

Harald H. Zimmermann

Zu den Grundlagen des ELSA-Projekts

Zielsetzung

Der Projektname *ELSA* steht für ‚*Prototypische multimedial basierte Entwicklung eines Elektronischen Archivs*‘. Ziel ist die informationstechnische Entwicklung eines *Prototyps* zur Speicherung und Präsentation literarischer und literatur-verwandter Archivalien.

Die Anwendung des zu erstellenden Systems erfolgt im Projektzeitraum an ausgewählten Materialien des *Literaturarchivs Saar-Lor-Lux-Elsass*, einem Teilbereich der Saarländischen Universitäts- und Landesbibliothek, die vom Anwendungs-Partner bereitgestellt werden. Wegen der großen Bandbreite der diesbezüglichen Archivalien wurden spezielle Teile des sog. Vorlasses des saarländischen Schriftstellers, Hörspielautors, Lieder- und Filmemachers *Alfred Gulden* zugrundegelegt. Eine Ausweitung auf andere Archivalien des Archivs sowie die spätere Anwendung in anderen Archiven der Öffentlichen Hand wie der Privatwirtschaft ist konzeptionell vorbereitet.

Das Projekt wird von der Fachrichtung Informationswissenschaft durchgeführt und verantwortet. Der Anwendungs-Partner betreut das Arbeitspaket ‚Datenerfassung‘ und wirkt bei der Strukturierung der Datenbank und der Gestaltung des Retrieval- und Präsentationskonzepts beratend mit. Das Projekt ELSA wird **in der gegenwärtigen Phase** im Rahmen der Landesinitiative Telekommunikation Saar durch das Saarland bezuschusst (**Projektnummer 0700**) und endet zum 30.9.2001. Die Entwicklungen erfolgen auf Werkvertragsbasis und werden größtenteils durch fortgeschrittene Studentische Hilfskräfte realisiert. Die Rechte der Nutzung liegen bei der Universität des Saarlandes als Rechtsperson. Die verantwortliche Projektleitung liegt beim Verfasser, die ausführende Projektleitung hat Dr. Heinz-Dirk Luckhardt. Die verantwortliche Leitung des Anwendungsbereichs hat Prof. Dr. Günter Scholdt, die ausführende **Projektleitung** hat hier Herrmann Gaetje M.A.

Vorgeschichte

Das Projekt ist eingebettet in die langjährigen Aktivitäten der Fachrichtung *Informationswissenschaft* der Philosophischen Fakultät III Empirische Humanwissenschaften an der Universität des Saarlandes im Schwerpunkt *Publikumsinformation*. Bereits seit Jahren existiert zu dieser allgemeinen Thematik an der Fachrichtung Informationswissenschaft ein Internet-Labor ‚MAIN - Multimedia-Angebote im Internet‘ (<http://www.phil.uni-sb.de/fr/infowiss/main/index.html>), dessen praktisches Resultat das Internet-Angebot zum Weltkulturerbe ‚Alte Völklinger Hütte‘ war und ist (<http://www.phil.uni-sb.de/fr/infowiss/huette/>). Ein Vorläufer ist das u.a. in zwei Seminaren erarbeitete) *Internet-Angebot* zu Alfred Gulden (<http://vili.phil.uni-sb.de/gulden/>). Man kann darüber hinaus weitere Teile des Informationsangebots der Fachrichtung im Internet zu den Vorläufern des ELSA-Projekts rechnen, etwa die Lösung ‚Virtuelle Lehre Informationswissenschaft‘ (ViLI): <http://vili.phil.uni-sb.de/vili/index.php>) oder den ‚Terminosaurus Rex‘ (<http://fs-infowiss.phil.uni-sb.de/trex/>).

Die Vorarbeiten zu ELSA haben in jüngster Zeit zu der Überlegung beigetragen, im Rahmen eines Forschungsseminars das Projekt ‚Nietzsche-Online‘ auf den Weg zu bringen, auch in der Hoffnung, aus den Ergebnissen des Seminars Rückschlüsse auf die Verfahrensweise bei ELSA ziehen zu können. Das Forschungsseminar sollte unter Mitarbeit der Studierenden u.a. klären, mit welcher Technologie man (derzeit) die Daten gestaltet und verarbeitet. Eine ‚reine‘

HTML-Lösung kam bereits hier nicht in Frage, so dass es weitgehend nur darum ging, ob man – informationstechnisch betrachtet – eine ‚reine‘ Datenbanklösung anstrebt oder SGML bzw. XML als Beschreibungsgrundlage für die Objekte (Daten) wählt.

Die Ergebnisse der Nutzung von XML im Rahmen von Nietzsche-Online (<http://bigmac.phil.uni-sb.de:82>) waren ermutigend, so dass nach einigen Diskussionen **für textbezogene Anwendungen** die Entscheidung auch bei ELSA für XML gefallen ist (s.u.). Dabei spielt bei ELSA - insbesondere im Zusammenhang mit dem **Retrieval der Daten / der Suche nach Daten** - eine Datenbank-Komponente (unter Nutzung der in den XML-**Materi- alien** entsprechend ausgezeichneten Merkmale) eine wesentliche Rolle.

Zu XML-Anwendungen liegen allgemein noch keine größeren Erfahrungen vor. Man ist also etwas ins ‚kalte Wasser‘ gesprungen. Das Vertrauen in diese Datenbeschreibungstechnik ist allerdings groß, da man in Forschungsprojekten der Fachrichtung mit der Verwendung von SGML bzw. analogen Dokumentauszeichnungen in früheren Jahren schon sehr positive Erfahrungen gemacht hat (EU-Projekt CANAL-LS).

Zur Frage der Nutzung des 'Internet'

Multimedialität, Interaktivität und Vernetzung sind Eigenschaften, die bei ‚Druckwerken‘ nicht verfügbar sind. Dies bedeutet nicht, dass die Erzeugung linearer gedruckter Lösungen keinen Sinn (mehr) macht. Dieser traditionelle Bereich bleibt aber aus vielerlei Gründen außerhalb des Projektziels, auch wenn später ggf. Schnittstellen zu einer entsprechenden Weiterverarbeitung zu entwickeln sein werden. Die Variante einer Off-line-Lösung, die gegenwärtig mit den Distributionsformen CD-ROM oder DVD verbunden wird, bleibt vorerst ebenfalls außer Betracht, obwohl man sich ‚hybride‘ Lösungen (CD-ROM / Internet) durchaus vorstellen kann. Das Werk von Alfred Gulden, zu dessen Repertoire u.a. Filme, Hörspiele und Lieder gehören (das also auch im herkömmlichen Sinne ‚multimedial‘ ist) erscheint ganz besonders geeignet, dem angestrebten Prototyp als materielle Grundlage zu dienen. Daher werden die multimedialen wie verknüpfungstechnischen Eigenschaften und Merkmale eines Online- Informations- und Kommunikationssystems (wie das heutige Internet mit dem Schwerpunkt des World Wide Web) optimal genutzt.

Nutzungsorientierung

Die ELSA-Lösung ist nutzer- und nutzungsorientiert. Daher bilden die ‚Einstiegs- bzw. Zugangsstellen‘ eines Nutzers / einer Nutzerin zum *Informationssystem* (Archivsystem) eine wichtige Gestaltungsgrundlage. Solche ‚Einstiege‘ sind etwa das *Register*, das *Glossar*, die *FAQs*, die *Werkübersicht* ... Ein weiteres Ziel ist die Schaffung von *Selektionsmöglichkeiten*, um dem Nutzer / der Nutzerin ein Instrument an die Hand zu geben, die Systemreaktionen auf diejenigen Bereiche zu orientieren, die seiner / ihrer Vorstellung entsprechen, wobei dies jeweils eine Option und keinen ‚Zwang‘ darstellt. Die Differenzierung der *Objekttypen* (Brief, Werk, Quellen ...) ist dazu eine Methode, aber auch eine Differenzierung bezüglich der ‚Tiefe‘ (- des Volumens: Volltext / Extrakt ... -) gehört hierher.

Die Förderung der *Kommunikation* der Nutzerinnen und Nutzer ‚untereinander‘ stellt ein weiteres Hauptziel dar. Im Rahmen des Projekts konnte bisher aus Zeitgründen dazu nicht alles praktisch gestaltet werden, auch zum Projektende wird erst ein Prototyp vorliegen. Es ist aber bedacht, dass (qualifizierte) Kommentierungen zugelassen werden, wobei die Frage der geeigneten Form einer ‚Moderation‘ - auch mit Blick auf die (Folge-)Kosten - noch weiterer

Überlegungen bedarf. Da das Gedankengut des (jeweiligen) Autors nicht ohne (Vor-)Wissen jedermann zugänglich ist, ist ein Instrumentarium zu schaffen, das den interessierten Laien wie den spezialisierten Wissenschaftlerinnen und Wissenschaftlern - natürlich auf unterschiedliche Weise - nützt, etwa in Form terminologischer Erläuterungen, spezifischer Verweise, 'Übersetzungen'. Dies gilt auch für die Ausgestaltung des / der Register(s), etwa in Richtung auf einen Thesaurus bzw. eine Terminologie-Bank.

Die Zielgruppen

Das Elektronische Literaturarchiv Saar-Lor-Lux-Elsass (als Modellfall) wird sich natürlich nicht auf *einen* Autor konzentrieren, es ist vielmehr offen für die Integration anderer Autoren / Autorinnen (der Region). Ausgangspunkt ist die Überlegung, dass es nötig und möglich ist, Werke und Leben der Autoren / Autorinnen einem breiteren Publikum systematisch nahe zu bringen. Dazu soll ein Online-Angebot, d.h. in heutiger Technologie auf der Basis des Internet, konzipiert und modellhaft realisiert werden.

Die Orientierung an einem breiteren Publikum, d.h. an dem sog. 'interessierten Laien', schließt nicht aus, dass auch (Literatur)-Experten dieses Angebot sinnvoll nutzen können. Aber auch die Nutzung des Archivs im *schulischen Unterricht* kann in mehrfacher Hinsicht lohnend sein: Einerseits wird die Jugend mit den literarischen Leistungen (hier: der Region) besser vertraut, andererseits lassen sich literarische Fragestellungen an Materialien, die von der Idee, dem ersten Entwurf – einer Notiz – über Materialsammlungen, Konzeptstudien, Vorfassungen bis zu einem Produkt (einer gedruckten Publikation, einem Film, einem Lied ...) reichen, hervorragend dokumentieren, Projektarbeiten gestalten u.a.m. Flankierend zur Entwicklung des Prototyps wird daher im Rahmen eines (weiteren) Seminars der Fachrichtung die entsprechende mögliche Nutzenanwendung behandelt und diskutiert.

Die 'Meinungen' über das Werk und die Wirkungsgeschichte eines Autors lassen sich nicht (immer) auf einen Nenner bringen. Es gilt also, die Möglichkeit der *Diskussion* (unter Fachleuten wie interessierten Laien) zu schaffen, ohne dass das Informationssystem Gefahr läuft, selbst wieder instrumentalisiert zu werden. Dies bedeutet, dass vor einer 'Freigabe' an das breite Publikum zumindest ein umfangreicher Grundstock der originalen Werke erschlossen und verfügbar ist. *Aus Gründen des Daten- und Persönlichkeitsschutzes müssen zudem Einschränkungen der Nutzung vorgesehen werden.* Eine Online-Präsentation des Gesamtwerks mit der Möglichkeit, sich zu jeder beliebigen Stelle zu bewegen, ist eine weitergehende Aufgabe. Sie kann allerdings im Rahmen der prototypischen Gestaltung nicht (mehr) erfolgen. Für den Aufbau wird Expertenwissen benötigt. Dies führte im Projektzeitraum zu einer intensiven arbeitsteiligen Zusammenarbeit mit den Experten des Literaturarchivs Saar-Lor-Lux-Elsass, aber auch – in diesem spezifischen Fall: mit dem ‚betroffenen‘ Autor. Das Konzept ist dabei offen für jede Art der Zusammenarbeit mit Fachleuten, wo sie sich auch immer befinden mögen.

Obwohl derzeit anhand der verfügbaren Mittel, die vom Saarland im Rahmen der Telekommunikations-Initiative zur Verfügung gestellt wurden, 'nur' ein Prototyp konzipiert und entwickelt werden kann, ist es doch das Ziel, eine dauerhafte Lösung zu realisieren, d.h. die *Nachhaltigkeit* der prototypischen Lösung zu sichern. Dies kann nicht in einem wissenschaftlichen universitären Rahmen geschehen. Deshalb wird ELSA als informationstechnische *Lösung* für das Literaturarchiv Saar-Lor-Lux-Elsass realisiert, dessen Aufgabe es sein wird, einen schrittweisen Ausbau des Online-Informationssystems und dessen Pflege sicherzustellen.

Zur Struktur des Online-Angebots

Die größte - bis lang nicht gelöste - Schwierigkeit des Internet kann mit der Anforderung beschrieben werden, die zu einer Problemlösung vorhandenen relevanten Daten einem *spezifischen* Nutzer *gezielt, präzise, vollständig und aktuell* verfügbar zu machen. Hierbei handelt es sich um ein typisches informationswissenschaftliches Problem, das sich auf die Berechnung des Recall und der Precision oder allgemein der Relevanz, bezogen auf eine spezifische Suchanfrage (hinter der sich ein spezifisches Problem verbirgt), abbilden lässt.

Unter *Recall* wird das (numerische) Verhältnis der gefundenen relevanten Belege zu allen in einem System vorhandenen relevanten Belegen verstanden, unter *Precision* das Verhältnis der gefundenen relevanten Belege zu allen gefundenen Belegen. Die Praxis hat andererseits gezeigt, dass das Erreichen der jeweiligen 'Idealwerte' (auf einer Skala von 0 bis 1 also jeweils 1,0) nicht nur kaum möglich erscheint, sondern sich die beiden Maße eher konkurrierend gegenüberstehen: Bei einer hohen Precision (viele der Treffer sind relevant) sinkt der Recall-Wert (nur wenige der im System vorhandenen relevanten Treffer werden 'gefunden') und umgekehrt.

Auch aus der Beobachtung des Nutzerverhaltens ergeben sich zudem erhebliche Probleme. Dies gilt sicherlich in besonderem Maß für den Laien-Benutzer. Die meisten Systeme (auch die Suchmaschinen des Internet) bieten Hilfen zur Formulierung komplexerer Suchanfragen an (etwa in Form von Klammerausdrücken in Verbindung mit booleschen Operatoren), doch werden diese in der Praxis nicht oder kaum benutzt. Die 'normale' Suchanfrage besteht aus der Aneinanderreihung von für relevant gehaltenen Wörtern und einer daraus abgeleiteten Volltextsuche. Aus diesen eher kritischen Anmerkungen sollte nun nicht automatisch der Umkehrschluss gezogen werden, dass auf eine (einfache) (Volltext-)Suche verzichtet werden könne. Es ist jedoch andererseits wenig fruchtbar, nur diese Methode als einzige Verfahrensweise zur Verfügung zu stellen. Daher muss nach alternativen bzw. ergänzenden Wegen gesucht werden, d.h. nach Verfahren, die dem Nutzer bei seiner Suche dahingehend helfen, seine Frage 'systemorientiert' zu gestalten, einen guten Startpunkt bei der Antwort auf seine Frage zu erhalten und diese weiter gezielt zu entfalten (d.h. zu modifizieren). Die Basis zu allem Verständnis der Vorstellungen und Überlegungen von Autoren – hier im Exemplan: von Alfred Gulden - sind ohne Zweifel seine *Werke*. Diese Werke elektronisch verfügbar zu machen und zu halten - seien es die Publikationen, Notizen, Entwürfe oder Fragmente, aber auch (ggf. selektierte) Äußerungen in dem Briefwechsel - ist ein wichtiges Desiderat. Zum Verständnis der Werke und einer Person kann seine Biographie einiges beitragen, aber auch Ergebnisse der Forschung sind hierbei einzubringen. Dennoch gilt letztendlich auch hier der Grundsatz: Nicht alles, was mit einem Informationssystem gemacht werden kann, muss schon vorstrukturiert sein ...

Für den Laien-Nutzer sind Hilfen und Zugangswege interessant, die sich an ausgewählten (vielleicht auch besonders diskutierten) Themen orientieren. Im Rahmen des Forschungsseminars zu Nietzsche wurde die Idee geboren und entwickelt, einen Zugangsweg über sog. 'Frequently Asked Questions' (FAQs) zu eröffnen, die zunächst in Form von 'Erwartens-Erwartungen' (also noch etwas 'aus der Retorte') gestaltet wurden. Sie wird auch in ELSA rudimentär eingebracht. Schließlich sollen die 'Themen' in Form komplexer Deskriptoren - durch einen Thesaurus erschlossen werden. Man kann sich damit - modern gesprochen - eine Nutzung in Form einer 'Guided Tour' vorstellen, auf der sich eine Nutzerin oder ein Nutzer ein Stück weit 'fortbewegen', sie aber jederzeit in der Lage sind, von diesem vorgegebenen Weg abzuweichen, etwa indem zu einem Wort eine 'Volltextsuche' gestartet wird (um weitere Belegstellen zu finden), man die Suche auf bestimmte Texttypen (z.B. Briefe) 'einschränkt' (oder entsprechend ausdehnt) oder indem man von einem Text- oder Brief-Zitat zum jeweili-

gen 'Volltext' verzweigt, Vorgänger oder Nachfolger sucht usf. Es können aber auch neue Formen vorgesehen werden (z.Zt. in Planung), etwa elektronische 'Kommentare', die 'nur' im Online-System verfügbar sind, auch (moderierte) Diskussionsforen sind vorstellbar.

Technische Realisierung

Für die elektronische Präsentationsform der zeichenkodierten Textquellen bildet zunächst die neue deutsche Rechtschreibnorm die Grundlage. Parallel wird jedoch ein Online-Zugang zur Schreibweise des Originals bzw. auch zu den Handschriften geschaffen. Im Zusammenhang mit der Entwicklung des Prototyps war zudem zu entscheiden, auf welcher Formatierungs-Grundlage die Datenerfassung und -speicherung erfolgen sollte. Das das Internet die Grundlage für den Nutzer-Zugang bilden sollte, wurde - wie eingangs schon angesprochen - nach einer kurzen Experimentierphase für den Prototyp das XML-Format gewählt, da es als Subset von SGML - beispielsweise über HTML hinausgehend - eine formal-inhaltliche Strukturierung erlaubt, Schnittstellen zur Präsentation in HTML (also auf gängiger Browser-Ebene) geschaffen werden können und Umsetzungen in relationale Datenbank-Formate (und damit die Nutzung von SQL) möglich sind. Gegenwärtig hat die entsprechende Implementierung begonnen. Es ist davon auszugehen, dass noch einige Erfahrungen gesammelt werden müssen.

Zum Prototyp

Am *Beispiel* von Alfred Gulden sollte möglichst problemnah die Thematik des *elektronischen Archivierens* 'erprobt' werden. Dabei muss davon ausgegangen werden, dass der Prototyp manche Fragen bzw. Lösungen nur 'andeuten' kann. So wird z.B. die Gestaltung von Foren zunächst ausgeklammert, bei Kommentierungen wird es allenfalls einige Beispiele geben. Der Prototyp wird modellhaft folgende (elektronische) Teile (= Dokumenttypen) umfassen, die durch Verweise miteinander vernetzt werden:

- Fragen mit Antworten, ggf. Verweise zu Zitaten usf.
- (elektronische), Extrakte zu Werken oder Briefen - weiter typisiert
- (zugehörige) Briefe von und an Gulden
- (elektronische) Werk-Extrakte
- Werkteile
- Quellenangaben, weitgehend mit Bezug zu den übrigen Objekten
- Personenbeschreibungen, Porträts, z.B. zu Brief-Autoren, erwähnten Personen, ...
- Ortsbeschreibungen
- Institute, z.B. Verlage, Museen, Gedenkstätten
- Links (annotiert) zu anderen Online-Quellen
- Thesaurus mit Register-Einträgen
- Glossar mit Begriffs-Erklärungen
- Photo(s)
- Film(e)

Zu diesen Dokumenttypen wurden / werden entsprechende *Document Type Definitions* (DTDs) erstellt und *Prozeduren* zum (automatischen) Erstellen von HTML-Präsentationen geschrieben. Das Konzept ist so angelegt, dass Objekte auch in anderen Sprachen als in Deutsch (etwa in Englisch oder Französisch) beschrieben bzw. unmittelbar eingebunden und vernetzt werden können.

Perspektiven

Ziel des Projekts ELSA 'nach außen' ist es, ein erstes Design eines systematischen Online-Angebots zu (literarischen) Archivdaten zu erstellen und mit dem Prototyp die Möglichkeiten zu veranschaulichen. Die Einbindung externer Verweise ('Links') soll verdeutlichen, dass sich das Online-System auch als *Drehscheibe* für Aktivitäten außerhalb des eigenen Systems versteht. Dabei ist es ein wichtiges Desiderat, die Entwicklungen in eine ernsthafte Diskussion um ein verteiltes, für nationale wie internationale Beteiligung offenes Informationsangebot zu (literarischen) Archivmaterialien münden zu lassen. Da sich das System auf die Online-Distribution konzentrieren wird, kann es Partnerschaften etwa im Hinblick auf CD- oder DVD-Lösungen oder Druckausgaben geben, die von entsprechenden Partnern (Verlagen ...) erstellt und vertrieben werden.

Einblick in die laufenden Arbeiten

Die laufenden Arbeiten können im Internet unter der Adresse <http://bigmac.phil.uni-sb.de> verfolgt werden (Passwort erforderlich). Es wird jedoch darauf verwiesen, dass es sich um Entwicklungen handelt, die ständig im Fluss sind. Änderungen sind gegenwärtig an der Tagesordnung.