. Giovanni Rubert Librelotto, Weber Souza, José Carlos Ramalho, and Pedro Rangel Henriques. Using the Ontology Paradigm to Integrate Information Systems. In *International Conference on Knowledge Engineering and Decision Support*, pages 497–504, Porto, Portugal, 2004.
5. Jack Park and Sam Hunting. *XML Topic Maps: Creating and Using Topic Maps for the Web*, volume ISBN 0-201-74960-2. Addison Wesley, 2003.
6. Steve Pepper and Graham Moore. XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification, August 2001. <http://www.topicmaps.org/xtm/1.0/>.