

# Neue Formen der Informationsvisualisierung, Navigation und kontextbezogener Textsuche für Bibliotheken und Archive

Kai Uwe Barthel<sup>1</sup>, Dirk Neumann<sup>1</sup>, Adriane Feustel<sup>2</sup>, Anja Thieme<sup>2</sup>

<sup>1</sup>) HTW Berlin, Wilhelminenhofstr. 75a, D-12459 Berlin

<sup>2</sup>) Alice Salomon Archiv der ASH Berlin, Karl-Schrader-Str. 7-8, D-10781 Berlin

barthel@htw-berlin.de

feustel@ash-berlin.eu

## Zusammenfassung:

INSAS ist der Kurztitel eines informations- und sozialwissenschaftlichen Forschungsprojektes, das die Hochschule für Technik und Wirtschaft und die Alice Salomon Hochschule gemeinsam in Kooperation mit regionalen Partnern aus der Wirtschaft und öffentlichen und privaten Informationsdienstleistern durchführt. Auf der Text- und Dokumentengrundlage des Alice Salomon Archivs der ASH Berlin wird eine neuartige kontextbezogene Suche, Navigation und Visualisierung von Informationen in Form eines "lernfähigen" Informationssystems entwickelt. Mit dem entwickelten System können Dokumente nicht nur präsentiert, sondern auch je nach Anfrage dekonstruiert und in neue Zusammenhänge gebracht werden. Als Prototyp ist eine Anwendungssoftware entstanden, die für alle Einrichtungen, Unternehmen und Einzelpersonen, die mit Text (und Bild) im Kontext arbeiten, geeignet ist und sie in ihrer Recherche erheblich unterstützen wird.

## Abstract:

INSAS is the short title of an information technology and social science research project from the University of Applied Sciences Berlin (HTW) and the Alice Salomon Hochschule Berlin (ASH). This project is conducted in close collaboration with regional partners from public and private information providers. Based on the texts and documents from the Alice Salomon Archive of the ASH Berlin a novel context-sensitive search system is developed. This system will allow an adaptive navigation and visualization of information. The system automatically presents suiting contexts according to search results. A prototype application has been developed, which is suitable for all institutions, companies and individuals who work with text (and images), and it is suitable to support considerably in their research.

## Problemstellung

Durch das Internet ist es auf einfache Weise möglich geworden, u.a. wissenschaftliche Texte einer großen Zahl von Nutzern zur Verfügung zu stellen. Die ursprüngliche Nutzung des Internets beschränkte sich meist auf ein einfaches „Anbieten und Abrufen der Informationen. Mit Techniken des „Web 2.0“ wie Blogs, Foren und Wikis wurde es den Nutzern möglich, selbst neue Inhalte zu kreieren bzw. bestehende Inhalte zu kommentieren oder zu verändern. Als Beispiel sei hier die Wikipedia-Online-Enzyklopädie genannt, die seit 2001 ständig gewachsen ist und inzwischen in Deutschland über mehr als 1,2 Millionen Artikel verfügt. Ein Problem der heutigen Internet-Recherche besteht jedoch darin, dass die großen Mengen von Informationen, die zu bestimmten Themen existieren, häufig nur schwer zu finden sind und dass vor allem keine semantische Einbettung existiert. Dies Problem wird durch die übliche Hypertextstruktur noch verstärkt. Die Verlinkungen zu weiteren Texten führen häufig dazu, den ursprünglichen Kontext zu verlieren.

Alice Salomon hat ein schriftliches Werk hinterlassen, das sich nicht nur durch seinen Umfang, sondern durch seine Vielfältigkeit und Komplexität auszeichnet. Wissenschaftlich und methodisch hat sie neue Wege beschritten. Ihre Schriften in ihren biographischen, zeithistorischen, sozial- und frauenpolitischen, philosophisch-wissenschaftlichen, religiös-ethischen, internationalen und anderen Kontexten online anschaulich zu zeigen, übersteigt die heute weitestgehend gängigen und möglichen Präsentationsformen. Die internetbasierten Recherche- und Forschungsmöglich-

keiten, die zunehmend an Gewicht gewinnen sind zu eingeschränkt und führen nicht selten zu reduzierten bis hin zu falschen Ergebnissen und Sichtweisen, da komplexe Zusammenhänge nicht erfasst werden können.

### **Forschungsstand**

Die Verknüpfung von semantischen Netzen mit neuen Visualisierungstechniken ist noch sehr wenig erforscht. Während das Auffinden von einzelnen Artikeln meist durch eine „geschickte“ Nutzung von Suchsystemen noch zu lösen ist, besteht ein großes Problem darin, dass es bis jetzt - bedingt durch die Natur des Internets - nicht möglich ist, einen kontextuellen Bezug eines bestimmten Beitrags zu anderen ähnlichen Artikeln bzw. sinnverwandten Themen zu erhalten. Wird versucht ähnliche Artikel mit Internet-Suchsystemen zu finden, so erfolgt dies typischerweise durch eine Anfrage mit Schlagwörtern. Werden nur wenige Schlagwörter verwendet, so ist die Ergebnismenge meist extrem groß, wobei viele der gelieferten Ergebnisse thematisch nicht zur Suchintention passen. Bei sehr präzisen Anfragen mit vielen Schlagwörtern werden meist nur sehr wenige Ergebnisse geliefert, unter denen sich die gesuchten Artikel u. U. nicht befinden. Das gleiche Problem besteht auch bei klassischen Datenbankanfragen. Gegenwärtig wird eine visuelle Repräsentation von Suchergebnissen nur bei Systemen eingesetzt, die nach visuell ähnlichen Bildern oder Produkten (z.B. [www.like.com](http://www.like.com)) suchen. Systeme, die eine visuelle und kontextuelle Navigation innerhalb der Suchergebnisse ermöglichen, existieren noch nicht.

Die vorhandenen Forschungen zu Alice Salomon ermöglichen, die Relevanz der Materialien und deren Kontext zu beurteilen und die Fehlstellen/Forschungsdefizite zu umreißen, die mit der zu entwickelnden Software deutlicher benannt werden können und zu deren Schließung Ansätze aufgezeigt werden sollen.

### **Methodisches Vorgehen**

Techniken des Information Retrieval machen es möglich, auch komplexere Ähnlichkeitsbeziehungen zwischen unterschiedlichen Dokumenten zu modellieren. Was global für alle Dokumente des Internets aufgrund zu vieler unterschiedlicher thematischer Bereiche nur schwer zu handhaben ist, kann in einem begrenzten Themenbereich durchaus gelingen und für eine neuartige Präsentationsform genutzt werden, die zugleich komplexe Recherchen unterstützt. Anstatt Basistexte und kontextrelevante Dokumente des Archivs in langen Listen aufzuführen, wurde ein Prototyp eines internetbasierten Suchsystems konzipiert, der die „Suchenden“ bei einer unscharfen Suche mit einer visuellen Navigation mit Kontextbezug unterstützt. Das Projekt greift als exemplarisches Material auf die Sammlungen des Alice Salomon Archivs zu Person und Werk Alice Salomons und die dreibändige annotierte Edition der ausgewählten Schriften (Luchterhand Verlag 1997-2004) zurück, mit denen Grundlagen und Methoden zur Kontextualisierung gegeben sind. Die Sammlungen enthalten neben den Schriften umfangreiche Materialien zu deren Erläuterung, wie Gesetzestexte, Verordnungen, Statistiken, Texte anderer Autoren (Romane, Gedichte, wissenschaftliche Texte), biografische Materialien, Fotos, Briefe u.v.a.m. Auf dieser Basis werden für den Prototyp geeignete Materialien ausgewählt und nach wissenschaftlichen Kriterien in Digitalisate umgesetzt. Das schließt die Lösung dabei auftretender Probleme ein - wie auch die Entwicklung von Konzepten zu Recherchierbarkeit und Visualisierung.

Das neue Suchsystem visualisiert mit einem ausgewählten Textkorpus beispielhaft Komplexität und Kontexte und macht diese erkennbar. Die Idee hierbei besteht darin, für all diese verschiedenartigen Dokumente und Medien anhand unterschiedlicher „Dimensionen“ wie Dokumentenart, Erscheinungsdatum, den thematischen Ausrichtungen, Schlagwörtern etc. die semantischen Beziehungen und Ähnlichkeiten bzw. den thematischen Kontext zu modellieren. Hierdurch entsteht ein semantisches Netzwerk, das redaktionell sicherlich korrigiert und ergänzt werden muss und soll. Dieses Netzwerk stellt die Grundlage für das neue Informationssystem dar. Mit modernen Visualisierungstechniken und Techniken des kollaborativen Filterns wurde ein Prototyp eines Informationssystems geschaffen, der eine visuelle Erkundung, Navigation und Recherche ermöglicht. Aufgrund der hochdimensionalen Vernetzungen werden die Suchenden aktiv in den Such- und Visualisierungsprozess eingreifen müssen, um zu den gewünschten

Ergebnissen zu gelangen. Ziel ist es hierbei immer, den Suchenden den Bezug zum jeweiligen Focus und Kontext zu präsentieren und zu neuen Fragestellungen anzuregen. Durch den Einsatz kollaborativer Techniken ist das Suchsystem so konzipiert, dass die Verknüpfungen und somit die Suchergebnisse im Laufe der Nutzung immer besser werden. Das bedeutet, dass die zunächst von den Projektmitarbeitern erstellten und geprüften Dokumentrelationen kontinuierlich und semi-automatisch durch die Nutzer ergänzt werden. Die Verwendung von Metadatenstandards wird die Anbindung an andere Systeme/Programme ermöglichen.

### Wissenschaftliche Bedeutung

Das Projekt ist noch nicht abgeschlossen, jedoch liegen bereits jetzt sehr vielversprechende Ergebnisse vor. In informationswissenschaftlicher Hinsicht werden neue Erkenntnisse über semantische verknüpfte Netzwerke und die Konzeption geeigneter Visualisierungsformen derartiger Netzwerke erwartet. Hierbei werden geeignete Featureextraktionsverfahren und entsprechende Metriken ermittelt, mit denen komplexe Suchverfahren und eine semi-automatische Generierung von semantischen Beziehungen zwischen den Dokumenten eines Archivbestandes möglich werden. Weiterhin ist mit interessanten Ergebnissen zur Konzeption und Gestaltung einer visuellen, kontextuellen Navigation in semantischen Netzen zu rechnen. Neuartig ist der Versuch, Dokumente aufgrund einer äußeren Struktur und beschreibender Kriterien automatisch einer Dokumentenart zuweisen zu lassen. In Bezug auf Leben und Werk Alice Salomons werden neue Erkenntnisse über Kontexte ihres Schaffens und Denkens in ausgewählten Themenbereichen erwartet. Dank der Zusammenarbeit mit regionalen Partnern werden neue Erkenntnisse möglich, hinsichtlich persönlicher Kooperationen und regionaler Bezüge sowohl im Bezirk Schöneberg, als auch im gesamtstädtischen Bereich von Berlin und nicht zuletzt auf überregionaler und internationaler Ebene, über die bislang nur wenig bekannt ist. Abbildung 1 zeigt eine erste Version des neuen Suchsystems, in dem verwandte Dokumente im Kontext mit dargestellt werden.

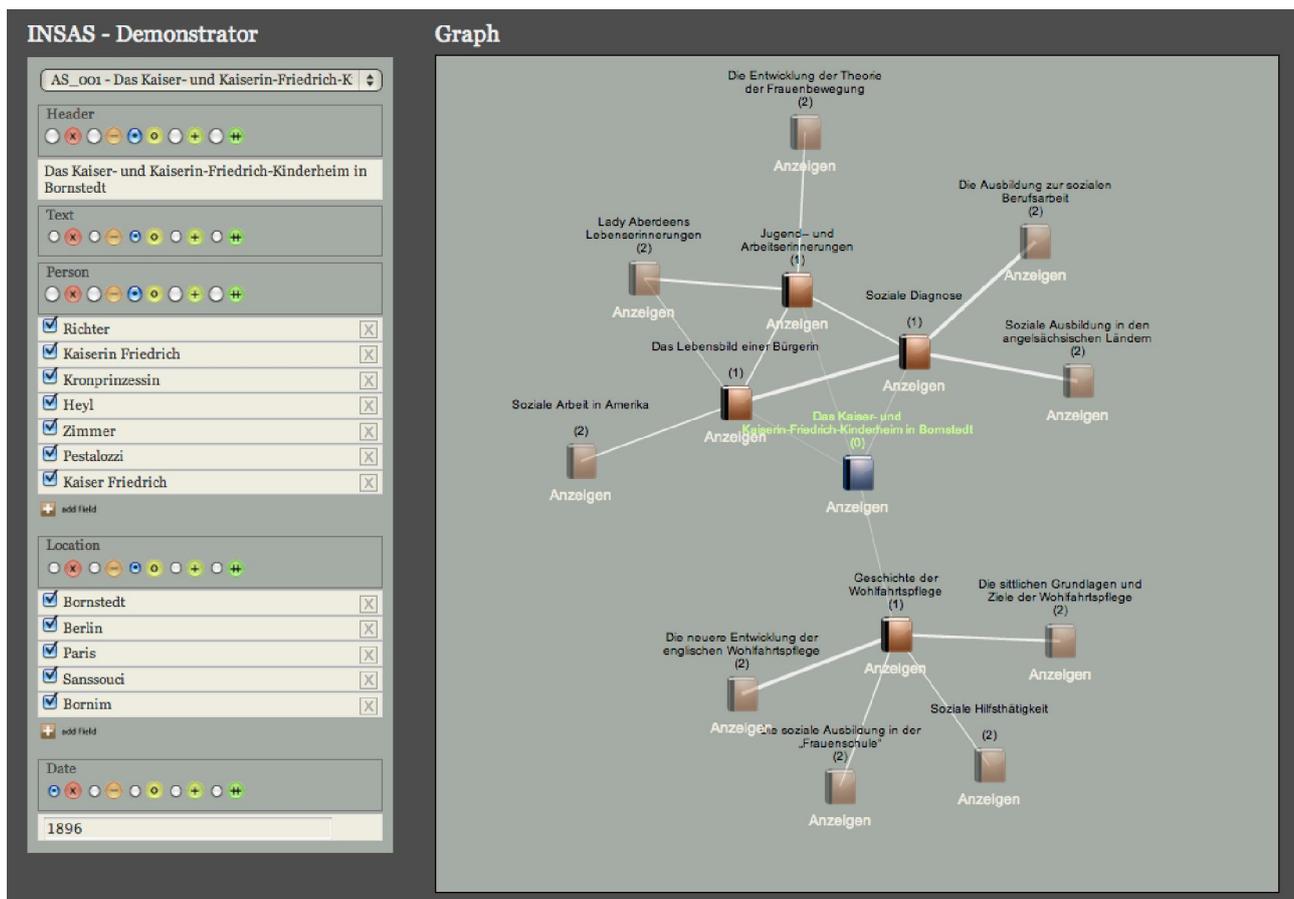


Abbildung 1: Prototyp des neuen kontext-basierten Suchsystems