XML Praktikum

Wintersemester 2001/02

Dan Olteanu, Tim Furche, Tobias Kiesling

xmlprakt@informatik.uni-muenchen.de
Contents

1 Introduction: What we want to do 3
2 The System: What we will work with 7
3 The Teams: What we can offer you 16
4 Wrap up: Planning for the future 21
Introduction: What we want to do

Why an XML tutorial?

- Hands-on experience in core XML technologies
- Looking at the XML promises
  - strengths and weaknesses
  - area of use
  - semi-structured data model
- Working on an existing project in a larger team
  - coordination of several team members as well as teams
  - use of development infrastructure
  - multi-platform development
  - commercial vs. open-source software
What we want to do

- Refine and enhance the **University Course Information System (UCIS)**
  - information about courses in an university
  - capability to manage course information for each student
  - developed last winter term by participants of the XML tutorial

- Enhancements:
  - support of aural media for output
  - increasing the managed information
  - support of schema

- Optimizations on:
  - production cycle
  - XPath
  - data model, database access layer
What you will learn (hopefully :-) )

- DSLs (domain specific languages) in XML:
  - Media-Markup: VoiceXML, WML
  - Page-Representation: APRIL
  - Programming: XSLT

- Core XML technologies:
  - Validation: XSchema, RELAX-NG, ...
  - Navigation: XPath
  - Transformation: XSLT

- XML Processing:
  - XSLT
  - XPath
  - using Java (DOM, TrAX)
  - ... and more
Software you will get to know

- XML-DBMS: Tamino (Software AG, commercial)
- XML web page production system: Cocoon (Apache Project, open-source)
- XSLT processor: Xalan (Apache Project, open-source)
- Presentation: HTML browsers, VoiceServer (IBM, commercial)
- Infrastructure:
  - versioning: CVS (open-source)
  - bug tracking: Bugzilla (Mozilla Project, open-source)
The System: What we will work with

We will present you UCIS from three perspectives:

- User view (Browser)
- Publishing view (Cocoon)
- Database view (Tamino)
User view

Course info
Person info
Personalization
Subscriptions

Navigation
Login
Publishing view

- Central tool for serving the data: **Cocoon1**
  - Java servlet
  - Separation of the production process into an arbitrary number of steps
  - Data is passed between steps as a DOM tree
  - Tutorial uses version 1.8

- Data is stored in Tamino database
Typical production process

```
<?xml version="1.0"?>
<page>
 <content>
  <xpl:query-insert>
   ...
  </content>
</page>
```

```
<?xml version="1.0"?>
<html>
 <head>
   <title>xxx</title>
 </head>
 <body>
   ...
 </body>
</html>
```
APRIL

- Abstract Page Representation Language
- Output-medium independent description of web pages
- Interface between content and presentation parts
- `<page>` element describes a page
- Three types of child elements:
  - `<content>`: contains the information to be included in the page
  - `<ui-group>`: contains possible interactions of a user with the system
  - `<meta>`: a named container for any kind of additional information
APRIL example

<?xml version="1.0" encoding="ISO-8859-1"?>

<april:page cmid="stundenplan"
xmlns:april="http://pms.informatik.uni-muenchen.de/xml/APRIL">

<april:ui-group cmid="navigation">
  <april:ui cmid="link_home" href="/"/>
  <april:ui cmid="link_vliste" href="/lehreinheiten"/>
  <april:ui cmid="link_pliste" href="/personen"/>
  <april:ui cmid="link_stundenplan" href="/stundenplan"/>
</april:ui-group>

<april:ui-group cmid="embedded_links" type="embedded">
  <april:ui cmid="vdetail" href="/vdetail">
    <april:ui-element>
      <april:location>//april:content/Lehreinheit</april:location>
      <april:param-name>id</april:param-name>
      <april:value type="XPath">@ID</april:value>
    </april:ui-element>
  </april:ui>
</april:ui-group>
Database view I

Tamino Manager

E-Business client

Browser

Internet

Tamino Server

Web Server

Web Applications

X-Node

Existing Applications

Data Sources

SQL Engine

X-Machine

SQL

Data Map

XML
Database view II
The Teams: What we can offer you

We propose the following teams:

- **Database**: Working with Tamino, modelling the data, adding schema support.
- **APRIL**: Enhancing production process using the page representation language.
- **XPath**: Finding and implementing general rules for rewriting XPath expressions.
- **Voice**: Implementing a voice dialog interface for UCIS.
Database team

Primary Goal: Enhancing the data access layer of UCIS.

- **Tamino**: Get experienced with Tamino, get an idea about strengths and weaknesses of Tamino.
- **Data model**: Improve the current data model, add more data (like literature information, change management).
- **Schema**: For validation of XML data, better schema support is needed (e.g. XML Schema, RELAX NG).

Primary Tool: Tamino, Schema Editing/Validation Tools.

Literature:

- Schema Languages @ OASIS Cover Pages: [http://www.oasis-open.org/cover/schemas.html](http://www.oasis-open.org/cover/schemas.html).
APRIL team

Primary Goal: develop new version of APRIL to solve the transformation problems
  - hard to understand, develop and maintain
  - possibly inefficient
  - dependent on embedded stylesheets that are not supported in Cocoon2

Primary Tool: Cocoon, APRIL, Xalan.

Literature:
XPath team

Primary Goal: Rule-based optimization for XPath.
- finding general rewriting rules (in)dependent on schema information
- implementing the XPath rewriter
- integrating the XPath rewriter within an existing XPath processor

Primary Tool: XPath, Java.

Literature:
- XPath 1.0 Spec.: http://www.w3.org/TR/xpath.
- Research papers on rule-based query optimizations for XML http://www.pms.informatik.uni-muenchen.de/mitarbeiter/olteanu/repository.html.
**Voice team**

**Primary Goal:** Learning to design an aural interface.
- hands-on experience with a Voice Dialog Markup Language and a Speech Grammar
- learning about the pieces necessary for a voice browser
- adding an aural interface to UCIS

**Primary Tool:** VoiceXML, IBM VoiceServer SDK.

**Literature:**
- W3C “Voice Browser” Activity: [http://www.w3.org/Voice/](http://www.w3.org/Voice/).
Wrap up: Planing for the future

Technicalities

- Main session: Thursday, 16-18:00, Z 1.09.
- Tutor sessions: Speak with your Tutor.
- Mailinglists and Groups: XMLALL, XMLDB, XMLAPRIL, XMLPATH, XMLVOICE.
- Homework: Preparation of the presentations for Session 2 and 3.
Session 1 – Development Environment

- Pieces of the environment
- Directories, testing, debugging, ...
- Versioning software CVS
- Cocoon administration: Tomcat, Cocoon, socket tunnel, ...
- Tamino administration: DB administration, interactive interface.
- Schema tools: Xerces.
- Voice tools: IBM VoiceServer SDK.
Session 2 – Presentation of the Tools

- Tamino
- Cocoon
- VoiceXML, VoiceServer
Session 3 – Presentation of the Tasks

- Changes to data model, which Schema languages to use.
- APRIL: What to do better?
- Design of an aural interface with VoiceXML.
- Ideas for XPath rewriting rules.
That’s it ...