LEAF – Linking and Exploring Authority Files

http://www.leaf-eu.org

LEAF wird gefördert innerhalb des Fünften Rahmenprogramms (1998–2002) der Europäischen Kommission, im Programm „Information Society Technologies“ (IST).

Das Projekt hat eine Laufzeit von 36 Monaten und startete im März 2001. Die Projektkoordination liegt bei der Staatsbibliothek zu Berlin, Teilnehmer sind 15 Archive, Bibliotheken, Dokumentationsstellen sowie Institute für Informationssysteme aus Deutschland, Frankreich, Großbritannien, Norwegen, Österreich, Portugal, Schweden, der Schweiz, Slowenien und Spanien.

Hintergrund

Durch die Bereitstellung von Bestandskatalogen im Internet heben sich aus Sicht der Benutzer zunehmend die Grenzen zwischen verschiedenen „Gedächtnisinstitutionen“ (Bibliotheken, Archive, Museen usw.) mit unterschiedlichen Erschließungs- und Katalogisierungstraditionen auf. Zugleich vervielfältigen sich durch die elektronischen Kataloge (OPACs) die Suchmöglichkeiten, da potentiell alle Elemente des elektronischen Katalogs als Sucheinstiege dienen können. Für die Benutzer ergeben sich gleichzeitig zwei Schwierigkeiten: zunächst nämlich müssen überhaupt Treffer erzielt werden, sodann muss gegebenenfalls die erzielte Treffermenge sinnvoll eingegrenzt werden können. Namen stellen das am häufigsten gewählte Suchargument dar, somit sind biographische Normdaten besonders gut geeignet, die Suchstrategien zu optimieren. In Personennormdateien werden Namen bestimmten Personen zugeordnet und mit anderen Namensformen dieser Person verknüpft, also etwa frühere und spätere Namen, Pseudonyme, fremdsprachige Schreibweisen, Transliterationen usw. Zudem enthalten Normdaten zusätzliche Informationen, wie Geburts- und Sterbedaten, Länderzugehörigkeit, Beruf usw., die geeignet sind, eine eindeutige Unterscheidung von anderen Personen gleichen Namens zu ermöglichen.

Normdateien sind zunächst konzipiert als ein Arbeitsinstrument in Bibliotheken für die Katalogisierung. Sie gewährleisten die Konsistenz der Kataloge und sind Voraussetzung für effiziente Datenerfassung in Bibliotheksverbünden. Auch in anderen Institutionen (Archiven, Museen usw.) gewinnen Normdateien wegen der Überführung der Findhilfen und Zettelkataloge in elektronische OPACs zunehmend an Bedeutung.

Während im Bibliothekssektor auf nationaler Ebene mehrere Namennormdateien nebeneinander bestehen und derzeit Versuche unternommen werden, diese miteinander zu verbinden, gibt es momentan keine Namennormdatei, die gleichermaßen von Bibliotheken, Archiven, Museen und Dokumentationszentren genutzt wird. Auch ist derzeit kein standardisiertes europäisches Format für Normdaten verfügbar. Die Nutzung existierender Normdateien ist im allgemeinen größeren Organisationen vorbehalten sowie solchen, die in Normdatenprojekte involviert sind oder waren. Kleinere Institutionen benutzen lokale Dateien, die speziell auf deren Bedürfnisse zugeschnitten sind und deren Inhalte im allgemeinen anderen Institutionen nicht zugänglich sind. Da die Normdatenerstellung den teuersten Aspekt der Katalogisierung darstellt, ist eine möglichst breite Zusammenarbeit aus der Sicht der Institutionen ein anstrebenswertes Ziel.

Ein beträchtliches Maß an Forschungs- und Entwicklungsarbeit wurde bisher im Bereich verteilter virtueller Katalogsysteme unternommen, jedoch wurde nicht versucht, die Daten im Hinblick auf einen einheitlichen Sucheinstieg zu verlinken. LEAF bietet einen völlig neuen Lösungsansatz für dieses Problem.

Projektziele

In LEAF wird eine Modellarchitektur für ein zentrales System entwickelt, in dem existierende Normdaten zu Personen und Körperschaften automatisch miteinander verlinkt werden.

Die Projektziele sind im wesentlichen:

- Verbesserung des Zugriffs auf Normdateninformation für alle involvierten Benutzergruppen.
- Verbesserung der Qualität von Normdateien.
- Aufbau eines „Zentralen Europäischen Normdatenpools“, der durch konkrete Benutzersuchabfragen generiert wird.
- Verbesserung der Suchmöglichkeiten in Online-Katalogen und anderen Netzapplikationen.
- Bereitstellung zusätzlicher (biographischer usw.) Informationen zu den Personen bzw. Körperschaften des Normdatenpools.

Lösungsansatz

Das LEAF-System wird eine zentrale Datenbank enthalten, in welche die lokalen Normdaten der LEAF-Partner geladen werden können. Die geplante Hybrid-Architektur erlaubt das Einbinden zusätzlicher Online-Datenquellen. Die angestrebte Lösung basiert auf modernen Internettechnologien (Applicationserver, Java, XML, XSLT).

Im LEAF-System werden die in unterschiedlichen Formaten (MAB2, MARC21, UNIMARC usw.) vorliegenden Daten in das XML-Format EAC konvertiert. EAC (Encoded Archival Context) wird gegenwärtig in einem internationalen Projekt, an dem auch LEAF-Partner beteiligt sind, entwickelt und wird als Standardformat für den Austausch und die Verlinkung von Normdaten zwischen Bibliotheken, Archiven und vergleichbaren Institutionen dienen. EAC erlaubt die Einbindung sowohl bibliothekarischer als auch archivalischer Normdaten.

Regelmäßige Updates werden gewährleisten, dass die Daten im zentralen LEAF-System stets aktuell sind. Im zentralen LEAF-System werden jene Datensätze, die dieselbe Person bzw. Körperschaft beschreiben, auf der Basis festgelegter Kriterien automatisch miteinander verlinkt. Dem Vergleich der Normdatensätze werden spezifische, im LEAF-Projekt zu definierende Kriterien zugrundegelegt. Jeder automatisch generierte Link kann manuell gelöscht werden, indem ein „negativer Link“ zwischen den Datensätzen generiert wird, der Priorität gegenüber den automatisch erzeugten Links hat. Zusätzlich können auch „positive“ Links manuell generiert werden.

Registrierte Benutzer des LEAF-Systems werden die Möglichkeit haben, den LEAF-Normdatensätzen Anmerkungen und Korrekturen hinzuzufügen. Es wird drei Klassen von Annotationen geben: erstens Korrekturen oder Ergänzungen zu einzelnen oder zu verlinkten Datensätzen. Ein Benutzer könnte z.B. ein fehlendes Geburtdatum ergänzen oder ein Sterbedatum korrigieren. Diese Korrekturen fließen selbstverständlich nicht direkt in die Datensätze ein. Vielmehr wird die Anmerkung an die Urheber-Institution des Datensatzes weitergeleitet. Falls die Anmerkung relevant ist, wird sie in den jeweiligen lokalen Datensatz eingearbeitet und wird mit dem nächsten LEAF-Update automatisch in den entsprechenden LEAF-Datensatz übernommen. Da nur registrierte Benutzer Anmerkungen vornehmen können, ist es für die Datenanbieter möglich, mit dem Benutzer Kontakt wegen etwaiger Rückfragen aufzunehmen.

Die zweite Klasse von Annotationen betrifft die Verlinkung von Datensätzen. Bemerkt ein registrierter Benutzer, dass Datensätze vom System fehlerhaft verlinkt wurden, kann er dies annotieren. Die Anmerkung wird überprüft und der Link manuell entfernt oder verändert.

Während die zwei genannten Klassen von Annotationen temporär sind und nach Bearbeitung gelöscht werden, besteht zusätzlich die Möglichkeit, den Daten persistente Annotationen hinzuzufügen. Eine kleinere Institution könnte so etwa auf eigene Bestände zu einer betreffenden Person hinweisen.

Das LEAF-System wird die technischen Voraussetzungen zur Verfügung stellen, um die gefundenen Normdatensätze aus dem LEAF-Format in ein Datenformat zu konvertieren, das in einem lokalen System weiterverwendet werden kann, also etwa in MAB2, UNIMARC, MARC21 oder EAD. Datensätze können in die lokalen Systeme der Benutzer exportiert werden. Ferner können registrierte Benutzer die Suchergebnisse in einem „LEAF User Workspace“ speichern und dort manuell weiterbearbeiten.

Schließlich können Benutzer das LEAF-Suchergebnis als Suchargument in anderen Systemen benutzen. Da Namen der wichtigste Sucheinstieg für verschiedene Netzapplikationen sind, werden sich auch andere externe Services mit LEAF verlinken können und so von dem Sucheinstieg über gemeinsam genutzte europäische Normdaten profitieren.

Als Beispiel wird eine Integration mit der MALVINE-Plattform (Manuscripts and Letters via Integrated Networks in Europe, www.malvine.org) durchgeführt, die eine verteilte Suche nach Informationen zu modernen Nachlässen und Autographen ermöglicht.

Projektpartner

Staatsbibliothek zu Berlin, Berlin (Koordinator)
Biblioteca de Universidad Complutense, Madrid
Biblioteca Nacional, Lissabon
Crossnet Systems Limited, Newbury
Institut Mémoires de l'Édition Contemporaine, Paris
Joanneum Research – Institut für Informationssysteme und Informationsmanagement, Graz
Narodna in Univerzitetna Knjiznica (National- und Universitätsbibliothek), Ljubljana
Österreichische Nationalbibliothek, Wien
Riksarkivet (Staatsarchiv), Stockholm
Schweizerische Landesbibliothek, Bern
Universitetet i Bergen – HIT-senteret (Universität Bergen – The Humanities Information Technology Research Programme), Bergen

Assoziierte Partner

British Library, London
Deutsches Literaturarchiv, Marbach am Neckar
Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie, Graz
Goethe- und Schiller-Archiv, Weimar