



In: Bekavac, Bernard; Herget, Josef; Rittberger, Marc (Hg.): Informationen zwischen Kultur und Marktwirtschaft. Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004), Chur, 6.-8. Oktober 2004. Konstanz: UVK Verlagsgesellschaft mbH, 2004. S. 481 – 483

## **Xylix Software Suite – "virtual archives" im Archiv für Zeitgeschichte**

*Daniel Nerlich*

Archiv für Zeitgeschichte [www.afz.ethz.ch](http://www.afz.ethz.ch)  
ETH-Zentrum  
8092 Zürich  
[nerlich@history.gess.ethz.ch](mailto:nerlich@history.gess.ethz.ch)

Das Archiv für Zeitgeschichte der ETH Zürich hat mit der Firma Xylix Software erste Schritte zur Umsetzung seiner Vision "virtual archives" realisiert. Optimierte für kleine und mittlere Unternehmen werden Technologien wie Grid Computing, konsequente Verwendung von XML zur Speicherung der Daten, sowie MySQL als Datenbank-Server eingesetzt.

Verfügbar ist nun eine Software mit äußerst geringem Ressourcenverbrauch: Hierzu gehört Xylix Retrieval System, ein leistungsfähiges probabilistisches Suchsystem, das beim getesteten Betrieb von mehreren Kollektionen mit über 70000 Dokumenten lediglich 1.5 Mb RAM belegt. Dies erlaubt es, dass das System flexibel nach Bedarf sowohl auf dem einfachen PC als auch auf Supercomputern eingesetzt werden kann. Zur Indexierung lassen sich mehrere Computer zu einem Rechner-Grid zusammenschließen.

Im Unterschied zum üblichen Verfahren, den Index möglichst im Arbeitsspeicher resident zu halten, wird eine Technik genutzt, die unter Verwendung von Memory Mapping einen äußerst schnellen Zugriff auf den auf der Harddisk abgelegten Index gestattet. Somit entfällt die Anschaffung eines speziellen Motherboard mit teurem Arbeitsspeicher, ohne dass Abstriche bezüglich Performance des Systems gemacht werden müssen.

Im System implementiert sind u.a. Wortsegmentierung, Stoppwort-Elimination, Treffer-Highlighting und Parsing verschiedener Dokumentenformate. Der Index ist als Suffix Tree aufgebaut und unterstützt auch Anfragen von Teilbegriffen.

Intranet und spezifisch zusammengestellte Kollektionen von Dokumenten aus dem Internet können so kostengünstig, intuitiv und schnell abfragbar gemacht werden.



Dieses Dokument wird unter folgender [creative commons](http://creativecommons.org/licenses/by-nc-sa/2.0/de/) Lizenz veröffentlicht:

<http://creativecommons.org/licenses/by-nc-sa/2.0/de/>

### Anwendungsbeispiel: Xylix Online Archives – Information Retrieval für archivische Metadaten

Mit Xylix Online Archives werden die nach errechneter Relevanz sortierten Suchergebnisse des Information Retrieval Systems stets im Kontext der Tektonik dargestellt. Dies eröffnet für die Archivwelt eine neue Dimension: Einzeltreffer werden nicht mehr isoliert, sondern in Verbindung mit der systematischen Bestandes-Gliederung angezeigt. Es stehen dem Benutzenden somit beide Recherchemöglichkeiten kombiniert zur Verfügung.

Die Daten müssen nicht aus der Datenbank exportiert und in einem zusätzlichen Arbeitsprozess für die Darstellung auf dem Internet aufbereitet werden, sondern lassen sich direkt *on the fly* aus der zentralen Archivdatenbank generieren. Um eine bestmögliche Sicherheit und Vertraulichkeit zu gewährleisten, sind Sperrmöglichkeiten auf mehreren Ebenen eingebaut worden: auf Bestandes- und Dossierebene sowie durch die Anmeldung mit Passwörtern.

Xylix Online Archives präsentiert die in einer Oracle-basierten zentralen Datenbank erfassten Metadaten zum Archivgut des Archivs für Zeitgeschichte in einem einfach zu bedienenden *Webfrontend*. Einschlägige Akten sowie auch audiovisuelle Quellenmaterialien können wahlweise über einen hierarchischen Baum per Mausklick *ermittelt* und/oder via Suchabfrage *recherchiert* sowie über die Funktionalität der Bookmarklist zu individuellen Rechercheergebnissen zusammengestellt werden. Soweit das Archivgut bereits in digitaler Form vorliegt und freigegeben ist, kann es über Xylix Online Archives angesteuert und eingesehen werden. Damit wird der virtuelle Lesesaal Realität. Darüber hinaus ermöglicht das oben beschriebene Xylix Retrieval System, ausgewählte Webinhalte zu indexieren und so ein "Webportal für zeitgeschichtlich Interessierte" aufzubauen.

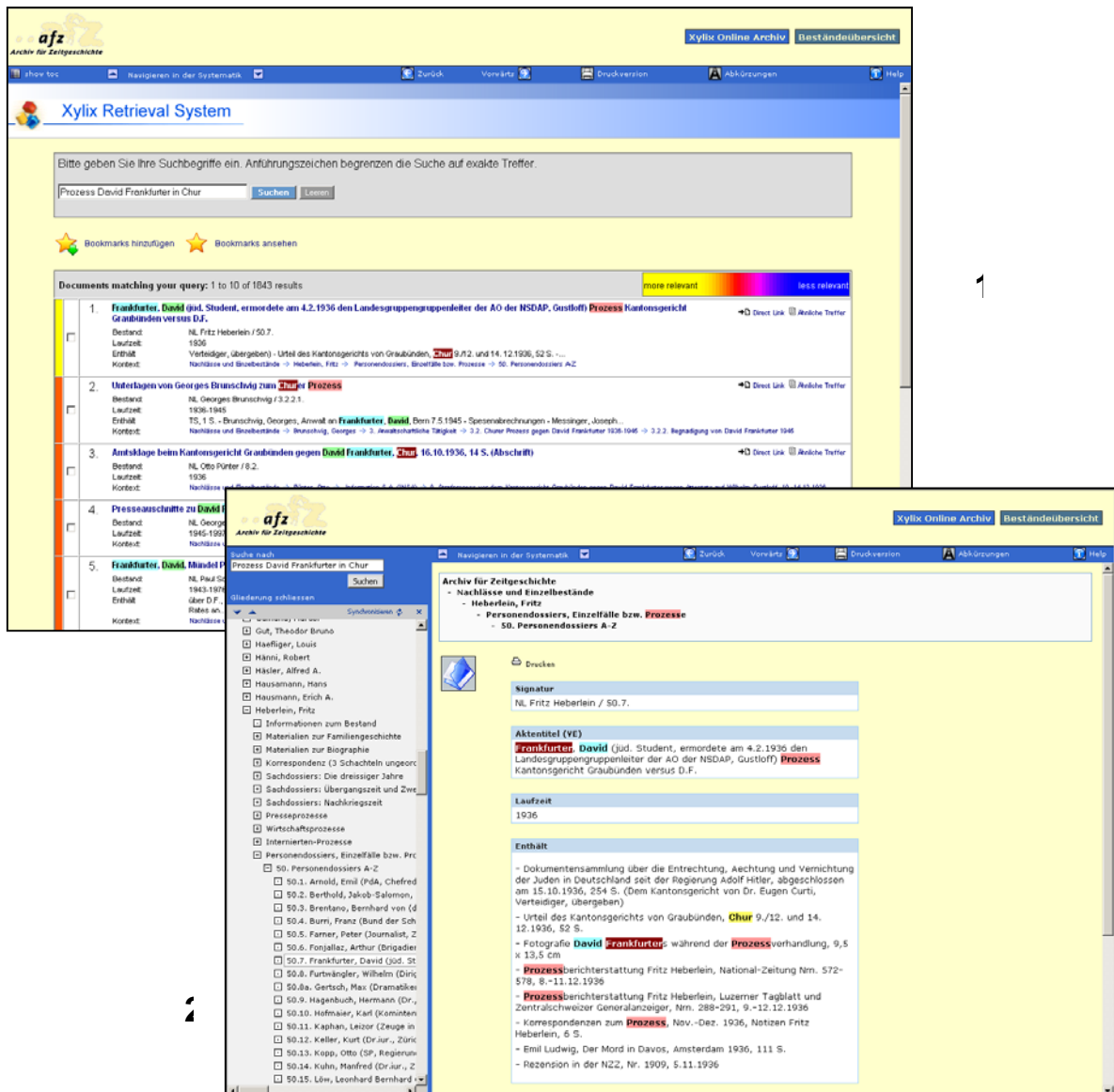


Abb. 1: Benutzeroberfläche Xylix Online Archives. Vernetzte Suchstrategien: Nach Relevanz geordnete, interaktive Trefferliste mit den hervorgehobenen Suchbegriffen (1) und Anzeige eines Einzeltreffers mit dem systematischen Kontext der Tektonik (2), die ein Weitersuchen in der systematischen Gliederung erlaubt.

Mit ihren vielfältigen Anwendungsmöglichkeiten leistet die Xylix Software Suite einen innovativen Beitrag zur Nutzung moderner Informationstechnologien. Die Aufschaltung der über Internet verfügbaren Version ist für Anfang 2005 geplant, während die Intranet-Version bereits im Einsatz ist.

