

Criação de Espaços Virtuais de Aprendizagem a partir de um repositório de documentos XML

Cristiana Araújo
Centro Algoritmi, Departamento de Informática
Universidade do Minho - Gualtar, Braga, Portugal
decrisianaaraujo@hotmail.com

Ricardo G. Martini
Centro Algoritmi, Departamento de Informática
Universidade do Minho - Gualtar, Braga, Portugal
rgm@algoritmi.uminho.pt

Pedro Rangel Henriques
Centro Algoritmi, Departamento de Informática
Universidade do Minho - Gualtar, Braga, Portugal
prh@di.uminho.pt

Maria João Varanda Pereira
Centro Algoritmi,
Instituto Politécnico de Bragança, Portugal
mjoao@ipb.pt

Palavras Chave – *Espaços Virtuais de Aprendizagem; Geração automática; DSL; Ontologias; XML; RDF.*

RESUMO

Para construir Espaços de Aprendizagem na Web, mais do que recolher e digitalizar informações, é necessário um poderoso mecanismo de busca e extração de dados e um sofisticado mecanismo de publicação na web. Neste artigo, é apresentado um sistema para construir automaticamente esses espaços de aprendizagem com base num repositório digital. O sistema extrai de um repositório de documentos XML a informação necessária para popular uma ontologia (a base de conhecimento que serve como núcleo do nosso sistema) criando uma representação interna tripla. Uma linguagem de domínio específico (CaVa^{DSL}) será usada para especificar os espaços de aprendizagem com base nessa ontologia. A descrição formal, escrita nessa linguagem, será processada pelo motor Cava^{gen} para gerar o espaço virtual de aprendizagem.

I. INTRODUÇÃO

As chamadas "memory institutions", como museus, arquivos ou bibliotecas, preservam hoje em dia as suas coleções ou recursos, como objetos digitais (bases de dados ou documentos anotados). Após digitalização e armazenamento, o objetivo imediato é explorar essas enormes fontes de relevante informação que constituem o patrimônio cultural da humanidade; Isso exige, pelo menos, mecanismos de pesquisa precisos e poderosos mecanismos de publicação na Web. Os museus virtuais - isto é, museus que não têm uma localização física e que não tem objetos físicos em exposição - surgiram neste contexto. Por outro lado, é suposto exibirem os objetos recolhidos nos repositórios digitais. Nesse caso, as salas de exposições (Espaços de Aprendizagem) são páginas Web; o visitante acede aos objetos através de um motor de busca [1].

Para criar um museu virtual na Web, é necessário consultar o armazenamento digital do repositório e processar (transformar e relacionar) a informação retornada antes de publicá-la em formato de página Web.

Como caso de estudo para implementação deste processo foram usados os recursos do Museu da Pessoa (MP) [2] [3] [4].

II. ARQUITECTURA DO SISTEMA

O núcleo ou o coração desta abordagem é uma ontologia que modela o domínio do conhecimento relacionado ao museu a ser construído. A plataforma que será introduzida, CaVa, divide o processo de construção num primeiro módulo, a função de ingestão, para extrair dados das fontes e carregar os triplos da ontologia (Figura 1) e um segundo módulo, o Generator (CaVa^{Gen}), para gerar automaticamente a consulta para cada sala de exposições (com base numa especificação formal e um subconjunto da ontologia principal) e organizar a informação retornada de forma a ser exibida adequadamente nas páginas Web (Figura 2).

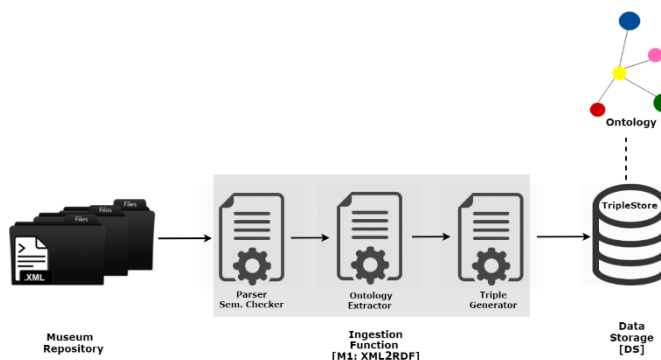


Figura 1. M1 – Extração de dados & Preenchimento dos Triplos

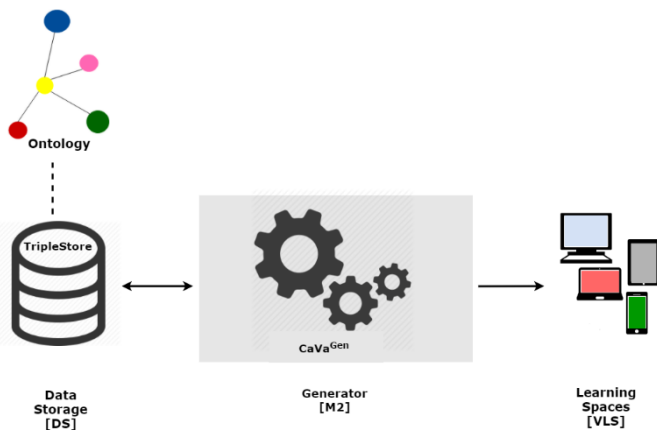


Figura 2. M2 – Geração dos Espaços Virtuais de Aprendizagem

A abordagem proposta pode então ser caracterizada por uma arquitetura (representada nas Figuras 1 e 2) que compreende: o repositório; a função de ingestão [M1] responsável por ler os documentos anotados, extrair e preparar os dados e armazenar as informações recolhidas; um armazenamento de dados (DS) que contém as instâncias da ontologia; uma ontologia que descreve o domínio do conhecimento que liga os conceitos através de um conjunto de relações [5]; o Gerador [M2] para receber e interpretar os pedidos de informações, aceder ao DS e retornar as respostas que são combinadas para configurar o Espaço Virtual de Aprendizagem final (VLS) [6] [7].

III. EXTRACÇÃO DE DADOS E POPULAÇÃO DA ONTOLOGIA: XML2RDF

O papel de [M1] – População da ontologia (Figura 1) é ler os documentos anotados, extrair e preparar os dados e armazenar as informações recolhidas.

Assim, para processar o repositório digital referido acima, construímos um filtro de texto para processar os dados de entrada automaticamente, produzindo um conjunto de triplos - XML2RDF [7] [8] [9]. Este filtro de texto foi desenvolvido usando o sistema gerador de compiladores AnTLR [11]. AnTLR gera um analisador léxico que implementa o filtro de texto desejado para extração de dados, com base em expressões regulares e num conjunto de regras de produção. Cada regra de produção é um par: no lado esquerdo, especificamos o elemento que queremos procurar - expressão regular (RE); O lado direito é um código que transforma os dados de entrada na respetiva saída [10].

Neste caso, a estrutura de entrada corresponde a XML (eXtensible Markup Language) e a saída será uma sequência de triplos (<sujeitos, predicados e objetos>). Em cada triplo (sujeito e objeto), os conceitos correspondem a dados que são o valor dos atributos de um elemento XML ou mesmo do conteúdo do elemento. As relações (predicados) que ligam os conceitos podem ser inferidas a partir dos elementos XML e da sua estrutura [10].

Na próxima secção, detalhamos a geração automática de espaços virtuais de aprendizagem virtual (VLS) para exibir

informações extraídas do tradutor XML2RDF num navegador Web.

IV. CAVA: GERAÇÃO AUTOMÁTICA DE ESPAÇOS VIRTUAIS DE APRENDIZAGEM

Como dito anteriormente, o CaVa é uma plataforma para gerar automaticamente Espaços Virtuais de Aprendizagem com base numa ontologia, que descreve repositórios de informações institucionais e numa especificação formal, escrita em CaVa^{DSL}, que define quais conceitos a ser exibidos e como eles devem estar organizados nas páginas Web.

O CaVa^{DSL} foi projetado para ser usado pelo curador de um museu, um arquivista ou um responsável por qualquer outra instituição cultural. A sintaxe da linguagem é simples, mas expressiva, permitindo que o utilizador final descreva as salas de exibição de forma fácil.

A estrutura CaVa^{DSL} é dividida em quatro blocos principais que descrevem a configuração principal, o cabeçalho, o conteúdo e o rodapé do espaço:

- A configuração principal (mainconfig): especifica o título do espaço e a descrição principal (por exemplo, o texto sobre a instituição cultural) e outros componentes gerais;
- Cabeçalho (menu): especifica o menu principal do espaço: a marca, o fundo e as cores do primeiro plano, o comportamento do menu e o tipo dos itens do menu (rótulo e link);
- O conteúdo (exposições): a lista de exposições. Para cada exposição especifica: um título, uma breve descrição e ícone; informações adicionais (título e descrição); comportamento (expandido ou colapsado); tipo de exibição ("permanente", "temporário", "futuro" ou "especial"); um operador de consulta ("all": que procura na ontologia todas as ocorrências do conceito especificado; "one": retorna apenas a primeira instância encontrada que corresponde ao procurado);
- O rodapé (rodapé): especifica uma área na parte inferior da página que compreende: imagens, data, nome da empresa ou instituição, comportamento e estilo;

CaVa^{Gen} é um conjunto de processadores que partindo de uma especificação CaVa^{DSL} e de uma ontologia populada produz um conjunto de ficheiros de saída com conteúdo estático e dinâmico do espaço virtual de aprendizagem. O Processador CaVa é o núcleo do CaVa^{Gen} e é descrito a seguir e o Processador CaVaSPARQLTriples trata da geração de conteúdo dinâmico, ou seja, a geração de consultas SPARQL. O Processador CaVa transforma uma especificação CaVa^{DSL} em vários ficheiros Web escritos em diferentes linguagens (HTML, PHP, JS, templates, CSS, etc.) que, quando colocados todos juntos, configuram várias páginas Web, ou seja, o espaço virtual de aprendizagem. A Figura 3 apresenta o esquema do Processador CaVa.

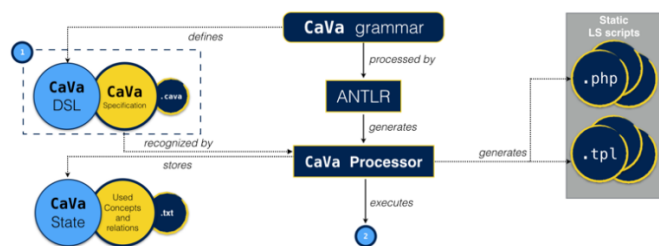


Figura 3. Processador CaVa

O retângulo identificado pelo número (1) está relacionado com o ficheiro de especificação (extensão ".cava") e descreve o espaço virtual de aprendizagem usando as regras do CaVa^{DSL}. É a principal entrada do Processador CaVa. A partir dessa especificação, o Processador CaVa gera o conteúdo estático do espaço (ficheiros .php e .tpl (template)). Como pode ser visto na Figura 3, o Processador CaVa é criado pelo Gerador de Compiladores ANTLR tomando como entrada a gramática da CaVa^{DSL}. O círculo CaVa State representa os ficheiros de estado necessários para armazenar os conceitos e relações da ontologia que vão ser usados na criação deste espaço. Este ficheiro .txt vai ser depois usado pelo Processador CaVaSPARQLTriples para gerar as queries necessárias para o conteúdo das páginas.

Para concluir, a abordagem proposta foi utilizada para implementar todos os VLS do Museu da Pessoa, cuja página principal é apresentada na Figura 4.

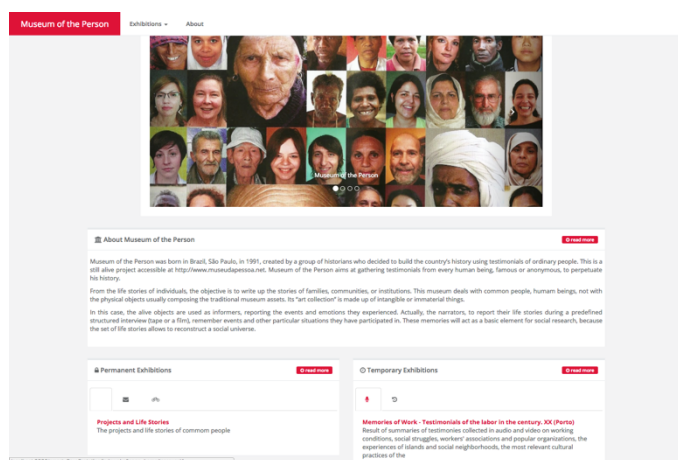


Figura 4. Museu da Pessoa – espaço virtual de aprendizagem gerado pelo CaVa

V. CONCLUSÃO

A descrição dos espaços virtuais de aprendizagem virtuais em linguagem simples e direcionada a utilizadores específicos, promove a geração de ambientes virtuais de aprendizagem e capacita o responsável da instituição cultural a se concentrar apenas na exposição do conteúdo (a estrutura das salas de exibição), organizando esse conteúdo da forma que considere

mais apropriada para o utilizador final. Isso elimina qualquer problema em relação à aprendizagem de várias linguagens de programação de propósito geral pela pessoa encarregada da instituição cultural, ou seja, aprender a linguagem CaVa^{DSL} será suficiente para especificar e gerar o espaço virtual de aprendizagem desejado.

Muitas novas ideias surgem ao longo do desenvolvimento deste projeto. No entanto, como trabalho futuro, o mais importante é realizar experimentos para avaliar a eficácia e usabilidade de nossa proposta, além de aplicar a abordagem de diferentes cenários e casos de estudo.

REFERÊNCIAS

- [1] W. Schweibenz, "The development of virtual museums, In Virtual Museums, 3rd ed., vol. 57(3), ICOM, 2004.
- [2] J. J. Almeida, J. G. Rocha, P. R. Henriques, S. Moreira, and A. Simões, "Museu da Pessoa – arquitetura," in Encontro nacional da associação de bibliotecários, arquivista e documentalistas, ABAD'01. BAD, 2001, URL <http://repositorium.sdum.uminho.pt/handle/1822/585>.
- [3] A. Simões and J. J. Almeida, "Histórias de vida + processamento estrutural = Museu da Pessoa," in XATA 2003 — XML: Aplicações e Tecnologias Associadas, page 16, Braga, Portugal, 2003, UM, URL <http://repositorium.sdum.uminho.pt/handle/1822/629>.
- [4] R. G. Martini, C. Araújo, J. J. Almeida, and P. R. Henriques, "OntoMP, An Ontology to Build the Museum of the Person", New Advances in Information Systems and Technologies, vol. 2, p.p 653–661. Springer International Publishing, Cham, 2016.
- [5] T. R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing". In International Journal of Human-Computer Studies, pp. 907–928. Kluwer Academic Publishers, 1993.
- [6] C. Araújo, R. G. Martini, P. R. Henriques, and J. J. Almeida, "Architectural Approaches to build The Museum of the Person," in Á. Rocha, L. P. Reis, M. P. Cota, O. S. Suárez, and R. Gonçalves, editors, Sistemas y Tecnologías de Información — Atas da 11ª Conferência Ibérica de Sistemas e Tecnologias de Informação, volume vol. 1 – Artículos de la Conferencia, pp. 383–388. AISTI–Associação Ibérica de Sistemas e Tecnologias de Informação, June 2016.
- [7] C. Araújo, "Building the Museum of the Person Based on a combined CIDOC-CRM/ FOAF/ DBpedia Ontology", MSc Thesis, Universidade do Minho, 2016.
- [8] C. Araújo, R. G. Martini, P. R. Henriques, and J. J. Almeida, "Building the Museum of the Person from RDF Triples and SPARQL", in Communications and Innovations Gazette (ComInG), 2nd ed., vol 1, pp. 1–14, 2016.
- [9] C. Araújo, R. G. Martini, P. R. Henriques, and J. J. Almeida, "Annotated documents and expanded CIDOC-CRM Ontology in the automatic Construction of a Virtual Museum", in New Advances in Informations Systems and Technologies, vol 2, Springer, 2018 (to appear soon).
- [10] C. Araújo, P. R. Henriques, and R. G. Martini, "Automatizing Ontology Population to drive the navigation on Virtual Learning Spaces in Á. Rocha, B. Alturas, C. Costa, L. P. Reis, and M. P. Cota, editors, Atas da 12ª Conferência Ibérica de Sistemas e Tecnologias de Informação, CISTI'2017, pp. 781–786. AISTI–Associação Ibérica de Sistemas e Tecnologias de Informação, June 2017.
- [11] Terence Parr, "The Definitive ANTLR 4 Reference" , 2nd ed., Pragmatic Bookshelf, 2013.