XATA 2010

XML, Associated Technologies and Applications

http://www.eseig.ipp.pt/conferencias/xata2010/

Abstracts Book

May 19–20, Vila do Conde

Escola Superior de Estudos Industriais e de Gestão

Instituto Politécnico do Porto

Wednesday — 19 May 2010

10h00 Conference Registration

11h00 Opening Session

Session I

Moderator: *José Carlos Ramalho*

11h30 Web Service for Interactive Products and Orders Configuration
António Arrais de Castro, Leonilde Varela, S. Carmo-Silva

Competition in the global economy is intensifying the implementation of the new paradigm of mass customization. This requires a substantial increase in the personalized interaction between client and producer. Agility in the order-production-delivery cycle optimization is a key element for enabling industrial enterprises to meet requirements of this paradigm. The use of agile methodologies requires improving the product specification process and data management. Under today's technology this tends to be predominantly carried out with the aid of the Internet using mostly web services. For a company, this also requires better integration of front-office processes (interaction with the outside) and back office (including production processes). Mass Customization scenarios are characterized by large product variety dependent on specific product requirements set by customers. In this process there is a need to provide the customer with tools for easy product specification, selection and/or configuration. Web-based configurators, can provide an opportunity for both, producer and customer, through an interactive process, for a more formal, faster, effective and better product and orders specification. In this paper, we propose an architecture and describe functionalities of a web service for interactive products and orders configuration. The proposed system can also be a valuable tool for supporting production and delivery.

12h00 Visual Programming of XSLT from examples
José Paulo Leal, Ricardo Queirós

Vishnu is a tool for XSLT visual programming in Eclipse — a popular and extensible integrated development environment. Rather than writing the XSLT transformations, the programmer loads or edits two document instances, a source document and its corresponding target document, and pairs texts between then by drawing lines over the documents. This form of XSLT programming is intended for simple transformations between related document types, such as HTML formatting or conversion among similar formats. Complex XSLT programs involving, for instance, recursive templates or second order transformations are out of the scope of Vishnu. We present the architecture of Vishnu composed by a graphical editor and a programming engine. The editor is an Eclipse plug-in where the programmer loads and edits document examples and pairs their content using graphical primitives. The programming engine receives the data collected by the editor and produces an XSLT program. The design of the engine and the process of creation of an XSLT program from examples are also detailed. It starts with the generation of an initial transformation that maps source document to the target document. This transformation is fed to a rewrite process where each step produces a refined version of the transformation. Finally, the transformation is simplified before being presented to the programmer for further editing.

12h30 XML Description for Automata Manipulations
José Alves, Nelma Moreira, Rogério Reis

GUItar is a visualization software tool for various types of automata (standard, weighted, push-down, transducers, Turing machines, etc.). It provides interactive manipulation of diagrams, comprehensive graphic style creation, multiple export/import filters, and a generic foreign function cal ls (FFC) interface with external systems. In this paper we describe GUItar's XML framework and show how it allows for extensibility, modularity and interoperability.

13h00 – Lunch

Soup

Creme de Abóbora / Pumpkin Cream

Main Course

Bacalhau Assado Lascado / Slashed Baked Codfish

Dessert

Buffet

Inclui água, sumo de laranja, vinho e café / Includes water, orange juice, wine and coffee.

Session II

Moderator: *Pedro Rangel Henriques*

14h30 Integration of repositories in Moodle

José Paulo Leal, Ricardo Queirós

Current Learning Management Systems focus on the management of students, keeping track of their progress across all types of training activities. This type of systems lacks integration with other e-Learning systems. For instance, learning objects stored in a centralized repository are unavailable throughout an organization for potential reuse. In this paper we present the interoperability features of crimsonHex - a service oriented repository of learning objects - highlighting the use of XML languages. Its interoperability features are compliant with the existing standards and we propose extensions to the IMS interoperability recommendation, adding new functions, formalizing an XML message interchange and providing also a REST interface. To validate the proposed extensions and its implementation in crimsonHex we designed two repository plugins for Moodle 2.0, the first of which is already implemented and is expected to be included in the next release of this popular learning management system.

15h00 Processing XML: a rewriting system approach

Alberto Simões, José João Almeida

Nowadays XML processing is performed using one of two approaches: using the SAX (Simple API for XML) or using the DOM (Document Object Model). While these two approaches are adequate for most cases there are situations where other approaches can make the solution easier to write, read and, therefore, to maintain. This document presents a rewriting approach for XML documents pro- cessing, focusing the tasks of transforming XML documents (into other XML formats or other textual documents) and the task of rewriting other textual formats into XML dialects.

These approaches were validated with some case studies, ranging from an XML authoring tool to a dictionary publishing mechanism.

15h30 XML to paper publishing with manual intervention

Oleg Parashchenko

Existing tools consider XML to paper conversion as one black box step and provide control only through predefined options. With this approach, tuning the layout of output documents is a burdensome task.

The paper advocates a new workflow for XML to paper conversion, in which a separate step allows the user to fine control the layout. The changes made by the user are remembered and later can be automatically re-applied during publishing an updated version of the document.

A possible technical implementation for the workflow is suggested. TEX is used as a typesetting engine. XML to TEX conversion is made using XSLT and TEXML. The management of changes is performed by diff and patch tools.

16h00 Coffee Break

Keynote

16h30 SAPO Services Bus

Marlene Marques

O Bus é a plataforma que suporta o Desenho, Aprovisionamento, Operação e Monitorização de mais de 200 serviços e 1000 operações no Sapo. Baseia-se nos standards abertos e em boas práticas como Contract-First development, gera 12.000 eventos de monitorização por segundo que alimentam um motor de Complex Event Processing, permitindo a análise em tempo real das suas condições de funcionamento e do estado dos serviços, e é actualmente a base da iniciativa de Orientação a Serviços na empresa. Nesta sessão, vamos apresentar os aspectos fundamentais da arquitectura da plataforma, focar no processo de desenho de serviços e mostrar como o uso de Domain Specific Languages e Software Factories pode contribuir decisivamente para a agilidade e organização na empresa.

Session III - Short Papers

Moderator: *Alberto Simões*

17h30 Parsing XML Documents in Java using Annotations

Renzo Nuccitelli, Eduardo Guerra, Clovis Fernandes

In Java language, to parse XML documents DOM, or some framework based on it, is a widely used solution. However, for large documents it is not possible use one of this approaches once they load the full XML tree, causing memory problems. In this context, it is advised the use of SAX, which is a pull parser and has an architecture based on Observer pattern. However, it has some drawbacks in its Java implementation, which make the parsing classes poorly extensible and granular. Intending to make pull parser development in Java easier, this paper proposes JColtrane, which is a framework that implements the parsing of XML documents based on code annotations. This paper also presents a case study which compares this approach with the regular SAX usage.

17h45 XML Archive for Testing: a benchmark for GuessXQ

Daniela Fonte, Pedro Carvalho, Daniela da Cruz, Alda Lopes Gançarski, Pedro Rangel Henriques

Aiming at making easier the process of information retrieval from structured documents, we developed an environment, GuessXQ, to provide query-by-example assistance. GuessXQ offers the user the chance to choose a family of documents, and to annotate graphically a sample document, picked from the collection. From the visual annotation, that exemplifies the user's

needs, GuessXQ infers a XQuery query that automatically applies to the collection to retrieve all the documents satisfying the information required by the user.

However, after developing a GuessXQ prototype, it is mandatory to validate pragmatically the effectiveness of the environment. This can be done, drawing a set of experiments. The success of such assessment strongly relies on the quality of collections used. This claim led to the construction of XAT, an archive of XML collections specially gathered in order to be useful for testing XML tools. Looking for a generic purpose, XAT was organized according to a classification criteria developed for this specific purpose.

This article is about the design and implementation of XAT.

18h00 First Day Closing Session

19h30 – Conference Dinner

The conference dinner will be held in Hotel Forte São João Baptista. This XVII century fort, originally built for the military defence of the River Ave, has recently been converted into a small, delightful hotel. The hotel offers luxury and comfort in superb surroundings. The Bartholomeu Restaurant at the Forte specializes in a wide variety of local dishes served in elegant surroundings.

There will be a shuttle from the conference venue to the restaurant, and from the restaurant to the conference venue and surrounding hotels.

Thursday — 20 May 2010

9h00 Warm-up and Certificates Delivering

Session IV

Moderator: *José Paulo Leal*

9h30 Test::XML::Generator: Generating XML for Unit Testing
Alberto Simões

To define a DTD or a Schema is not a trivial task. It can be compared to the task of preparing a data structure or, in some cases, to program that data structure adding some semantic. This makes this task error prone. It is common that a final Schema/DTD supports some special XML structures that should not be considered correct, or that, although these special structures are correct, they are not being correctly managed by the application parsing them.

We defend that the possibility to automatically generate XML documents based on a Schema/DTD can help preventing these situations. The generated documents can be used for unit testing and help tuning the Schema/DTD or fixing application problems. They can also assist on benchmarking issues as sometimes developers does not have access to full featured real world documents.

In this article we discuss a Perl module (Test::XML::Generator) that provides different mechanisms for automatically generate XML documents based on a DTD and a set of controlling parameters.

10h00 A Performance-based Approach for Processing Large XML Files in Multicore Machines
Filipe Felisberto, Ricardo Silva, Patricio Domingues, Ricardo Vardasca, Antonio Pereira

Due to its ubiquity, XML is used in many areas of computing, contributing to partially solve the problem of universal data representation across platforms. Although the parsing of XML files is a relatively well studied sub ject, processing large XML files with more than hundreds of megabytes pose many challenges. In this paper, we tackle several approaches focusing on how the performance can be improved when parsing very large XML files (hundreds of megabytes or even some gigabytes). We present a multithreaded block strategy that yields a roughly 2.19 relative speedup in a quad core machine when processing a 2.6 GB XML file.

10h30 A Refactoring Model for XML Documents
Guilherme Salerno, Marcela Pereira, Eduardo Guerra, Clovis Fernandes

Code refactorings are structural changes in the code without alterations on the external behavior. They are well known, studied and largely used. Many times it's also necessary to change the structure where data is stored. Data refactoring is more complicated because it involves, besides the structural change, migration of the existent data. The existing methods and tools developed for this purpose are focused only on database refactoring, but since XML is turning into the leading model for data formatting in several fields, it is also important to be considered. This work proposes a refactoring method specific for XML, which modifies the base XML Schema and updates the XML document through XSLT transformations. A tool was developed in order to validate the technique applicability. This methodology enables XML documents refactoring minimizing any collateral damage that may affect its consumer applications, since it is possible to obtain the document on their recognized format.

11h00 Coffee Break

Session V - Short Papers

Moderator: *Cristina Ribeiro*

11h30 XML, Annotations and Database: a Comparative Study of Metadata Definition Strategies for Frameworks
Clovis Fernandes, Douglas Ribeiro, Eduardo Guerra, Emil Nakao

The use of metadata in frameworks contributes for a cleaner code with properties unbounded from the implemented logic, helping for a better program development. An incorrect choice of metadata definition form can compromise the framework's extensibility and usage, but there is no study yet that compares different metadata definition forms. To address this issue, the present work compares three metadata definition strategies in face of different requirements, by developing frameworks that uses them and comparing the difficulty for its implementation and its usage. The conclusion of this work clarifies how and with which strategy metadata should be used according to the requirements to avoid future problems. As a result, developers can prevent implementation troubles and unnecessary work because of a bad choice for the metadata definition strategy.

11h45 A Semantic Representation of Users Emotions when Watching Videos
Eva Oliveira, Teresa Chambel, Nuno Ribeiro

One of the greatest strengths of video is its power to provoke emotions and induce states of mind, and is an excellent tool for displaying affective information. A need for appropriate tools to allow the characterization of video scenes with information about induced and expressed emotions is becoming more and more evident. Having video emotionally classified will enable video access and search through emotions, and it will also enable users to find interesting affective information in unknown or unseen videos. In this paper, we propose a set of semantic descriptors based on both user physiological signals, captured while watching videos, and on video low-level features extraction. These descriptors based on XML contribute to the creation of automatic affective meta-information that will not only enhance a video recommendation system based in

emotional information, but also enhance search retrieval of videos affective content from both users classification and content classification.

12h00 Integrating SVG and SMIL in DAISY DTB production to enhance the contents accessibility in the Open Library for Higher Education - Discussions and Conclusions
Bruno Giesteira, Inês Gomes, Alice Ribeiro, Diamantino Freitas

Following BAES investment in the use of DAISY DTB, it has become notorious the lack of authoring tools and user agents based in the most recent updates of the DAISY standard. The software in use nowadays does not allow the integration of SVG and SMIL files, two formats that can improve DTB, making contents richer and more accessible. It would be an important step for Portuguese Higher Education Institutions to be able to produce DTB with SVG images and SMIL videos.

12h15 CardioML: Integrating Personal Cardiac Information for Ubiquous Diagnosis and Analysis
Luis Coelho, Ricardo Queirós

The latest medical diagnosis devices enable the performance of e-diagnosis making the access to these services easier, faster and available in remote areas. However this imposes new communications and data interchange challenges. In this paper a new XML based format for storing cardiac signals and related information is presented. The proposed structure encompasses data acquisition devices, patient information, data description, pathological diagnosis and waveform annotation. When compared with similar purpose formats several advantages arise. Besides the full integrated data model it may also be noted the available geographical references for e-diagnosis, the multi stream data description, the ability to handle several simultaneous devices, the possibility of independent waveform annotation and a HL7 compliant structure for common contents. These features represent an enhanced integration with existent systems and an improved flexibility for cardiac data representation.

12h30 Closing Session

13h00 – Lunch

Soup

Creme de Ervilhas com Natas / Soup of Green Peas with Cream

Main Course

Rolo de carne ao molho de tomate / Meat loaf with tomato sauce

Chinês de Legumes / Chinese Vegetables

Arroz Branco / White Rice

Dessert

Tarte de amêndoa decorada com fruta / Almond tart decorated with fruit

Inclui água, sumo de laranja, vinho e café / Includes water, orange juice, wine and coffee.

Lunch and Coffee Breaks were prepared by the students on the course in hotel management.
`http://www.eseig.ipp.pt`

XATA 2010

XML, Associated Technologies and Applications

Quick Programme

Wednesday – May 19	
10h00	Conference Registration
11h00	**Opening Session**
	Session I
11h30	Web Service for Interactive Products and Orders Configuration
12h00	Visual Programming of XSLT from examples
12h30	XML Description for Automata Manipulations
13h00	**Lunch**
	Session II
14h30	Integration of repositories in Moodle
15h00	Processing XML: a rewriting system approach
15h30	XML to paper publishing with manual intervention
16h00	**Coffee Break**
	Keynote
16h30	SAPO Service Bus
	Session III – Short Papers
17h30	Parsing XML Documents in Java Using Annotations
17h45	XML Archive for Testing: a benchmark for GuessXQ
18h00	First Day Ending
19h30	**Conference Dinner**

Thursday – May 20	
9h00	Warm-up and Certificates Delivering
	Session IV
9h30	Test::XML::Generator: Generating XML for Unit Testing
10h00	A Performance-based Approach for Processing Large XML Files in Multicore Machines
10h30	A Refactoring Model for XML Documents
11h00	**Coffee Break**
	Session V – Short Papers
11h30	XML, Annotations and Database: a Comparative Study of Metadata Definition Strategies for Frameworks
11h45	A Semantic Representation of Users Emotions when Watching Videos
12h00	SVG Integrating SVG and SMIL in DAISY DTB production to enhance the contents accessibility in the Open Library for Higher Education - Discussions and Conclusions
12h15	CardioML: Integrating Personal Cardiac Information for Ubiquous Diagnosis and Analysis
12h30	**Closing Session**
13h00	**Lunch**