

INTEGRAÇÃO DAS FONTES DE DADOS NO MUSEU DA EMIGRAÇÃO PORTUGUÊS¹

INTEGRATION OF DATA SOURCES AT THE PORTUGUESE EMIGRATION MUSEUM

**Guilherme Chagas Kurtz², José Carlos Ramalho³ e Pedro Rangel
Henriques³, Giovanni Rubert Librelotto⁴**

RESUMO

O presente artigo apresenta uma aplicação do *Metamorphosis* ao caso do Museu Virtual da Emigração de Portugal. 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 homogênea desses recursos por meio da criação de um *topic map* que represente esse universo de discurso. Para isso, usa-se o *Oveia* para a extração 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 projetista.

Palavras-chave: Museu Virtual, documentos, visão integrada.

ABSTRACT

This article presents an application of Metamorphosis to the case of the Portuguese Emigration Virtual Museum. The museum has information sources composed of XML documents and a relational database. The function of Metamorphosis is to allow a homogeneous vision of these resources through the creation of a topic map that represents this discourse range. For that, Oveia is used for the building of a topic map and Ulisses for the navigation on the dominion knowledge. The result is an integrated vision of the Emigration Museum, according to the ontology projected by the designer.

Keywords: *Virtual Museum, documents, integrated vision.*

¹ Trabalho de Iniciação Científica - UNIFRA.

² Acadêmico do Curso de Ciência da Computação - UNIFRA.

³ Coorientadores - Universidade do Minho, Portugal.

⁴ Orientador - UNIFRA.

INTRODUÇÃO

Devido ao fato de as pessoas viverem em constante mudança, o ritmo de 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 dos quais participavam. Também era usual fazer um diário de viagem quando se realizava uma jornada por vários locais.

O *Museu da Emigração - Comunidades e Luso-Descendentes* (ME) foi criado para reunir e disponibilizar à comunidade informações diversas acerca dos emigrantes, em particular, daqueles que emigraram para o Brasil.

Nesse sentido, é importante o fato da ME ser um museu virtual. Para isso, a informação a ser exibida no ME é recolhida a partir de diversas fontes de dados, tais como registros de passaportes, almanaques, jornais, diários de viagem e certidões de nascimento. Essa 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 que contêm dados sobre os almanaques, diários de viagem e passaportes, pode-se disponibilizar uma quantidade significativa de informação sobre um determinado indivíduo, reconstruindo-se, assim, o seu percurso de vida que irá contextualizar e enquadrar o seu processo de emigração.

Pensando agora no Sistema de Informação digital (SIME) que irá suportar o Museu Virtual da Emigração, observa-se que, pela natureza dos documentos aqui envolvidos e por razões práticas, as fontes de informação estão armazenadas em suportes heterogêneos: algumas são guardadas em bases de dados relacionais, enquanto outras estão na forma de textos anotados em XML (HAROLD; MEANS, 2001).

Neste artigo, busca-se integrar as *fontes nominativas históricas heterogêneas*, isto é, as armazenadas de formas diversas em suportes diferentes pelo uso do *Metamorphosis*, o qual vai representar o conhecimento extraído do SIME na norma ISO 13250 *Topic Maps*. O objetivo é propiciar à comunidade em geral a extração do conhecimento a partir das fontes de dados do Museu da Emigração, por meio da navegação conceptual sobre a informação contida nas seguintes fontes heterogêneas.

Este artigo encontra-se dividido em quatro seções. A seção 2 descreve as fontes de dados que fazem parte do SIME, apresentando sua estrutura e conteúdo.

A seção 3 define os passos a serem seguidos para obter a interoperabilidade sobre os recursos de informação do Museu da Emigração pelo uso do *Metamorphosis*. A conclusão do estudo em questão apresenta-se na seção 4.

SISTEMA DE INFORMAÇÃO DO MUSEU DA EMIGRAÇÃO

Normalmente, a informação relevante que se pretende extrair de qualquer sistema está armazenada em formatos distintos. As próximas subseções descrevem a estrutura e exemplificam o conteúdo de cada fonte de dados pertencente ao SIME.

NOTAS BIOGRÁFICAS CONTIDAS EM ALMANAQUE

Os almanaques são publicações, normalmente periódicas e nacionais (embora, por vezes, possam ter um caráter 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, eles não se encontravam escritos de uma forma facilmente manipulável (de um ponto de vista informático), pois suas histórias não possuíam um conteúdo semelhante entre si, podendo abranger um grande leque de descrições sociais, psicológicas, físicas, além de conter informação ambígua e parcial.

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

```
<Notas_Biograficas>
  <fonte>Almanaque de Fafe</fonte>
  <titulo>José Pereira Leite</titulo>
  <tipoMemoria valor="outro"/>
  <MEMORIA>
    <PESSOA ID="ID000006">
      <nome>José Pereira Leite</nome>
      <profissao>Comércio</profissao>
    </PESSOA>
  </MEMORIA>
</Notas_Biograficas>
```

```

<HISTORIA_A>
  <notasAdicionais>
    <perfil valorPerfil="psicologico">Sincero e Dedicado</perfil>
  </notasAdicionais>
    <localNasc>Natural de Chacim, Cabeceiras de Basto</localNasc>
    <paisEmigrou>Brasil</paisEmigrou>
    <notasAdicionais>
      <perfil valorPerfil="profissional">Graças à sua força de vontade e
        atividade, possui uma boa casa de negócio</perfil>
    <relSociais>Tem ao seu serviço o Sr. Joaquim de Macedo Ferraz e
      Souza, conhecidíssimo na vila de Fafe</relSociais>
    </notasAdicionais>
    <docAnexo>foto6</docAnexo>
  </HISTORIA_A>
</MEMORIA>
</Notas_Biograficas>

```

Essa história, anotada em XML, possui a mesma informação sobre o emigrante da página de um almanaque em papel. A informação está estruturada e explicitamente marcada (para permitir uma interpretação objetiva, não ambígua), tornando viável o seu arquivo eletrônico e o processamento desse documento por um *parser* XML.

DIÁRIOS DE VIAGEM

Um *Diário de Viagem* é um documento em que uma pessoa escreve suas jornadas, ou seja, notas de embarque, de repouso e o que fez ou não em certos locais. Essas notas são acompanhadas da data do respectivo apontamento. Resumindo, é um diário que se refere a uma dada viagem efetuada por uma pessoa, contendo informações como a data, o local e uma descrição textual de sítios visitados ou episódios vivenciados. Os diários de viagem são muito genéricos, sucintos e podem ocorrer em 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 suficiente, o que tornou possível a catalogação de toda a informação contida num diário. Esse esquema define a estrutura lógica, em anotação XML, para um diário de viagem, o qual é exemplificado pelo documento abaixo:

```

<Notas_Bio>
  <PESSOA ID="ID000006">
    <nome>José Pereira Leite</nome>
      <idade>25</idade>
  </PESSOA>
  <HISTORIA_V ID="HIST0100">
    <nota data="1905-01-07">Início da viagem à Europa: Amanhã.</nota>
    <nota data="1905-01-08">Sahida hoje do Rio, para a Bahia, ...</nota>
      <nota data="1905-01-09">...</nota>
      <nota data="1905-01-10">...</nota>
  </HISTORIA_V>
</Notas_Bio>

```

Em termos do SIME, todos os diários de viagem recolhidos e pertencentes ao âmbito do ME encontram-se anotados em XML, como no exemplo acima.

REGISTROS DE PASSAPORTES

Os passaportes contêm informações referentes às viagens efetuadas 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).

As informações referentes aos passaportes estão estruturadas seguindo o modelo relacional em quatro tabelas: *Passaportes* (dados básicos dos passaportes); *Viagem* (dados referentes às viagens efetuadas por determinado emigrante); *Acompanhantes* (dados sobre os acompanhantes dos emigrantes em cada viagem) e *Pessoa* (dados pessoais de cada emigrante). A figura 1 apresenta essas quatro tabelas e os relacionamentos entre elas, no diagrama entidade-relacionamento (ER) dessa base de dados.

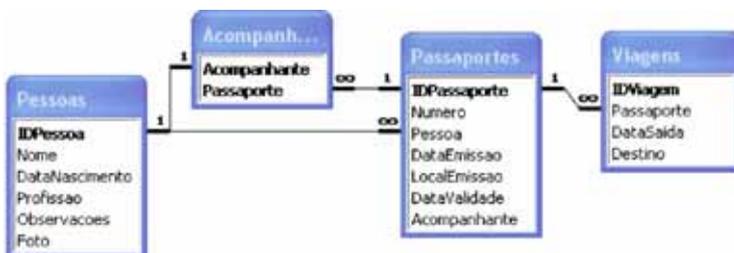


Figura 1 - Diagrama ER da base de dados dos Passaportes.

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.

PASSOS PARA A INTEGRAÇÃO DE FONTES HETEROGÊNEAS

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

O *Metamorphosis* (LIBRELOTTO et al., 2006) 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, por meio do uso de uma ontologia representada em *Topic Maps*. A partir da descrição das fontes heterogêneas de informação e da especificação da ontologia (GRUBER, 1993), o *Oveia* (um dos componentes do *Metamorphosis*) extrai automaticamente o respectivo *topic map*. Depois de guardado, a componente *Ulisses* gera uma interface *Web* para manipular o *topic map* extraído.

O processo que envolve a geração dessa visão é composto por uma série de etapas. As primeiras etapas requerem um especialista no universo do 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 Map*, desse modo, assim, tornar a especificação o mais abrangente possível.

POR QUE *TOPIC MAPS*?

Topic Maps (BIEZUNSKY et al., 1999) 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 suas associações e ocorrências. Uma coleção desses tópicos e associações é designada *topic map*. Assim, eles podem ser vistos como um paradigma que permite organizar, manter e navegar por meio da informação, permitindo transformá-la em conhecimento.

CONSTRUÇÃO DO *TOPIC MAP* PELO *OVEIA*

Esta seção tem por objetivo demonstrar, passo a passo, o processo de geração do *topic map*, a partir dos dados extraídos das fontes heterogêneas de informação. O *topic map* gerado segue a ontologia descrita na especificação XS4TM (*XML Specification for Topic Maps*) (LIBRELOTTO et al., 2006) que descreve o domínio do ME.

Especificação das fontes em XSDS

Essa etapa é responsável por definir a especificação XSDS (*XML Specification for Data Sources*) (LIBRELOTTO et al. 2006) 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 seleção das partes de informação, contidas nesses 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 Almanques e Passaportes:

```
<resources>
  <datasources>
    <!-- Declaração da Fonte de Dados dos Almanques -->
    <datasource extratorDriver="br.unifra.tmbuilder.drivers.XMLFile"
name="Almanaque">
      <parameter name="pathDocument">C:\\MUSEU\\Almanques\\Alm.xml</
parameter>
    </datasource>
    <!-- Declaração da Fonte de Dados dos Passaportes -->
    <datasource extratorDriver="br.unifra.tmbuilder.drivers.DataBase"
name="Passapt">
      <parameter name="connectionURL">jdbc:access://localhost/passapt</
parameter>
      <parameter name="password"/>
      <parameter name="user"/>
      <parameter name="jdbcDriver">sun.jdbc.odbc.JdbcOdbcDriver</parameter>
    </datasource>
  </datasources>
</datasets/>
</resources>
```

Para o acesso aos documentos XML, informa-se o caminho na estrutura de diretórios do sistema operacional. Aliado a isso, define-se o nome de cada *dataset* que será criado (no atributo *name*) e informa-se o *driver* responsável pela extração (no atributo *extratorDriver*). Para o acesso na base de dados relacional Microsoft Access®, determina-se um nome para o *dataset* a ser criado e para o *driver* de extração. Além disso, é necessária a criação de quatro parâmetros: a URL de conexão, o utilizador, sua senha e o *driver* JDBC.

Uma vez definidas as fontes de dados em **<datasource>**, 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 extração.

```
<resources>
  <datasources>
    ...
  </datasources>
  <datasets>
    <dataset name="Alm-Emigrante" database="XML_Almanaque">
      <columns>
        <column identify="true">@ID</column>
        <column>nome</column>
        <column>idade</column>
        <column>profissao</column>
      </columns>
      <statement>//Notas_Biograficas/MEMORIA/PESSOA</statement>
    </dataset>
    <dataset>
      ...
    </dataset>
    ...
  </datasets>
</resources>
```

O *dataset*, acima apresentado, efetua uma extração no documento XML dos Almanques. O conjunto de dados extraídos será referenciado pelo nome de *Alm-Emigrante* e conterá 4 elementos, referidos nos elementos **<column>**. Cada um desses elementos será encontrado fisicamente por meio do caminho XPath, declarado no elemento **<statement>**. Para exemplificar, o *dataset* chamado *Alm-Emigrante* formará uma tabela, em que cada linha conterá as seguintes informações extraídas da fonte *XML_Almanaque*:

- //Notas Biograficas/MEMORIA/PESSOA/@ID – será o identificador do *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:

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

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

Definição da ontologia do universo de discurso

Esta subseção apresenta a definição da ontologia para o Museu da Emigração, contendo a representação de todo o conhecimento desse domínio. A definição da ontologia consistiu, inicialmente, em determinar quais conceitos eram considerados importantes no domínio dos Almanques, dos Diários de Viagem e dos Passaportes e seguiu a metodologia definida por Gruber (1995). Esse passo consiste na definição dos relacionamentos entre os conceitos. Então, o especialista identifica quais conceitos estão relacionados e cria uma associação entre ambos.

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

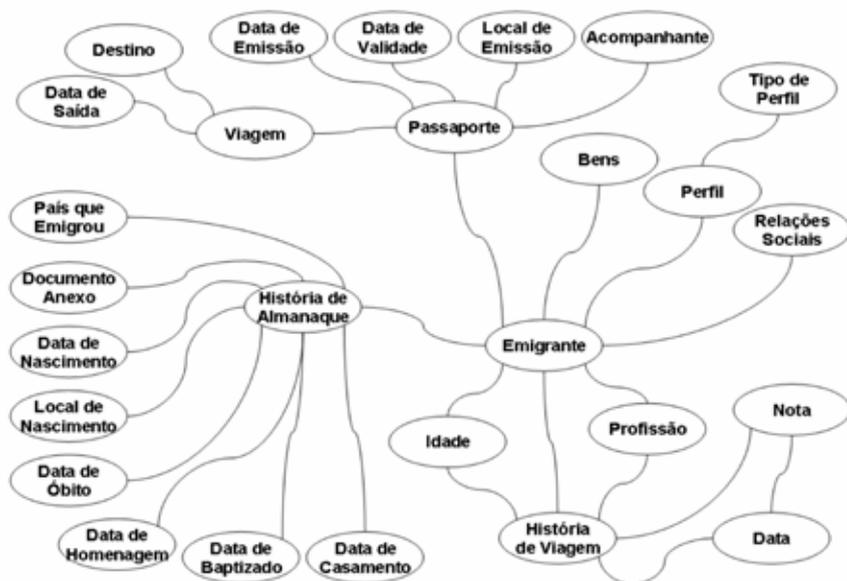


Figura 2 - Ontologia para o Museu da Emigração.

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

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

Especificação da Ontologia em XS4TM

A partir do momento que todos os conceitos e seus relacionamentos foram definidos, foi especificado, em XS4TM, a ontologia apresentada na subseção anterior. A primeira etapa da especificação consistiu na codificação dos tipos de tópicos (os conceitos). A partir da sua definição, partiu-se para a definição dos tipos de associações. Assim como os tipos de tópicos, os tipos de associações são definidos como subelementos do elemento <ontologies>.

Ao final desse passo, a ontologia do Museu da Emigração estará definida no elemento <ontologies>, que consiste basicamente na definição dos tipos de tópicos, dos tipos de associações e dos papéis de atuação em associações.

Instanciação da Ontologia em XS4TM

Uma vez definida a ontologia, o próximo passo consiste em relacioná-la com os *datasets* especificados em XSDS. Essa relação entre conceitos e fonte de informação é realizada no elemento <instances> da especificação XS4TM, em que cada elemento *//instances/topic* e *//instances/association* possui um atributo chamado *dataset* que determina de qual *dataset* os dados serão obtidos para a construção do tópico ou da associação em questão.

Como exemplo, apresenta-se, a seguir, a especificação da extração dos tópicos que possuem o tipo *Emigrante*. No exemplo, o tópico *Emigrante* é definido para a fonte de dados *almanaques* (*dataset Alm-Emigrante*). As demais fontes (*diários de viagem* (*dataset DViagem-Emigrante*) e *passaportes* (*dataset Passaporte-Emigrante*) terão uma representação similar.

```
<xs4tm xmlns="http://www.topicmaps.org/xtm/1.0/">
  <ontologies>
    <!-- Aqui vai a definição da ontologia do SIME -->
  </ontologies>
  <instances>
    <topic dataset="Alm-Emigrante" id="@Alm-Emigrante.@ID">
      <instanceOf>
        <topicRef xlink:href="#Emigrante"/>
          </instanceOf>
          <baseName>
            <baseNameString>@Alm-Emigrante.nome</baseNameString>
          </baseName>
          <occurrence>
            <scope>
              <topicRef xlink:href="#Idade"/>
            </scope>
            <resourceData>@Alm-Emigrante.idade</resourceData>
          </occurrence>
        </topic>
      </instances>
    </xs4tm>
```

No processo de extração 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 (BIEZUNSKY et al., 1999). Desse modo, se um mesmo emigrante é encontrado nas três fontes de dados do SIME, um único tópico conterá toda a informação acerca do emigrante.

Essa especificação determinará a extração 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.

Extração do *topic map* com o *Oveia*

O processo de extração por intermédio do *Oveia* foi realizado em 3 etapas. Na etapa 1 são definidas as informações referentes ao projeto de extração, como a localização dos ficheiros XSDS e XS4TM. Após definir o projeto, na etapa 2 definem-se as configurações de saída do *topic map* gerado pelo *Oveia*: ou uma base de dados relacional (SQL Server, MySQL, Oracle, etc.) ou um ficheiro XTM (PARK; HUNTING, 2003).

A etapa 3 consiste no processo de extração e armazenamento do *topic map*. No caso do SIME, 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 uma base de dados, a qual contém o mesmo número de tópicos e associações que a versão XTM.

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.

Desse modo, toda a informação referente a esse emigrante está apresentada de uma forma homogênea, pois para o utilizador que navega no *website* pouco importa de qual fonte vieram os dados responsáveis por definir uma associação de um conceito. Nesse caso, o importante é obter toda a informação disponível nas fontes de dados sobre o conceito desejado.

Museu do Emigrante

Museu do Emigrante Ontologia Índice Hierarquia de Classes

Sérgio Carvalho

Recursos de informação:

Estudante

Tópicos relacionados:

- é citado em uma nota com a data [1985-08-13](#)
- é citado em uma nota com a data [1985-08-20](#)
- é citado em uma nota com a data [1985-08-24](#)
- é citado em uma nota com a data [1985-02-13](#)
- é citado na nota [Chegou ontem à capital moçambicana, Oito, onde me foi para trabalhar](#)
- é citado na nota [Lindete Lima, Tereza e mais uma vez a passear por Oite. As passadas têm sido boas, como sempre.](#)
- é citado na nota [Houa foi a casa e como sempre hábito das uma grande festa. Convidei a Rosário Simões, a Rita Pereira, a Huel, Inês e Daniel Martins, entre outros.](#)
- é citado na nota [Meli, uma criança de 4 anos de idade, com os seus dois pais, Rui e Ana, estão de regresso.](#)
- é citado na nota [O barco atracou no porto de Coimbra, onde separamos para uma breve visita à cidade. Depois um jantar tranquilo e a noite.](#)
- é citado na nota [Parti do porto de Lisboa, no bairro Fátima a caminho da Fátima, para do trabalho, onde me foi pedir a ajuda.](#)
- é citado em notas do tipo [passagem](#)
- é citado em notas do tipo [passagem](#)
- nasceu em [Lisboa](#)
- nasceu no dia [02-04-1979](#)

Figura 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.

Por meio dos *links* que descrevem cada associação, pode-se navegar conceitualmente pelo grafo formado pelo *topic map* do ME. Portanto, cada *link* representa um arco que conecta dois nodos relacionados às páginas que descrevem os conceitos.

CONCLUSÃO

Esta pesquisa 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 facilitar 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 descrita num *topic map* que contém a rede semântica, formada pelos conceitos e seus relacionamentos. Esse *topic map* pode ser intercambiado posteriormente com outras aplicações, pois a sintaxe XTM é a mais difundida e aceita pela comunidade acadêmica e empresarial para a representação de *Topic Maps*.

O intercâmbio do *topic map*, gerado com outras aplicações, pode permitir que o conhecimento acerca desse universo de discurso possa ser integrado com outros *topic maps* que representem domínios relacionados ao Museu da Emigração, aumentando, desse modo, a riqueza semântica da rede resultante. Também é importante o fato de se poder gerar um olhar completamente diferente dos dados retirados das mesmas fontes, bastando, para isso, reformular a ontologia, definindo-a em XS4TM.

REFERÊNCIAS

BIEZUNSKY, M., BRYAN, M.; NEWCOMB, S. **ISO/IEC 13250 - Topic Maps. ISO/IEC JTC 1/SC34**. 1999. Disponível em: <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf>.

GRUBER, T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In: GUARINO, N.; POLI, R., editors, **Formal Ontology in Conceptual Analysis and Knowledge Representation**, Deventer, The Netherlands. Kluwer Academic Publishers, 1999.

GRUBER, T. R. How to Design an Ontology. **Web page as part of the Ontolingua guided tour**, 1995.

HAROLD, E. R.; MEANS, W. S. **XML in a Nutshell**. O'Reilly & Associates.

LIBRELOTTO, G. R.; RAMALHO, J. C.; HENRIQUES, P. R. ; Metamorphosis - A Topic Maps Based Environment to Handle Heterogeneous Information Resources. **Lecture Notes in Computer Science**, v. 3873, p. 14–25, 2001. Springer-Verlag GmbH.

PARK, J.; HUNTING, S. **XML Topic Maps: Creating and Using Topic Maps for the Web**, ISBN 0-201-74960-2. Addison Wesley, 2003.

