



Digital Culture & Heritage Patrimoine & Culture Numérique



Haus der Kulturen der Welt, BERLIN

Aug. 31st - Sept. 2nd, 2004
31 Août - 2 septembre 2004

ZUCCARO - EIN UNIVERSELLES FRAMEWORK FÜR NETZBASIERTE FORSCHUNGSDATENBANKEN IN DEN GEISTESWISSENSCHAFTEN

Bernd Kulawik, Martin Raspe

Bibliotheca Hertziana

**Published with the sponsorship of the
French Ministry of Culture and Communication**

Actes publiés avec le soutien de la Mission de la Recherche et de la
Technologie du Ministère de la Culture et de la Communication, France

Interprétation simultanée du colloque et traduction des actes réalisées
avec le soutien de l'Agence Intergouvernementale de la Francophonie

Zusammenfassung (DE)

Mit ZUCCARO wird es erstmals möglich, auf der Basis vollständig freier und unter Open Source Lizenzen stehender Software, standardisierte, zukunftssichere Web-Datenbanken für die Forschung und Dokumentation in den historischen und geisteswissenschaftlichen Fächern kostengünstig zu erstellen, indem ein vorhandenes Modell an die jeweiligen Anforderungen eines spezifischen Projekts angepasst wird, und zugleich den Datenaustausch mit verwandten Datenbanken und so Kosten minimierenden Forschungsk Kooperationen über das Internet zu ermöglichen. Die vollständige Web-Basierung solcher Datenbanken erlaubt den Mitarbeitern nicht nur ohne die Installation einer speziellen Software (ein standardkonformer Browser auf einer beliebigen Betriebssystemplattform genügt) über das Internet von jedem beliebigen Standort aus Daten einzugeben und zu bearbeiten, sondern durch die Content Management Funktionen auch die wissenschaftliche Qualität der Daten zu sichern. Durch die Fähigkeiten zur Ein- und Ausgabe der Daten in XML wird es darüber hinaus ermöglicht, Daten schon während der Eingabe maschinell zu validieren. Die schnelle Entwicklung von ZOPE und Plone, das zuletzt durch die Gründung der Plone Foundation in enger Zusammenarbeit mit dem Software-Konzern Computer Associates (CA) auch finanziell und organisatorisch auf ein neues Niveau gehoben wurde, ermöglicht es den mit ZUCCARO realisierten Web-Datenbank-Projekten auf der Grundlage des Open Source Charakters gezielt Hilfe aus einem großen und ständig wachsenden Pool von freien Entwicklern wie auch Firmen in Anspruch zu nehmen, ohne sich in einseitige und - wie die Vergangenheit in z. T. großen Datenbank-Projekten gezeigt hat - möglicherweise fatale Abhängigkeit von einem einzelnen Unternehmen zu begeben.

Schlüsselwörter: Webbasierte Datenbank, Open Source, Geisteswissenschaften, Forschungsk Kooperationen, XML, ZOPE, Plone.

Résumé (FR)

Avec ZUCCARO il est possible pour la première fois de créer des bases de données sûres et pérennes sur le web, pour la recherche et la documentation en histoire de l'art, en ayant recours à des standards et à des logiciels *Open Source* libres. Dans un esprit et avec une exigence scientifique, le modèle peut être adapté à d'autres projets ou corpus, d'une manière qui permettra l'échange de données et la coopération à des coûts réduits.

L'adoption de technologies respectant les protocoles d'internet pour ces bases de données n'exige pas des utilisateurs d'autre logiciel qu'un navigateur web standard installé sur n'importe quelle plate-forme (Mac OS, Linux, Windows, etc.). L'usage de XML permet en outre aux collaborateurs de valider automatiquement les données pendant la saisie.

ZUCCARO bénéficie du support d'une large communauté de développeurs de logiciels libres, ce qui lui épargnera les problèmes de dépendance inhérents aux développements de grandes bases de données tributaires d'un prestataire particulier.

Name und Konzept

ZUCCARO steht als Akronym für "Zope-based Universally Configurable Classes for Academic Research Online". Damit sind mehrere Aspekte der Software und des Projektes benannt:

- _ ZUCCARO basiert auf dem freien Open Source Web Application Server ZOPE (= Z Object Publishing Environment)
- _ ZUCCARO bietet ein Set von vordefinierten Klassen zur Erfassung und Beschreibung von Objekten der "realen" Welt (im weitesten Sinne), und
- _ ZUCCARO ist besonders für den Einsatz in der akademischen Forschung zum Aufbau web-basierter Forschungsdatenbanken gedacht.

Der Name ist zugleich eine Reverenz an den italienischen Maler und Kunsttheoretiker Federico Zuccaro bzw. Zuccari, in dessen selbsterbautem Haus die Bibliotheca Hertziana seit ihrer Gründung residiert.

ZUCCARO wurde zunächst als Framework konzipiert für zwei Datenbank-Projekte, die an der Bibliotheca Hertziana seit 2001 bzw. 2002 entwickelt werden, und soll als allgemeines Informationssystem zur Kunstgeschichte Roms dienen. Das Framework selbst zielt aber auf eine allgemeinere Verwendbarkeit nicht nur an unserem Institut allein, sondern auch in anderen Bereichen der historischen Geisteswissenschaften; Anwendungen zur Dokumentation in Archiven haben wir auch ins Visier genommen.

Von Anfang an war eine Hauptforderung, dass die zu erstellenden Forschungsdatenbanken über das Internet genutzt werden können - und zwar sowohl zur Datenabfrage als auch zur Eingabe, Kommentierung und Verlinkung der Einträge, also zur Administration. Erst dadurch, sowie durch die automatische Dokumentation aller Änderungen, können Web-Datenbanken den Anspruch erheben, *Forschungsdatenbanken* zu sein. Das heißt aber auch, dass grundlegende Eigenschaften eines Content Management Systems gefordert sind:

- _ Definition differenzierter Zugriffsrechte für Benutzer mit verschiedenen Rollen, etwa anonyme Benutzer, Autoren, Redakteure, (technische) Administratoren;
- _ Beschränkung der Zugriffs- und Bearbeitungsmöglichkeiten für einzelne Datensätze oder Dokumente entsprechend der Rolle des jeweiligen Benutzers;

- _ Dokumentation der "Geschichte" eines Datensatzes und seiner Veränderungen;
- _ strukturierter Workflow für die kontrollierte Erstellung, Bearbeitung und Änderung sowie Freischaltung von Daten

Eine weitere wesentliche Anforderung an die Softwarebasis wie auch die fertige Projektsoftware ist außerdem, dass diese frei erhältlich und Open Source Software im Sinne der GNU GPL oder vergleichbarer Lizenzen sind: Nur diese Form von Software scheint uns die größtmögliche Zukunftssicherheit zu gewährleisten, weil sie unabhängig von Hardware-, Betriebssystem- und Softwarehersteller bleibt und Lizenzkosten entfallen. Dazu gehört auch, dass Daten, Strukturen und Beziehungen untereinander in einem nicht-proprietären, zukunftssicheren Format wie XML speicherbar sein sollen - die vollständige Erfüllung dieser letzten Anforderung ist noch in Arbeit.

Software-Basis

Aufgrund der genannten Anforderungen entschlossen wir uns, den freien Web Application Server Zope (= Z Object Publishing Environment) einzusetzen und an unsere Bedürfnisse anzupassen. Dabei stellten wir mehrfach fest, dass Probleme, die wir bisher nur mittels eigener Programmierarbeiten bewältigen konnten, inzwischen durch neue Module gelöst sind, die die Entwicklungsgemeinschaft mitunter für ganz andere Zwecke bereit gestellt hat

Zope enthält neben einem Webserver auch einen FTP- und einen WebDAV-Server sowie eine eigene, objekt-orientierte Datenbank samt flexibel erweiterbarer Benutzerverwaltung; es können aber auch andere, vor allem (objekt-)relationale (SQL-)Datenbanken mit Hilfe spezieller Module angebunden werden. Die Ablage der Daten erfolgt i.d.R. in einer hierarchischen Struktur aus Ordnern und darin enthaltenen Dokumenten, wie sie jedem Computerbenutzer vertraut ist. Ein Vorteil dabei ist, dass jedes Objekt, also jeder Datensatz, über eine eigene URL direkt angesprochen werden und sein Inhalt so z. B. auch von Suchmaschinenrobotern indexiert werden kann.

Zope ist fast komplett in der freien, objekt-orientierten und leicht zu erlernenden Programmiersprache Python geschrieben und bietet daher gute Voraussetzungen zur Erweiterung und Anpassung an die Bedürfnisse einzelner Projekte. Das ist allerdings auch ein Problem - denn viele Benutzer oder potentielle Anwender, gerade aus dem Bereich der historischen und Geisteswissenschaften, werden nicht bereit oder in der Lage sein, sich diese

Kenntnisse anzueignen. Andererseits verfügen solche Projekte, wie wir sie mit ZUCCARO realisieren möchten, i. d. R. nicht über so große Finanzmittel, dass man professionelle Entwickler mit der Programmierung beauftragen könnte. Deshalb erschien uns die Bereitstellung eines grundlegenden Klassen- bzw. Datenmodells, welches sich mit graphischen Hilfsmitteln anpassen läßt, am geeignetsten, um die notwendige, aber kostspielige Beratung durch Profis auf ein Minimum zu reduzieren.

ZUCCARO wird außerdem das auf ZOPE aufsetzende Web Content Management System Plone verwenden. Für dieses komplexe, modular aufgebaute Software-Pakete existiert mit den sogenannten *Archetypes* eine Möglichkeit, eigene Content-Typen [= Klassen; entsprechen in etwa Formularen zur Erfassung von Daten] zu definieren und "on-the-fly" dem Benutzer zur Verfügung zu stellen. Aber auch dieser Weg zur Erstellung neuer Content Types erfordert noch erhebliche Vorkenntnisse. Deshalb entwickelte die Firma Bluedynamics mit ArchGenXML einen Parser, der aus mit Hilfe graphischer Modellierungswerkzeuge erzeugter XMI-Dateien Archetypes-Klassen generiert, die sofort in ZOPE/Plone eingesetzt werden können. Beispiele für Modellierungstools aus dem Bereich der freien Software sind ArgoUML, Poseidon oder Eclipse) zur Erstellung von XMI-Dateien mittels UML zu definieren.

Implementierung

Am Anfang der Entwicklung stand die schwierige und infolgedessen recht langwierige Definition der Datenmodelle, die den beiden Datenbank-Projekten unseres Instituts jeweils zugrunde liegen. Auf Grundlage einer Verallgemeinerung/Abstraktion dieser Modelle entstand das nun ZUCCARO zugrundeliegende Datenmodell, das für weitere Projekte ein gemeinsames Set an Klassen bereitstellt. Diese Klassen können durch Modifikationen an den entsprechenden UML-Diagrammen den spezifischen Bedürfnissen des jeweiligen Projektes angepaßt werden.

Auf der Grundlage dieses Datenmodells sowie eines daraus entwickelten XML-Schemas entstand in gemeinsamer Entwicklung mit Mitgliedern der Plone Community ein UML-Modell. UML-Diagramme können natürlich auch dazu benutzt werden, um das ZUCCARO-Basismodell zu modifizieren.

Die aus dem UML-Diagramm generierten XMI-Dateien werden nun mit Hilfe von

ArchGenXML in Archetypes für Plone übersetzt. Die Archetypes mit ihren eigenen Interfaces stehen nach dem Kopieren der erzeugten Dateien in das entsprechende Verzeichnis des ZOPE-Servers (ohne Neustart) sofort zur Verfügung stehen.

Damit erlaubt es die Kombination dieser Tools, die sämtlich frei verfügbar sind und unter einer Open Source Lizenz stehen, Web-Datenbank-Projekte mit den für Forschungsdatenbanken unabdingbaren Content Management Fähigkeiten in kürzester Zeit zu erzeugen.

Datenaustausch und Zusammenarbeit mittels XML

An Forschungsdatenbanken muß die Forderung gestellt werden, dass ihre Daten sich ohne Verlust auf andere, neuere bzw. zukünftige Systeme migrieren lassen und dass sie die Zusammenarbeit mit anderen, inhaltlich ähnlich gelagerten Projekten erlauben.

ZOPE-basierte Systeme können prinzipiell ihre Daten untereinander austauschen und abgleichen - und zwar über alle Hardware-Plattformen hinweg und per Netzwerkverbindung. Dadurch ist z. B. möglich, dass ein Mitarbeiter eine Kopie - einen "Clone" der Datenbank - auf seinem Laptop in einem Archiv ohne Netzanbindung benutzt und die neuen Daten anschließend in die Originaldatenbank überspielt bzw. von dort inzwischen neu eingegebene Daten übernimmt. Für ZUCCARO stand aber von vornherein - vor allem, aber nicht nur mit Blick auf die Zukunftssicherheit der Daten - fest, dass die Daten samt ihrer Strukturierung und den internen Beziehungen auch unabhängig von ZOPE lesbar sein müssen. Als Format hierfür kommt im Hinblick auf die aktuelle Entwicklung der Webstandards nur XML in Frage. Ein Ziel von ZUCCARO ist es daher, sämtliche Daten in einem standardisierten XML-Format ausgeben und einlesen zu können, die dem XMI-Schema des UML-Modells entspricht.

Dadurch wird es zugleich möglich, dass Datenbank-Projekte, die sich für eine gänzlich anders geartete Software-Basis entschieden haben (z.B. objekt-relationale Datenbanken mit Internetanbindung durch einen eigenen Webserver), ZUCCARO-basierte, daher auch ZUCCARO-konforme und also untereinander kompatible Datenbanken aufbauen können, deren Daten mit anderen, ZUCCARO-basierten Lösungen austauschbar sind. Auf diese Weise wird nicht nur die zukünftige Migration auf andere Softwaresystem für eine Zeit "nach ZOPE" von vornherein gewährleistet, sondern auch die Möglichkeit zum Datenaustausch

zwischen verschiedenen Projekten geschaffen: So sollen unsere Datenbanken zur Architekturzeichnung des Barock und zur frühbarocken Malerei Roms gemeinsam an einem Datenpool relevanter Personen, Bau- und Kunstwerke oder Körperschaften sowie Literatur und Archivalien teilhaben und vermeiden so zeit-, kosten- und personalintensive Doppeleingaben.

In ähnlicher Weise können wir uns eine Kollaboration zwischen ZUCCARO-basierten Datenbank-Projekten über das Internet vorstellen.

Da aber jedes Objekt - jeder Datensatz - in einer ZOPE-Datenbank über einen URL anzusprechen und abzurufen ist, ist es nicht einmal notwendig, dass miteinander zusammen arbeitende Projekte ihre Daten untereinander physisch - etwa durch das Versenden von XML-Dateien - austauschen, sondern es genügt, auf die interessierenden Datensätze der einen Datenbank per Hyperlinks in den Datensätzen der anderen zu verweisen. Durch die Weiterentwicklung des bisher nur innerhalb einer ZOPE-Anwendung arbeitenden Link- bzw. Relations Manager dürfte es in naher Zukunft sogar möglich sein, die Validität solcher Links regelmäßig zu prüfen und z.B. das Löschen von Objekten, die noch mit anderen verlinkt sind - sicherlich die wohl häufigste Ursache für "404"-Anzeigen im Internet - zu verhindern.