

WALU — Eine Annotations- und Lern-Umgebung für semantisches Tagging

Andreas Wagner and Marc Rössler

Universität Duisburg-Essen
Computerlinguistik
Lotharstraße 65
D-47048 Duisburg
`andreas.wagner@uni-due.de`
`marc.roessler@uni-due.de`

Zusammenfassung WALU (WIKINGER Annotations- und Lern-Umgebung) ist eine Software zur Annotation und (semi-)automatischen Erkennung von Eigennamen sowie Instanzen anderer semantischer Kategorien in Texten. Ziel der Entwicklung von WALU ist die Realisierung eines komfortablen Werkzeugs, das von Experten unterschiedlichster Domänen ohne computerlinguistische und informatische Vorkenntnisse eingesetzt werden kann. Dies unterscheidet WALU von existierenden Annotations- und Lern-Umgebungen im Bereich Informationsextraktion, die auf andere Tasks zugeschnitten oder multifunktionell ausgelegt sind, was einen erheblichen Konfigurationsaufwand erfordert. WALU ist Teil der kollaborativen Wissens-Infrastruktur, die im eScience-Projekt WIKINGER entwickelt wird. Darüber hinaus ist es als Stand-Alone-Tool einsetzbar. Dieser Beitrag spezifiziert die Design-Prinzipien von WALU, skizziert die aktuell implementierten Funktionalitäten und beschreibt seinen Einsatz im WIKINGER-Kontext.

1 Einleitung

WALU (WIKINGER Annotations- und Lern-Umgebung) ist eine Software zur Annotation und (semi-)automatischen Erkennung von Eigennamen sowie Instanzen anderer semantischer Kategorien in Texten. WALU wird im Rahmen des Projekts WIKINGER entwickelt.

Dieser Beitrag gliedert sich wie folgt: Abschnitt 2 skizziert das Projekt WIKINGER. Abschnitt 3 spezifiziert die Anforderungen an die Eigennamenerkennungsstrategien am Allgemeinen und an das zu verwendende Werkzeug im Besonderen. Abschnitt 4 beschreibt den aktuellen Entwicklungsstand von WALU und zeigt weitere Perspektiven auf. Abschnitt 5 befasst sich mit der Anwendung von WALU für die WIKINGER-Pilot-Domäne. Abschnitt 6 enthält einen generellen Vergleich von WALU mit anderen einschlägigen Annotations-Tools. Abschnitt 7 enthält Zusammenfassung und Ausblick.

2 WIKINGER

Das Projekt WIKINGER (WIKI Next Generation Enhanced Repository), vgl. – Literatur– wird im Rahmen der eScience-Initiative des BMBF gefördert (Förderungs-Zeitraum: 10/05–09/08). Ziel des Projekts ist die Entwicklung einer intelligenten Infrastruktur für einen effizienten Austausch wissenschaftlicher Ergebnisse. Diese Infrastruktur wird als semantisches Wiki organisiert, d.h. als Repositoryum von Dokumenten und Informationen, die mit Techniken des Semantic Web kodiert und vernetzt sind. Dieses Repositoryum wird von den Mitgliedern einer Forschungscommunity online erstellt und modifiziert. Dies ermöglicht den verteilten Auf- und Ausbau eines Informationsnetzes, welches Fachdomänen-Wissen effizient zugänglich macht. In WIKINGER wird zunächst die Pilot-Domäne Katholische Zeitgeschichte behandelt. Im Sinne der Nachhaltigkeit der entwickelten Infrastruktur ist ihre Wiederverwendbarkeit, d.h. ihre Portierbarkeit auf andere wissenschaftliche Domänen und industrielle Anwendungen jedoch ein wichtiges Desiderat.

Ein entscheidendes Projektziel ist die Entwicklung von Verfahren zum (semi-)automatischen Aufbau eines solchen semantischen Wissensnetzes aus einschlägigen Dokumenten der jeweiligen Domäne. Hierfür werden in einem ersten Schritt Named Entities (z.B. Personen, Orte, Organisationen) sowie die Vorkommen anderer für die Domäne wesentlicher semantischer Kategorien (z.B. Bedeutendes Ereignis) in den Texten erkannt und klassifiziert (d.h. semantisch getaggt). In einem nächsten Schritt werden diese Entitäten mit semi-automatischen Methoden zu semantischen Netzen verknüpft, indem ihre Kookkurenzen und die zugehörigen Kontexte in den Texten analysiert werden. Sowohl die Ergebnisse des semantischewn Taggings als auch die daraus resultierenden semantischen Netze unterliegen dem Feedback der Community-Mitglieder, die über das Wiki Zugriff darauf haben und ggf. Korrekturen und Ergänzungen vornehmen können.¹ Dieses Feedback wird zur Verfeinerung der automatischen Extraktionsverfahren eingesetzt. Abbildung 1 gibt einen schematischen Überblick über die Prozesse im WIKINGER-System.

An WIKINGER sind das Fraunhofer Institut für Intelligente Analyse- und Informationssysteme (IAIS), Sankt Augustin, die Computerlinguistik der Universität Duisburg-Essen, Duisburg, sowie die Kommission für Zeitgeschichte (KfZG), Bonn, beteiligt. Die Duisburger Computerlinguistik befasst sich mit der Eigennamen-Erkennung bzw. dem semantischen Tagging. Ein Kernstück der diesbezüglichen Aktivitäten ist die Entwicklung von WALU.

Weitere Informationen über das Projekt befinden sich unter <http://www.wikinger-esience.de>

3 Anforderungen an Annotations-Strategien

Aus dem Ziel der Wiederverwendbarkeit der in WIKINGER entwickelten Infrastruktur, d.h. ihrer Adaptierbarkeit an neue Domänen, ergeben sich wesentliche

¹ Außerdem besteht die Möglichkeit, neue Daten in das Wiki aufzunehmen.

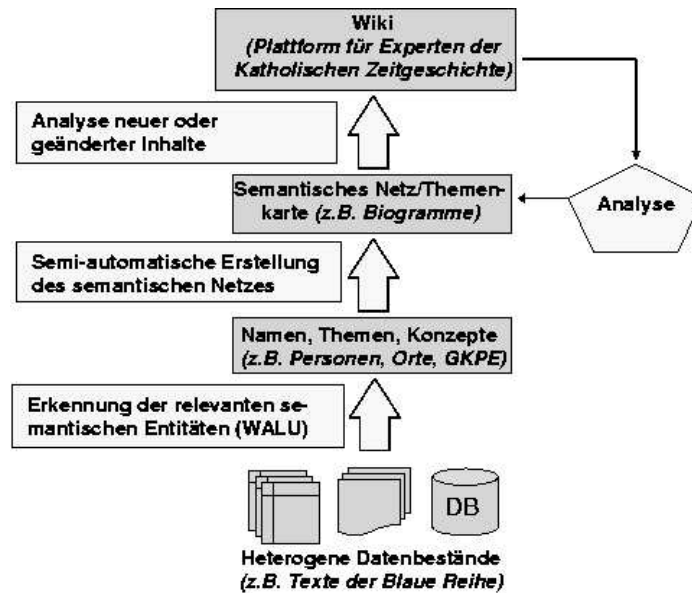


Abbildung 1. Überblick WIKINGER

Anforderungen an die eingesetzten Strategien zur semantischen Annotation. Dies betrifft sowohl die prinzipiellen Verfahren als auch die konkrete Ausgestaltung der verwendeten Werkzeuge.

3.1 Annotationsverfahren

Die Fülle der zu erschließenden Daten, die für eine bestimmte Domäne gewöhnlich vorliegen, macht den Einsatz automatischer und semi-automatischer Annotationsverfahren unerlässlich. Für NER existieren sowohl regelbasierte Verfahren als auch statistische (maschinelle) Lernverfahren. (– Literatur –) Im Hinblick auf das Desiderat der Adaptierbarkeit auf beliebige Domänen haben maschinelle Lernverfahren signifikante Vorteile. Bei der Anpassung von regelbasierten Verfahren an eine neue Domäne und/oder Sprache ist es i.d.R. erforderlich, das Set von Erkennungsregeln substanziell zu überarbeiten und schlimmstenfalls neu aufzusetzen. Da ein adäquates Regelwerk für praktische Anwendungen sehr komplex ist (sowohl hinsichtlich der Zahl der Regeln als auch hinsichtlich ihrer Interaktion), setzt diese Aufgabe profunde Kenntnisse über die neue Domäne einerseits und Wirkungsweise und das Zusammenspiel der Erkennungsregeln andererseits voraus. Dies erfordert eine zeitintensive und aufwändige Zusammenarbeit zwischen den jeweiligen Domänenexperten und Computerlinguisten. Dagegen ist für die Anpassung von maschinellen / statistischen Lernverfahren im Wesentlichen die Erstellung neuer Trainingsdaten erforderlich. Diese können durch die Annotierung einschlägiger Texte durch die Domänenexperten gewonnen werden;

der hierfür benötigte fachübergreifende Kooperationsaufwand mit Computerlinguisten ist vergleichsweise gering.

Dieser beispielbasierte Ansatz – die für das Lernverfahren benötigten Informationen werden durch Beispiele übermittelt, nicht durch explizite Regeln und Definitionen – erhöht die Domänen-Adaptivität auch bei dem Schritt, der dem eigentlichen Lernen vorausgeht: der Definition der zu lernenden semantischen Kategorien. Durch die manuelle Annotation verschiedener Instanzen einer Kategorie grenzen die Domänenexperten die Verwendung dieser Kategorie in ihrer Domäne ein (z.B.: Gilt eine Kirchengemeinde als geographische Entität?) und liefern so eine implizite Definition. Dies erspart die aufwändige Erarbeitung einer expliziten Definition. Ebenso ermöglicht das empirische Arbeiten mit den relevanten Texten im Zuge des Annotationsprozesses eine adäquatere Festlegung des Kategorien-Inventars: Neue Kategorien können eingeführt und praktisch irrelevante Kategorien eliminiert bzw. zusammengefasst werden, wenn während der Annotation ein entsprechender Bedarf festgestellt wird.

Im Sinne der Adaptivität sind also lernbasierte Ansätze zu bevorzugen. Besonders interessant sind Verfahren, die versuchen, den initialen Annotationsaufwand möglichst zu minimieren, z.B. durch Active Learning. Jedoch ist in begrenztem Umfang auch der Einsatz regelbasierter Verfahren sinnvoll, wenn sie mit geringem Aufwand gute Erkennungsleistungen erzielen und/oder domänenübergreifend einsetzbar sind. Beispielsweise bieten sich für die Erkennung von Datums- und Zeitangaben reguläre Ausdrücke an; diese sind vergleichsweise einfach definierbar und domänen- (wenn auch nicht sprach-) unabhängig. Unter das regelbasierte Paradigma fällt insbesondere die Verwendung von allgemeinen oder domänenspezifischen Listen von Kategorie-Instanzen (z.B. Personen oder Bistümern). Diese können entweder aus den annotierten Texten oder aus externen Quellen gewonnen werden. Wir halten die Anwendung von Listen für eine zufriedenstellende Erkennungsrate sowie zur Unterstützung bei der manuellen Annotation für unverzichtbar.

3.2 Werkzeug

Damit die Anpassung des Systems, d.h. die Erstellung von Trainingsdaten und das Trainieren und Anwenden von Lernverfahren, weitestgehend selbstständig von den jeweiligen Domänenexperten durchgeführt werden kann, ist es unabdingbar, hierfür ein geeignetes Werkzeug zur Verfügung zu stellen. Entscheidend ist, dass dieses Tool von von Experten unterschiedlichster Domänen ohne computerlinguistische und informatische Vorkenntnisse intuitiv bedienbar ist. Es sollte eine komfortable Oberfläche für den gesamten Annotations- und Lernzyklus bereit stellen. Dazu gehört der einfache Zugriff auf das Repositorium der zu annotierenden Texte, eine benutzerfreundliche Mechanismen zur manuellen Annotation, eine flexible Verwaltung des Kategorien-Inventars, vielfältige, konfigurierbare Anzeigemöglichkeiten, die nahtlose Integration listenbasierter Annotation sowie die Anbindung von automatischen Annotationsverfahren. WALU ist auf diese Anforderungen zugeschnitten. Im Sinne der Wiederverwendbarkeit ist WALU in Java implementiert und damit betriebssystem-unabhängig einsetzbar.

4 WALU – aktueller Entwicklungsstand

Die Prioritäten bei der Entwicklung von WALU spiegeln die Erfordernisse in WIKINGER wider. In der bisherigen Phase stand die manuelle Annotation durch die Domänenexperten im Vordergrund. Mit den so gewonnenen Daten werden Listen erstellt und maschinelle Lernverfahren trainiert, die in den kommenden Phasen sukzessive für die automatische Annotation eingesetzt werden. Dementsprechend stand bisher für WALU die Realisierung komfortabler Mechanismen zur manuellen Annotation, einschließlich der Anbindung listenbasierter Annotatoren, im Vordergrund.

4.1 Manuelle Annotation

WALU bietet eine komfortable Annotations-Oberfläche (vgl. Abbildung 2). Ein Default-Inventar semantischer Kategorien ist vorgegeben; darüber hinaus können, auch textspezifisch, neue Kategorien definiert werden. Eine Annotation erfolgt durch Markieren einer Instanz im Text und der Auswahl der entsprechenden Kategorie (über Kontextmenü, Buttons oder Shortcuts). Die Annotationen werden im Text farblich markiert (jeder Kategorie entspricht eine bestimmte Farbe, die frei definiert werden kann) sowie in einer separaten Liste neben dem Textfeld angezeigt. Jede Kategorie lässt sich im Text und/oder in der Liste ein- und ausblenden. Eine Annotation kann mit einem Kommentar versehen werden. Neben der Annotation ermöglicht WALU die manuelle Editierung von Dokumenten.

4.2 (Semi-)automatische Annotation

Aus den annotierten Texten extrahiert WALU Listen von Kategorie-Instanzen, die für die automatische Annotation weiterer Vorkommen dieser Instanzen eingesetzt werden können. Eine automatische Annotation wird zunächst als “unchecked” erfasst und dargestellt; eine manuelle Bestätigung führt zum Status “checked”, der zu einer manuellen Annotation äquivalent ist. Die Listen werden bei der Durchsicht der Annotationen interaktiv angepasst, indem beim Löschen einer falschen Annotation auch der entsprechende Listen-Eintrag entfernt werden kann (jedoch nicht muss). So wird die Qualität der Listen sukzessive erhöht.

Ein weiterer Mechanismus zur automatischen Annotation sind reguläre Ausdrücke. Z.Zt. ist ein regulärer Annotator für Datums- und Zeitangaben integriert. Weitere vordefinierte und konfigurierbare reguläre Annotatoren werden hinzukommen.

4.3 Qualitätskontrolle

WALU ist mit einer einfachen Zeichenketten-basierten Suchfunktion ausgestattet. Darüber hinaus ist ein spezieller Such- und Sortier-Modus für Annotationen implementiert. In diesem Modus werden die annotierten Entitäten (Types) mit den entsprechenden Kategorien in einer Liste angezeigt (sortiert nach Häufigkeit oder Alphabet). Klickt man auf ein Entitäts-Kategorie-Paar, werden die

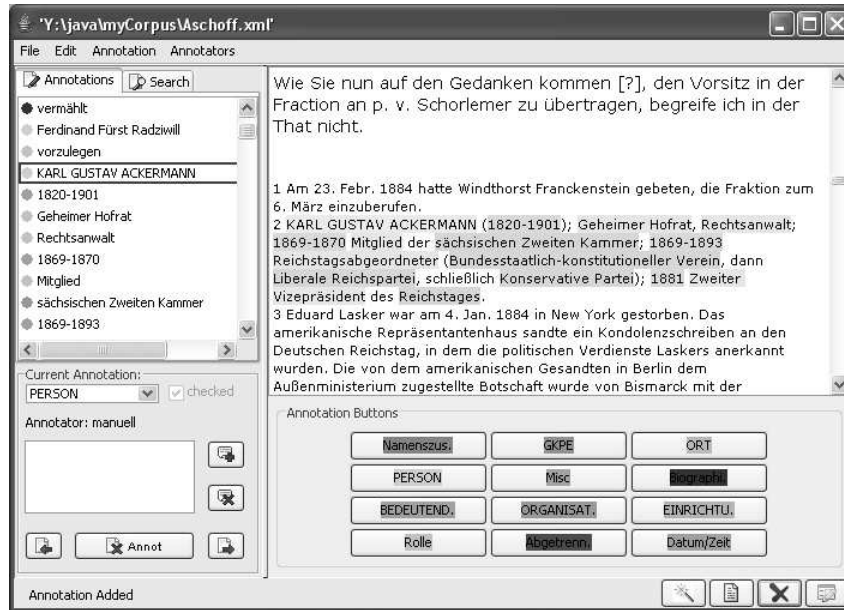


Abbildung 2. WALU-Screenshot

entsprechenden Vorkommen mit ihren Kontexten im KWIC-Format angezeigt und können direkt bearbeitet werden. So können die einzelnen Annotationen bestätigt, gelöscht oder die Kategorie geändert werden. Dies ermöglicht eine effiziente Kontrolle und Korrektur automatischer Annotationen, insbesondere im Hinblick auf Ambiguitäten. Werden beispielsweise durch listenbasierte Annotation alle Vorkommen von „Singen“ als Ort markiert, können diese Annotationen im Sortier-Modus zusammen aufgelistet werden und eventuelle Fehler (z.B. Fälle, in denen „Singen“ einem Personennamen oder der nicht zu annotierenden Tätigkeit entspricht) korrigiert werden. Zudem ermöglicht diese Überblicksdarstellung ein verlässlicheres Urteil darüber, inwieweit der Listen-Eintrag „Singen“ als Ort überhaupt hilfreich ist.

4.4 Import und Export

Beliebige Dokumente im Textformat können direkt in WALU importiert und annotiert werden. Die Importfunktion beinhaltet interaktive Möglichkeiten zur Auflösung von Mehrspalten-Text. Heuristisch wird zwischen Fließtext, Überschriften, Fußnoten sowie Kopf- und Fußzeilen unterschieden, die jeweils in unterschiedlichen Schriftgrößen dargestellt werden. WALU unterstützt Unicode.

Annotierte Dokumente werden in einem Standoff-XML-Format gespeichert.² Damit ist es möglich, diese Dokumente auch außerhalb von WALU zu bearbeiten. Abschnitt 5 geht näher auf das Standoff-Format ein.

4.5 Kommunikation mit der WIKINGER-Infrastruktur

Mit den bisher skizzierten Funktionalitäten ist WALU als Stand-Alone-Tool verwendbar. Wie jedoch in Abschnitt 2 dargestellt, bildet WALU zudem einen Teil der in WIKINGER entwickelten Infrastruktur. Die Einbettung in diese Infrastruktur wird durch spezielle Kommunikations-Mechanismen realisiert. Die in WIKINGER verwendeten Dokumente werden in einem Dokumenten-Repository verwaltet, die zugehörigen Annotationen sowie weitere Informationen im sog. Matadata Repository.³ Mit diesen Repositories, die i.Allg. auf einem entfernten Server betrieben werden, muss WALU Daten austauschen. Zum jetzigen Zeitpunkt ist das Dokumenten-Repository bereits implementiert (als gekapselte relationale Datenbank, mit der über Web Services kommuniziert wird⁴. Für das Metadata Repository existieren erste Design-Entwürfe (die ebenfalls auf relationalen Datenbanken basieren). WALU ist momentan in der Lage, die Dokumente im Repository anzuzeigen und von dort einzulesen. Für die vollständige erforderliche Konnektivität wird der Zugriff auf das Document Repository erweitert und flexibilisiert werden, sodass einerseits lediglich Teile von Dokumenten (z.B. alle Fußnoten) gelesen und andererseits ggf. editierte Dokumente wieder zurückgeschrieben werden können. Analoge Zugriffe werden für die Annotationen im Metadata Repository realisiert werden, sobald dieses implementiert ist. Diese werden mit den entsprechenden Dokument-Zugriffen synchronisiert.

5 Einsatz in der Pilotdomäne

Die Fachexperten der KfzG stellen einen Großteil der zu erschließenden Wissensquellen bereit: Die von der KfzG herausgegebene „Blaue Reihe“ umfasst mehr als 150 Bände zur Katholischen Zeitgeschichte. Alle diese Bände wurden im Rahmen des Projekts digitalisiert. Das domänenspezifische Inventar semantischer Kategorien wurde ausgehend von –Literatur– erarbeitet. Bei der manuellen Annotation exemplarischer Werke kristallisierte sich ein adäquates Inventar heraus, das wie folgt festgelegt wurde:

GKPE (Geographische Kirchliche/Politische Einheit), Ort, Einrichtung, Organisation, Person, Abgetrennter Namensbestandteil (z.B. in Registern), Namenszusatz, Rolle/Funktion, Biographisches Ereignis, Datum / Zeitangabe, Bedeutendes Ereignis

² Aus Gründen der Verarbeitungseffizienz wird ein Dokument und die zugehörigen Annotationen WALU-intern durch Java-Objekte repräsentiert.

³ Die Repositories beinhalten eine Versionskontrolle.

⁴ Das Mapping zwischen Java-Objekten und relationalen Datenbank-Strukturen erfolgt mittels Apache OJB.

Da die Zeithistoriker der KfZG im Rahmen der WIKINGER-Aktivitäten insbesondere an biographisch-bibliographischen Informationen über katholische Persönlichkeiten im 19. und 20. Jahrhundert interessiert sind, konzentriert sich die Annotation zunächst auf die sog. Biogramme, d.h. Kurzbiographien, die i.d.R. in Fußnoten präsentiert werden.

Listing 1.1 zeigt ein Biogramm und einen Teil der zugehörigen Annotationen im XML-Export-Format. Die XML-Struktur bildet das einfache und verarbeitungseffiziente interne Dokumentformat ab. Ein Text besteht aus einer Folge von Paragraphen, die unterschiedlichen Layout-Kategorien angehören (fn steht für eine Fußnote). Diesen folgt eine Liste der Annotationen für das Dokument. Jede Annotation kodiert die semantische Kategorie, die annotierte Entität, das Anfangs- und Endoffset des entsprechenden Vorkommens im Text, den Annotator – z.Zt. Human, listMatcher (listenbasierte Annotation) oder REtime (regulärer Ausdruck für Zeitangaben – sowie die Angabe, ob sie manuell überprüft wurde. Die Standoff-Annotation ermöglicht überlappende bzw. eingebettete Annotationen, z.B. die Annotierung der Städte im biographischen Ereignis „Studium der Medizin in Würzburg, Straßburg, Heidelberg“ als GKPE.

6 Vergleich mit existierenden Tools

WALU ist speziell auf die Annotation semantischer Kategorien ausgerichtet. Wie in Abschnitt 3.2 angeführt, ist es ein entscheidendes Ziel, ein komfortables Werkzeug zu implementieren, das von Experten unterschiedlichster Domänen ohne computerlinguistische und informatische Vorkenntnisse verwendet werden kann. Dies unterscheidet WALU von existierenden Annotations- und Lern-Umgebungen, die im Bereich Informationsextraktion eingesetzt werden, z.B. GATE ?, WordFreak –Literatur–, PALinkA ? oder MMAX –Literatur–. Diese sind i.d.R. für Benutzer mit (computer-)linguistischem Hintergrund (oder zumindest mit entsprechendem nachhaltigen Support) konzipiert. Dies hat zur Folge, dass sie entweder auf andere, komplexere Tasks zugeschnitten sind (z.B. PALinkA und MMAX für Koreferenz- und Diskurs-Annotation) oder im hohem Maße multifunktional ausgelegt sind (z.B. GATE, WordFreak). Diese Multifunktionalität erlaubt einerseits einen flexiblen, auf komplexe Bedürfnisse zugeschnittenen Einsatz, ist jedoch andererseits mit einem erheblichen Konfigurationsaufwand und einer für „Laien“ mitunter unintuitiven Bedienung verbunden.⁵

Zudem ist WALU sowohl als Stand-Alone-Werkzeug als auch als integraler Bestandteil der WIKINGER-Infrastruktur konzipiert. In letzterer Eigenschaft sind spezifische Web-Kommunikations-Module implementiert bzw. vorgesehen.

7 Zusammenfassung und Ausblick

WALU ist ein in der Entwicklung befindliches Werkzeug zum manuellen und (semi-)automatischen semantischen Tagging von Eigennamen und anderen Ka-

⁵ Z.B. muss in GATE beim Laden einer XML-Datei zusätzlich ein Dokument-Name angegeben werden, oder das System weist einen kryptischen Bezeichner zu.

```

<paragraph name="fn">55 Ferdinand Stemmler * 30.7.1868
Cronberg (Taunus), Gymnasium Montabaur, Studium der Medizin
in Würzburg, Straßburg, Heidelberg, Sanitätsrat
und Badearzt in Bad Ems, Facharzt für Balneologie,
MdL (Preußen) 1921–1928 für die Zentrumspartei.</paragraph>
...
<markup semCat="PERSON" cont="Ferdinand Stemmler"
  offset="229525" end="229543"
  annotator="listMatcher" checked="true"/>
<markup semCat="Biographisches Ereignis" cont="* 30.7.1868"
  offset="229544" end="229555"
  annotator="Human" checked="true"/>
<markup semCat="Datum/Zeit" cont="30.7.1868"
  offset="229546" end="229555"
  annotator="REtime" checked="true"/>
<markup semCat="GKPE" cont="Cronberg (Taunus)"
  offset="229556" end="229573"
  annotator="Human" checked="true"/>
<markup semCat="EINRICHTUNG" cont="Gymnasium Montabaur"
  offset="229575" end="229594"
  annotator="listMatcher" checked="true"/>
<markup semCat="GKPE" cont="Montabaur"
  offset="229585" end="229594"
  annotator="Human" checked="true"/>
<markup semCat="Biographisches Ereignis" cont="Studium
der Medizin in Würzburg, Straßburg, Heidelberg"
  offset="229596" end="229650"
  annotator="Human" checked="true"/>
<markup semCat="GKPE" cont="Würzburg"
  offset="229619" end="229627"
  annotator="listMatcher" checked="true"/>
<markup semCat="GKPE" cont="Straßburg"
  offset="229629" end="229638"
  annotator="listMatcher" checked="true"/>
<markup semCat="GKPE" cont="Heidelberg"
  offset="229640" end="229650"
  annotator="listMatcher" checked="true"/>
<markup semCat="Rolle" cont="Sanitätsrat und
Badearzt in Bad Ems"
  offset="229652" end="229687"
  annotator="Human" checked="true"/>
<markup semCat="GKPE" cont="Bad Ems"
  offset="229680" end="229687"
  annotator="listMatcher" checked="true"/>
<markup semCat="Rolle" cont="Facharzt für Balneologie"
  offset="229689" end="229713"
  annotator="Human" checked="true"/> ...

```

Listing 1.1. Beispiel XML-Exportformat

tegorien. Es wird im Kontext des Projekts WIKINGER entwickelt und bildet einen Teil der dort aufgebauten Infrastruktur. Ebenso ist WALU als Stand-Alone-Tool einsetzbar. Ein primäres Entwicklungsziel ist die nachhaltige Wiederverwendbarkeit, innerhalb und außerhalb der WIKINGER-Plattform. Dies wirkt sich sowohl auf die Wahl der automatischen Erkennungs-Verfahren (Beispielbasiertheit) als auch auf die konkrete Gestaltung des Tools (Komfortabilität und intuitive Bedienbarkeit geht vor Multifunktionalität) aus. WALU wird erfolgreich in der WIKINGER-Pilotdomäne eingesetzt.

Die nächsten Entwicklungsschritte konzentrieren sich auf die Anbindung maschineller Lernverfahren (SVM, MaxEnt) an WALU. Diese Verfahren sollen mit Active-Learning-Strategien verfeinert werden, um den Aufwand der Trainingsdaten-Erstellung zu minimieren. Des Weiteren werden wir uns der Vervollständigung der WIKINGER-Kommunikationsmodule sowie dem Aufbau des Metadata Repositorys widmen.