Schlussbericht zu Nr. 3.2 BNBest-BMBF 98

Zuwendungsempfänger: Universität Duisburg-Essen Fakultät für Ingenieurwissenschaften	Förderkennzeichen: **01C5966**

Vorhabensbezeichnung:

WIKINGER – Wiki Next Generation Enhanced Repository

Laufzeit:
01.10.2005 bis 31.12.2008

Berichtszeitraum:
01.10.2005 bis 31.12.2008

I. Kurze Darstellung zu

1. Aufgabenstellung

Ziel des Projekts WIKINGER war es, eine Plattform zu schaffen, die es Wissenschaftlern ermöglichen sollte, effizient und ortsunabhängig in Wissensbasen ihres jeweiligen Fachgebiets zu recherchieren und kollaborativ über das Internet neues Wissen zu generieren. Im Detail wurden Verfahren zur automatischen Eigennamenerkennung und zur semi-automatischen Erstellung von semantischen Netzen entwickelt, um Eigennamen (Personen-, Zeit-, Ortsangaben und ggf. Namen historischer Ereignisse) aus einer umfangreichen textuellen Datenkollektion zu extrahieren und diese computerunterstützt zu einem semantischen Netz zu verknüpfen. Bei der Generierung des Netzes wurden Techniken des Semantic Web eingesetzt, um eine einfache Verbreitung und Nachnutzung des gesammelten Wissens über das Internet in Zukunft zu ermöglichen.

Als exemplarischer Forschungsbereich wurde der von der Kommission für Zeitgeschichte[1] vertretene Fachbereich zeitgeschichtlicher Katholizismusforschung gewählt, um die Verfahren auf einer begrenzten Domäne erproben zu können.

[1] Die Kommission für Zeitgeschichte ist die zentrale Anlaufstelle für zeitgeschichtliche Katholizismusforschung in Deutschland. Sie hat seit 1962 in doppelter Hinsicht ein „Verbundsystem" und eine deutschland- und europaweite Vernetzung mit befreundeten Einrichtungen entwickelt: a) Zwischen Archiv, Bibliothek und Forschung sowie b) durch die Verbindung von Forschungsstelle und Wissenschaftlicher Kommission. Den ursprünglichen Forschungsschwerpunkt „Drittes Reich" hat die Kommission bis heute beibehalten und neue Aufgaben zusätzlich übernommen: Forschungen zur Geschichte der Bundesrepublik Deutschland, zum Katholizismus in SBZ und DDR und zum Katholizismus in Europa. In einer eigenen, aktuell (Juni 2009) 167 Bände umfassenden wissenschaftlichen Reihe – den Veröffentlichungen der Kommission für Zeitgeschichte – werden die Forschungsergebnisse publiziert. Die Forschungen konzentrieren sich ergänzend auch auf tages- oder geschichtspolitisch aktuelle Brennpunktthemen, wie z.B. das Thema „Katholischer Kirche und Holocaust" oder „Katholischer Kirche und Zwangsarbeiter". Vgl. insgesamt: http://www.kfzg.de.

Die Präsentation und die kollaborative Weiterverarbeitung der extrahierten und vernetzen Daten kann nun künftig über ein Wiki-System erfolgen, das allen autorisierten Nutzern des Wissenschaftsbereichs zur Verfügung steht.

Die Einzelziele des Projektes und dessen Verwertung lassen sich in folgender Übersicht zusammenfassen:

- Einsatz und Weiterentwicklung von Verfahren zur Extraktion Entitäten bestimmter semantischer Kategorien, wie z.B. Personen, Orten, Historischen Ereignissen, Daten oder Zeitangaben mittels Eigennamenerkennung (Named Entity Recognition)
- Entwicklung von halbautomatischen Verfahren zum Aufbau semantischer Netze zu Persönlichkeiten der modernen Kirchengeschichte mittels Techniken des Semantic Web oder mittels Topic Maps
- Aufbau und Weiterentwicklung einer kollaborativen Wissensplattform auf der Basis von Wiki-Systemen zur wissenschaftlichen Kommunikation, Recherche und verteilten Wissensgenerierung, um anschließend ein „Biographisch-bibliographisches Handbuch für das katholische Deutschland" erstellen zu können.
- Verbreitung und Nutzung der WIKINGER-Plattform innerhalb eines ausgesuchten Personenkreises zur Steigerung des Vernetzungspotentiales in dem genannten Forschungsbereich.

Die **Universität Duisburg-Essen (UDE)** entwickelte die Arbeitsumgebung WALU (Wikinger Annotations- und Lernumgebung), die es Fachwissenschaftlern komfortabel gestattet, Texte mit farblich markierten Annotationen zu versehen. Diese sog. Trainingsdaten bilden die Ausgangsbasis für Algorithmen des maschinellen Lernens, die dann die relevanten Kontexthinweise für die Entitätsklassen ermitteln. Hieraus entsteht ein Modell, das auf Rohtexte angewandt werden kann und dann die Annotationen automatisch erzeugt.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

a. Datenbasis und Vorarbeiten

Als primäre Datenbasis dienten zunächst die biografischen Informationen, die aus den »Veröffentlichungen der Kommission für Zeitgeschichte« (»Blaue Reihe«) gewonnen werden konnten.
In der »Blauen Reihe« sind zahlreiche wissenschaftliche Editionen zeitgenössischer Quellen (Reihe A) sowie wichtige sozial- und mentalitätsgeschichtliche Einzelstudien zur Geschichte des deutschen Katholizismus im 19. und 20. Jahrhundert (Reihe B) erschienen, die die Analyse und historische Beschreibung der durch Subkultur (Konfession, Glaubens- und Frömmigkeitsleben) und Substruktur (Vereine, Presse, Parteien, Führungseliten) gekennzeichneten katholischen Gesellschaftsgruppe ins Zentrum stellen, die sich in Deutschland seit der