

SALSAH - System for Annotation and Linkage in Arts and Humanities

Patrick Ryf, Lukas Rosenthaler, Tobias Schweizer
Imaging & Media Lab
Universität Basel
Bernoullistrasse 32, CH-4056 Basel
Tel: +41 61 267 04 89
www.iml.unibas.ch

Zusammenfassung

SALSAH ist eine mehrsprachige und browserbasierte virtuelle Forschungsumgebung, welche Werkzeuge zur Betrachtung, Annotation und Verlinkung von digitalisiertem Material sowie von Bestandteilen desselben bietet.

Weltweit wird durch Bibliotheken, Archive und andere Institutionen eine grosse Zahl an Digitalisaten und Meta-Informationen den Forschenden zur Verfügung gestellt, nicht jedoch die Werkzeuge, diese auch effizient nutzen zu können. Diese Lücke zwischen den Angeboten der Datenprovider¹ und der effektiven Nutzbarkeit dieser Daten in der alltäglichen Forschungsarbeit und Lehre versucht SALSAH zu schliessen.

Abstract

SALSAH is a multilingual and browser-based virtual research environment that offers tools for examining, annotating and linking of digitalized materials and components thereof.

A huge amount of digital images and meta-information is provided to researchers worldwide, but tools that offer/allow for an efficient utilization of this information are still lacking. SALSAH aims at closing this gap between the proposals of providers of data and the effective daily usability of the data.

Hintergrund

Die beschleunigte Entwicklung in den Kommunikations- und Informationstechnologien Mitte der 1990er-Jahre und ihr Transfer in alle Gesellschaftsbereiche hat zu einer nachhaltigen Veränderung der Organisation und Arbeitsabläufe in allen wissenschaftlichen Disziplinen geführt. In den Geistes- und Kunstwissenschaften führte dies beispielsweise zum Aufbau digitaler Informationsangebote und virtueller Fachbibliotheken. Obwohl die Zahl der Digitalisierungsprojekte in den letzten Jahren fortwährend stieg und die Fülle an Material exponentiell zunahm, blieb die Wiedergabe desselben bisher hinter den Anforderungen geisteswissenschaftlicher Forschung zurück. Während sich in den letzten drei bis vier Jahren interaktive Web 2.0-Anwendungen in unterschiedlichen Online-Communities² zusehends durchgesetzt haben, fehlen solche Darstellungen für die digitalisierten Bestände von Bibliotheken und Archiven weiterhin. Die Darstellungen orientieren sich insbesondere an archivalischen Findmitteln oder an traditionellen Katalogen der Bibliotheken. Beide Konzepte sind jedoch auf die Suche und das Finden von Quellen und Büchern mit Hilfe umfangreicher Meta-Daten ausgerichtet, nicht jedoch auf die wissenschaftliche Bearbeitung deren Inhaltes.

An diesem Punkt setzt SALSAH an. Im Gegensatz zu den bestehenden Plattformen von Archiven und Bibliotheken versteht sich SALSAH nicht als Online-Archiv oder Online-Bibliothek für Digitalisate, sondern als virtuelle Forschungsumgebung, welche den Forschenden Werkzeuge zur Nutzung der bereits im Web zugänglichen Digitalisate oder Texte zur Verfügung stellt.

¹ das sind in diesem Fall Bibliotheken, Archive und alle Institutionen, welche online Digitalisate, Meta-Daten oder andere Informationen zur Verfügung stellen

² insbesondere 'social networks' wie Facebook.

Werkzeuge vom SALSAH

Besonderes Augenmerk wurde bei der Entwicklung von SALSAH auf die Benutzerfreundlichkeit gelegt. Ein dem Benutzer vertrautes Fenster-System³ ermöglicht abgebildete Elemente mit Drag&Drop zu verschieben, zu minimieren und sowohl das Digitalisat als auch die umgebende Arbeitsfläche zu vergrössern oder zu verkleinern. Dadurch wird die Applikation für jede Bildschirmgrösse oder Auflösung individuell gestaltbar und der begrenzte Platz des Browserfensters optimal genutzt. Durch dieses interne Fenster-System wird für SALSAH immer nur ein einziges Browser-Fenster benötigt. Die intuitive Anordnung der einzelnen Elemente machen SALSAH übersichtlich und ohne grossen Lernaufwand einsetzbar. Digitalisate können je nach Art des Inhaltes mit unterschiedlichen Werkzeugen betrachtet werden⁴. Ausserdem steht dem Benutzer ein stufenfreier Zoom zur Verfügung, welcher jede gewünschte Auflösung darstellen kann. Das Fenster-System und der Zoom ergänzen sich ideal und ermöglichen ein optimales an den Benutzer angepasstes Arbeiten.

Während das Fenster-System innerhalb des Browsers und der stufenfreie Zoom der Verbesserung der Benutzerfreundlichkeit dienen, sind die personalisierten Annotationen und Verlinkungen von Inhalten neue Elemente im Umgang mit Digitalisaten. Der angemeldete Benutzer kann Annotationen hinzufügen, welche nur für ihn sichtbar sind. Solche Annotationen beschränken sich jedoch nicht auf ganze Digitalisate, sondern auch auf Teile derselben. Mit einer Bereichsauswahl kann ein Ausschnitt eines Textes oder Bildes mit einer Annotation versehen werden, welche über die personalisierte Suche⁵ wiedergefunden werden kann.

Neben der einfachen Annotation, können auch mehrere Digitalisate oder Ausschnitte derselben mit einer gemeinsamen Annotation versehen werden. Damit können untersuchte Bereiche direkt im System miteinander verlinkt und annotiert werden. Sowohl einfache Annotationen als auch Verlinkungen können über die personalisierte Suche oder über das systemeigene Annotations- und Verlinkungsmodul wiedergefunden werden. Sowohl Annotationen als auch Verlinkungen sind mit eigenen lokal gespeicherten, aber auch mit anderen im Web frei zugänglichen Inhalten möglich, auf welche dann referenziert wird.

Die Suche in SALSAH stützt sich auf Schnittstellen, welche es erlauben die eigenen, auf der SALSAH-internen Infrastruktur gespeicherten Datensätze, als auch die Datensätze externer Datenprovider zu durchsuchen. Aufgrund des auf verwendeten Datenmodells sind sowohl Freitext-Suchen als auch strukturierte Suchen möglich. Der Suchbereich kann dabei auf beliebige Kombinationen von internen und externen Datenprovidern⁶ eingeschränkt werden.

Verwaltung von SALSAH

SALSAH verfügt über ein differenziertes Berechtigungssystem, welches erlaubt Gruppenrechte für die Betrachtung, das Editieren und die Webfreigabe zu vergeben. Dieses System fördert das kollaborative Arbeiten in Forschungsgruppen und bietet gleichzeitig die Möglichkeit einer gezielten Qualitätskontrolle der Einträge.

Neben der Nutzung als Forschungsinstrument erlaubt die Verwaltung von SALSAH auch den Einsatz als Lehrinstrument. Gruppenrechte für Studenten erlauben es, Studentinnen und Studenten direkt am Forschungsprozess teilhaben zu lassen, ohne dass für die Dozierenden ein Mehraufwand entsteht.

³ das Fenster-System von SALSAH enthält viele Elemente, welche von Windows oder MacOS X her bekannt sind.

⁴ damit soll den unterschiedlichen Anforderungen aus verschiedenen Fachbereichen Rechnung getragen werden.

⁵ neben den öffentlich zugänglichen Daten, können auch eigene Einträge gesucht werden.

⁶ Bibliotheken, Archive oder andere Datenquellen.

Implementation & Datenmodell

Die Implementation des Systems basiert auf LAMP (Linux, Apache, MySQL, PHP) sowie auf dem Javascript Framework jQuery⁷, AJAX⁸ und HTML5. Durch die Verwendung von HTML5 benötigt SALSAAH zwar moderne Webbrowser, verzichtet jedoch auf Plugins wie beispielsweise Flash oder Zotero.

Im Gegensatz zum verbreiteten relationalen Datenbankmodell verwendet SALSAAH überwiegend ein detailliert ausgearbeitetes RDF-Modell⁹. Dieses Datenmodell hat insbesondere Vorteile bei der Verwendung von ‚weichen Definitionen‘¹⁰, wie sie in den Geisteswissenschaften häufig vorkommen. Zudem ist das RDF-Datenmodell besonders geeignet, um vernetzte Zusammenhänge zu repräsentieren. Allerdings musste das RDF-Datenmodell, welches vom World Wide Web Consortium (W3C) für das „Semantische Web“¹¹ entwickelt worden ist, in einigen wesentlichen Punkten erweitert werden, um den geisteswissenschaftlichen Bedürfnissen zur Datenrepräsentation zu genügen.

Externe Datenbestände können in SALSAAH eingebunden werden, wenn die externen Datenprovider¹² einen entsprechenden Zugang zur ihrer Datenbasis Verfügung stellen. Der Zugang erfolgt in der Regel von Server zu Server über einen Webservice gemäss dem SOAP oder RESTful-Standard. Dank der flexiblen Datenmodellierung in SALSAAH können externe Metadaten-Schemas sehr einfach in SALSAAH eingebunden werden. Dadurch werden die Funktionen für Annotation und Verknüpfung auch für externe digitalen Quellen verfügbar, ohne dass eine redundante Verdopplung der Datenhaltung benötigt wird. Die digitalen Quellen und ihre Metadaten liegen nach wie vor beim externen Datenprovider, während Annotationen und Verknüpfungen in der RDF-Struktur innerhalb von SALSAAH verwaltet werden.

Fazit

Die Hauptmerkmale von SALSAAH sind:

- die webbasierten Annotations- und Verknüpfungswerkzeuge, mit welchen die virtuelle Forschungsumgebung durch die Verwendung moderner Programmiermethoden die Benutzerfreundlichkeit einer normalen Desktop-Applikationen erreicht.
- Die Einsatzmöglichkeit in praktisch allen Disziplinen der Geisteswissenschaften. Die Verwendung des erweiterten RDF-Datenmodells erlaubt eine äusserst flexible Adaption an die verschiedensten Bedürfnisse durch eine Art „Simulation“ externer Datenmodelle.
- Die Möglichkeit digitale Archive, Datenprovider und Forscher zu vernetzen, indem die Werkzeuge von SALSAAH sowohl auf intern als auch extern vorhandene Datensätze angewandt werden können und durch das Benutzerkonzept kollaborative Arbeitsweisen unterstützt werden.

Damit erlaubt SALSAAH den konsistenten Zugriff auf geografisch und semantisch disjunkte Datenquellen. Die digitalen Quellen können mit den zur Verfügung gestellten Werkzeugen einfach kontextualisiert werden. Es entsteht eine Art Informations- oder Wissensnetz, das einerseits Synergien

⁷ <http://jquery.com>

⁸ Asynchronous server request; Eine Möglichkeit, eine Webseite zu aktualisieren, ohne dass die ganze Seite neu geladen werden muss.

⁹ das Resource Description Framework bezeichnet Standards des World Wide Web Consortiums zur Beschreibung von Informationen über Objekte.

¹⁰ in den Geisteswissenschaften lässt sich ein Objekt oft unterschiedlich einordnen, da einheitliche Definitionen fehlen. Dies kann dazu führen, dass für dasselbe Objekt mehrere Definitionen bestehen. In einem relationalen Datenbankschema liesse sich ein solches Objekt nicht mehr einfach und konsistent abbilden. Mit dem RDF-Schema ist dies kein Problem.

¹¹ http://de.wikipedia.org/wiki/Semantisches_Web

¹² die Daten können aus Bibliotheken, Archive oder aus anderen Quellen stammen.

zwischen Forschungsprojekten zu Tage fördern kann und andererseits die nachhaltige Nutzung von Forschungsergebnissen erlaubt.

SALSAH geht als webbasierte Forschungsumgebung wesentlich weiter, als dies Bibliotheken oder Archive mit ihren Online-Beständen tun. Unter Verwendung der neusten Webtechnologien und einer in hohem Masse integrativen Anwendung derselben bietet SALSAH Forschungswerkzeuge an, welche sich in erste Linie an den Bedürfnissen der geistes- und kunstwissenschaftlichen Forschung orientiert. SALSAH stellt einen Schreibtisch zur Verfügung, welcher Notizen, Verweise, Meta-Daten und Bilder vereint, den individuellen Bedürfnissen der Benutzer angepasst werden kann und damit diesem eine Mehrwert bietet.