Protocol: Software Development meeting

Presentation of eXist

Date: 13.10.2008, 13-15 pm
Participants: Robert Casties, Alexander Czmiel (BBAW), Malcolm Hyman, Wolfgang Schmidle, Markus Schöpf (BBAW), Josef Willenborg, Dirk Wintergrün
Protocol writer: Josef Willenborg

Presentation of eXist by Alexander Czmiel and Markus Schöpf

1. About BBAW (Berlin Brandenburgische Akademie der Wissenschaften – bbaw.de) and its document base:
- rapid prototyping at the beginning
- project online at telota.de
- 11.000 small XML documents (with less than 100 KB)
- a few big XML documents (about 15 MB)
- texts and images are transcribed by 1 Euro jobbers in MS-Word format
- these texts are tranformed to XML documents (TEI P5)

2. Experiences of eXist (exist.sourceforge.net):
- eXist is used since 3,5 years in BBAW
- eXist has a team of 5 main developers, head of the main developers is Wolfgang Meier
- development of desired extensions of eXist could be accelerated by money (e.g. 10000 Euro for versioning of documents in eXist)
- own extensions of eXist are not practiable because it is too difficult, but external modules could be implemented
- stability problem: different performance in different versions (sometimes some functions become slower)
- memory problems when many documents are created (each month)
- one time eXist was inconsistent but with the backup (journals) it could be repaired fast

3. Functionality of eXist:
- fultext querying: many and all desired functions
- structural querying: many and all desired functions
- document formats: only xml, no other formats like pdf, doc, etc.
- own extension for lemmatizing in indexing process is not possible to integrate
- extension of a query for all full word forms is easy programmable (with XQuery)
- Oxygene (see www.oxygene.com) could be used for editing XML documents directly in the eXist WebDAV directory
- no integrated SQL database, but an external database could be accessed by XQuery directly
- good XQuery support, extensible by own Java code (in Java modules)
- Java-Triggers are possible (e.g. on delete, update or create)
- XML documents in eXist could be based on a xsd-format but must not
- backup of eXist: a collection of XML resources in a directory

4. Performance of eXist
- Hardware: Opteron-Server, 6GB RAM, Suse Linux
- fulltext query (on fulltext index): 0,3 sec. (result with 17.000 items)
- XML query (on structural index): 2 sec. (result with 17.000 items)
- select one specific page: 1 sec., select 100 specific pages: 100 sec.
- /archimedes//s is faster as /archimedes/text/chap/p/s
- server configuration possible: how many queries are processable at once
-