Einsatz einer semantischen Textanalyse zur Unterstützung von Redakteuren in Content Management Systemen

Sabina Maria Fröhlich

MASTERARBEIT

eingereicht am
Fachhochschul-Masterstudiengang

Interactive Media

in Hagenberg

im September 2013

Erklärung

Ich erkläre eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benutzt und die den benutzten Quellen entnommenen Stellen als solche gekennzeichnet habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Hagenberg, am 30. September 2013

Sabina Maria Fröhlich

Inhaltsverzeichnis

Kurzfassung

Die ständig wachsende Menge an Informationen im Internet brachte die Entwicklung des Semantic Web hervor, welches die Informationen strukturiert und nicht nur für Menschen aufbereitet, sondern auch maschinenlesbar macht. Dadurch ergeben sich neue Forschungsbereiche, wie die der semantischen Textanalyse.

Diese Arbeit behandelt verschiedene Ansätze, die Entwicklung des Semantic Web insbesondere für Redakteure einzusetzen, um sie bei ihrem täglichen Umgang mit Content Management Systemen zu unterstützen und dadurch die Usability zu verbessern. Dafür werden Ansätze für die automatische Erstellung einer Tag-Cloud, eine automatische Textzusammenfassung, die Eingliederung von Artikeln in die Navigation bzw. die Erstellung einer Navigation anhand der Artikel sowie eine semantische Textanreicherung näher betrachtet.

Anhand des letzten Beispiels – der semantischen Textanreicherung – wurde ein Prototyp entwickelt, um Redakteure bei der Erstellung von Artikeln zu unterstützen. Durch den Einsatz einer Semantic API werden Schlüsselwörter aus dem Text extrahiert. Mit Hilfe dieser Schlüsselwörter werden dem Redakteur verschiedene Bilder, weiterführende Links, verwandte Artikel oder ein Kartenausschnitt aus internen und externen Quellen vorgeschlagen. Diese können per Klick in den Text an eine beliebige Stelle eingefügt werden und ermöglichen somit eine schnelle Anreicherung des Textes. Eine Usability-Studie in Kombination mit einem Fokusgruppen-Interview hat gezeigt, dass Redakteure mit Hilfe des Prototypen in ihren täglichen Aufgaben gut unterstützt werden können und Prozesse dadurch verkürzt und erleichtert werden. Außerdem konnten im Zuge der Studie Vorschläge für Erweiterungs- und Verbesserungsmöglichkeiten gesammelt werden.

Abstract

The ever-growing amount of information on the internet causes the development of the Semantic Web, which structures the information and makes it not just understandable to humans, but also amenable to computer processing. This results in new research areas such as semantic text analysis.

This thesis discusses various approaches using the development of the Semantic Web particularly to support content-editors in their usual dealing with a Content Mangement System and try to thereby improve usability. Different approaches will be considered in more detail, like the automatic creation of a tag cloud, automatic text summarization, the classification of articles in the navigation or the creation of a navigation based on the articles, as well as a semantic text enrichment.

Based on the last example – the semantic text enrichment – a prototype was designed to support editors in writing and enriching articles. Through the use of a Semantic API, keywords are extracted from the text. With the aid of these keywords several pictures, related links, related articles or a map from internal and external sources are proposed to the editor. These things can be added at an arbitrary point in the text by clicking. This allows a quick enrichment of the text. A usability study, in combination with a focus group interview has shown that content-editors are well supported in their daily tasks using the prototype: above all the processes are shortened and made easier. Moreover, suggestions for additional features and possible improvements could be found.

Kapitel 1

Einleitung

1.1 Motivation

Redakteure in *Content Management Systemen* (CMS) sind immer wieder mit den selben Aufgaben konfrontiert. Sie recherchieren neue Themen, verfassen Artikel, reichern diese mit Bildern oder weiterführenden Links an und ordnen sie anschließend in die Navigation ein. Dabei müssen sie oft ihre gewohnte Arbeitsumgebung verlassen, um neue Anregungen zu finden, Bilder zu suchen und weiterführende Links zu sammeln.

Gleichzeitig bringt die Entwicklung des Semantic Web eine Menge an Daten in eine strukturierte, maschinenlesbare Form. Außerdem ermöglichen Semantic APIs durch eine semantische Analyse, Texte mit Metadaten anzureichern.

Vereint man diese beiden Thematiken, ergeben sich Wege, Redakteure bei den beschriebenen Aufgaben mit Hilfe einer semantischen Textanalyse zu unterstützen. Die durch die Analyse extrahierten Schlüsselwörter können für verschiedene Bereiche der Textanreicherung genutzt werden.

1.2 Zielsetzung

Ziel dieser Arbeit ist es, herauszufinden, ob mit Hilfe einer semantischen Textanalyse die Usability und die Zufriedenheit der Redakteure in Content Management Systemen verbessert werden kann. Dazu sollen verschiedene Ansätze für eine mögliche Unterstützung von Redakteuren in Zusammenhang mit semantischen Technologien vorgestellt werden.

Die konkrete Implementierung einer dieser Ansätze für ein CMS wird umgesetzt und bildet die Grundlage für die anschließende Evaluierung. Die Ergebnisse dieser Evaluierung soll die Zufriedenheit der Anwender mit der entwickelten Anwendung ermitteln und verschiedene Verbesserungsvorschläge oder Möglichkeiten einer Weiterentwicklung aufdecken.

1.3 Inhaltlicher Aufbau

Der inhaltliche Aufbau dieser Arbeit besteht aus fünf Hauptteilen. Nach der Einleitung bietet Kapitel 2 eine Einführung in das Semantic Web. Die Grundlagen des Semantic Web geben nicht nur einen Einblick auf unterschiedliche Technologien, sondern behandeln auch verschiedene Schnittstellen, mit denen man auf Services des Semantic Web Zugriff erlangt. Des Weiteren wird in Kapitel 3 in die Grundlagen der Usability eingegangen. Hier werden besonders Usability-Guidelines und Methoden der Usability-Evaluation betrachtet.

In Kapitel 4 werden die in den beiden vorhergehenden Kapiteln behandelten Themen – Semantic Web und Usability – zusammengeführt. Es werden verschiedene Ansätze vorgestellt, welche durch den Einsatz des Semantic Web und dessen Technologien Redakteure bei ihren Tätigkeiten in einem CMS unterstützen können. Die Erstellung einer *Tag-Cloud* in Abschnitt 4.1 bedeutet für den Nutzer viel Aufwand, da hierfür Schlüsselwörter gefunden und gewichtet werden müssen. Ein weiteres Beispiel bietet die automatische *Textzusammenfassung* in Abschnitt 4.2. Auch die Eingliederung geschriebener Artikel in die *Navigation* in Abschnitt 4.3 ist eine der wiederkehrenden Aufgaben eines Redakteurs. Ein weiterer wichtiger Bestandteil stellt die semantische *Textanreicherung* in Abschnitt 4.4 dar, welche dem Nutzer nicht nur Zeit, sondern auch sehr viel Aufwand erspart. Bei allen Beispielen wird auch ein Überblick über bestehende wissenschaftliche Arbeiten zu dem jeweiligen Themenfeld gegeben.

Kapitel 5 bildet den praktischen Hauptteil dieser Arbeit. Hier wird auf das Beispiel der semantischen Textanreicherung näher eingegangen und ein Prototyp für das TYPO3 CMS vorgestellt. Im Anschluss daran erfolgt die Evaluierung des entwickelten Prototypen. Kapitel 6 beschreibt den Ablauf der Evaluierung genau. Dabei werden die Vorbereitung, Testumgebung, Testpersonen, eingesetzte Methoden und die daraus resultierenden Ergebnisse vorgestellt. In den Schlussbemerkungen in Kapitel 7 wird ein abschließendes Fazit gezogen und ein Ausblick auf mögliche zukünftige Forschungsarbeiten bzw. Weiterentwicklungen gegeben.

Kapitel 2

Semantic Web

Die Verbreitung und das rasante Wachstum des *World Wide Web* (WWW) bringt eine ständig wachsende Menge an Informationen und frei verfügbaren Daten mit sich, welche durch das Aufkommen des *Social Web* noch einmal enorm gestiegen ist und an Vielfalt gewonnen hat. Der Begriff des Web 2.0 hat sich etabliert und es besteht die Möglichkeit der sozialen Interaktion.

In der verfügbaren Form sind diese Daten allerdings auf den Menschen als Endnutzer ausgerichtet, welcher die Bedeutung einer Information erfassen und zu anderen Informationen in Beziehung setzen kann. Eine Maschine kann dies in der Regel nicht. Daraus ergibt sich das Problem, dass aus einer Fülle vorhandener Daten eine bestimmte Information nur schwer gefunden werden kann, bzw., dass Informationen nicht explizit im Web zu finden sind, sondern nur implizit aus der Schlussfolgerung gegebener Fakten gegeben ist [20].

Aus dieser Problematik heraus, ergab sich die Idee des Semantic Web. Tim Berners-Lee, Begründer des WWW, hat im Jahr 1999 seine Vision über ein Semantic Web geprägt. Diese beschrieb er wie folgt [4, Kap. 12]:

> I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A "Semantic Web", which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The "intelligent agents" people have touted for ages will finally materialize.

Das Semantic Web hat das Ziel, von Menschen erstellte Inhalte zu maschinenverwertbaren und interpretierbaren Informationen zu verarbeiten und diese zu strukturieren und sinngemäß zu verknüpfen. So wird es ermöglicht, besser nach brauchbaren, an die Bedürfnisse der Nutzer angepassten Informationen aus der Menge an Daten zu filtern.

Das Semantic Web ist dabei kein separates Web, sondern eine Erweite-

rung des aktuellen Webs. Webinhalte werden mit Metadaten angereichert, damit Maschinen in einer Art und Weise mit den Informationen umgehen können, die aus menschlicher Sicht sinnvoll und nützlich erscheint. Somit wird eine Zusammenarbeit zwischen Mensch und Computer ermöglicht [5, 20].

2.1 Konzepte

Das Semantic Web beruht auf zwei grundlegenden Konzepten. Dazu zählen *Metadaten*, die zur Beschreibung der Informationsressourcen eingesetzt werden und *Ontologien* zur Strukturierung und Darstellung komplexer Wissensbeziehungen. Wichtig dabei ist auch die Unterscheidung zwischen *Syntax* und *Semantik*.

Unter Syntax versteht man im Allgemeinen die Struktur von Daten. Eine Menge von Regeln dient der Strukturierung von Zeichen und Zeichenketten. Semantik steht hingegen für die Bedeutung von Wörtern, Phrasen oder Symbolen. Im Semantic Web beschreibt sie insbesondere die Bedeutung von Zeichen(-ketten) und deren Beziehungen untereinander [20].

2.1.1 Metadaten

Webinhalte werden aktuell für den Menschen aufbereitet und mittels (X)-HTML oder als PDF-Dokument ausgegeben. Maschinen können dabei nicht klar erkennen, dass es sich beispielsweise bei einer Datumsangabe auch um eine solche handelt. Metadaten sind Daten über Daten und bilden die Grundlage zur Beschreibung von Semantik im Web. Sie werden eingesetzt um Informationsressourcen zu beschreiben und Beziehungen zwischen den Ressourcen herzustellen [8].

Das Hinzufügen solcher Metadaten bezeichnet man als „auszeichnen“ oder „annotieren“ [20]. Im Semantic Web werden Metadaten mit Hilfe von *RDF* strukturiert, worauf in Abschnitt 2.2.1 näher eingegangen wird.

2.1.2 Ontologien

Im Kontext des Semantic Web sind Ontologien Informationssammlungen, mit deren Hilfe komplexe Wissensbeziehungen in Form eines Netzes aus Hierarchien dargestellt werden.

Ontologien wurden im Umfeld der künstlichen Intelligenz entwickelt und stellen eine Basiskomponente des Semantic Web dar. Mit ihnen kann Wissen einer Domäne modelliert und unabhängig von Programmen wiederverwendet werden. Ontologien beschreiben also Klassen einer Wissensdomäne und die Beziehungen zwischen diesen Klassen. Somit wird ermöglicht, dass Maschinen Webinhalte interpretieren können [8].

Nach Blumauer und Pellegrini [8, S. 12] werden Ontologien entwickelt und eingesetzt, um

- den Datenaustausch zwischen Programmen zu ermöglichen,
- die Vereinheitlichung und Übersetzung zwischen verschiedenen Wissensrepräsentationsformen zu ermöglichen,
- Services zur Unterstützung von Wissensarbeitern zu entwickeln,
- Theorien abzubilden,
- die Semantik strukturierter und semi-strukturierter Information auszudrücken und
- die Kommunikation zwischen Menschen zu unterstützen und zu erleichtern.

2.2 Technologien

Das Semantic Web basiert auf einer Reihe von Technologien, zu denen Wissensrepräsentationssprachen für Ontologien und Methoden und Werkzeuge zur Erstellung, Wartung und Anwendung von Ontologien gehören. Interoperabilität ist grundlegend für ein Semantic Web, das heißt, einheitliche Standards einzuführen, um Informationen zwischen verschiedenen Plattformen austauschen zu können. Die Standards müssen dabei flexibel und erweiterbar sein um möglichst alle Anwendungsfälle abbilden zu können [20].

Das *World Wide Web Consortium*[1] (W3C), Gremium zur Standardisierung der WWW betreffenden Technologien, hat bereits Standards für beispielsweise XML, RDF, RDFS und OWL geschaffen, welche im Semantic Web Stack in Abbildung 2.1 dargestellt werden. Diese bilden auch die Basis-Sprachen zur Repräsentation des Semantic Web.

2.2.1 RDF

Mittels XML können Daten flexibel in einer einheitliche Sprache strukturiert werden. XML-Namespaces erlauben es, XML-Vokabulare weltweit eindeutig zu definieren und mit XML-Schemata kann die Syntax von Vokabular detailliert festgelegt werden. Um Informationen über diese Daten selbst anzureichern und in Beziehung zu setzen, ist die hierarchische Struktur von XML schlecht geeignet [6].

Das *Resource Description Framework* (RDF) ist eine formale Sprache zur Beschreibung strukturierter Informationen und ist ein grundlegender Baustein des Semantic Web. Es beschreibt die Beziehung zwischen Ressourcen und ermöglicht so, Daten nicht nur strukturiert darzustellen sondern diese auch mit anderen Informationen in Beziehung zu setzen. Diese Beschreibung wird von RDF durch eine *Aussage* oder *Triple* (englisch: *statement*)

[1] W3C http://www.w3.org/

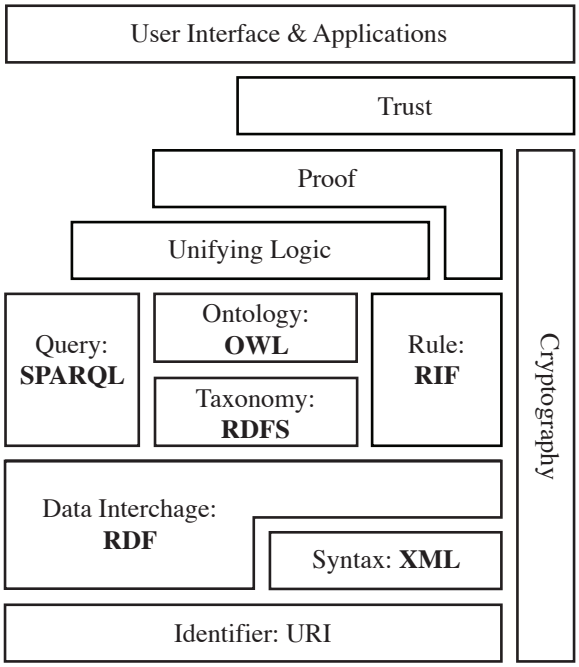

Abbildung 2.1: Semantic Web Stack, nach [63].

ermöglicht. Ein Triple besteht aus einem Subjekt, einem Prädikat und einem Objekt. Das Objekt kann dabei entweder ein Wert (Literal) oder eine weitere Ressource und Subjekt für weitere Triple sein. Durch die Verknüpfung dieser Triple erhält man einen gerichteten Graphen. Die Knoten eines RDF-Graphen bilden dabei Subjekt und Objekt, welche durch eine gerichtete Kante (Prädikat) verbunden sind. Eine Ressource wird über eine eindeutige Namensgebung – einem *Uniform Resource Identifier* (URI) – identifiziert. Durch diese URIs wird es ermöglicht, Aussagen aus verschiedenen Quellen zu verbinden. Ein Beispiel für einen RDF-Graph ist in Abbildung 2.2 zu sehen. Literale werden dabei rechteckig umrandet um sie von den oval umrandeten URIs zu unterscheiden. Der abgebildete RDF-Graph sagt aus, dass die Ressource „http://example.org/SemanticWeb“ den Titel „Semantic Web – Grundlagen“ hat und diese Ressource beim Verlag mit dem Namen „Springer-Verlag“ verlegt wird [6, 20].

2.2.2 RDF-Schema

Mit RDF können bereits einfache Aussagen über Ressourcen gemacht werden. Um mit diesen Aussagen eine Ontologie aufzubauen, benötigt man allerdings Möglichkeiten, Begriffe zu gruppieren und Aussagen darüber zu machen. *RDF Vocabulary Description Language* (RDFS oder RDF-Schema), eine weitere W3C-Empfehlung, bildet eine semantische Erweiterung von RDF

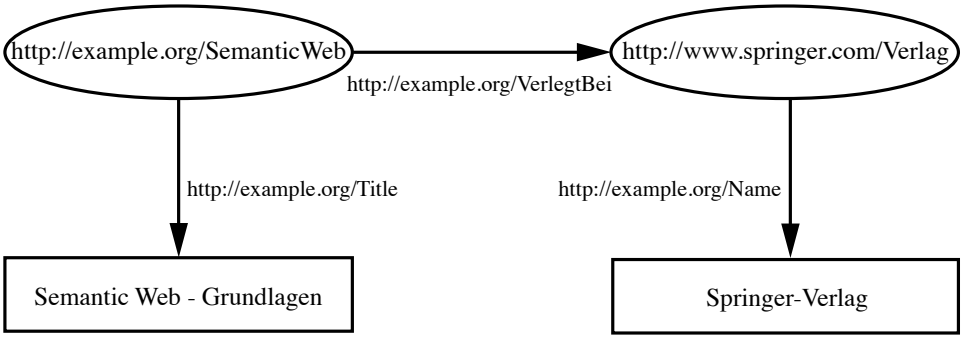

Abbildung 2.2: Ein RDF-Graph mit Literalen zur Beschreibung von Datenwerten, nach [20].

und definiert ein Vokabular zur Beschreibung der Eigenschaften und Klassen von RDF-Ressourcen. RDFS erlaubt es in RDF-Klassen von Objekten und Subklassen von Klassen zu definieren. Außerdem können Eigenschaften, Werte- und Gültigkeitsbereiche zugeordnet werden [6, 55, 20].

2.2.3 OWL

RDF kann in Kombination mit RDFS einfache Ontologien abbilden, ist jedoch für komplexere Modellierungen bzw. für die detaillierte Darstellung von komplexem Wissen nicht immer ausreichend. Die *Web Ontology Language* (OWL) ist eine Vokabular-Erweiterung zur Beschreibung von Eigenschaften und Klassen und hat durch ihre Ausdrucksmächtigkeit mehr Möglichkeiten Inhalt und Semantik auszudrücken, als dies von XML, RDF oder RDFS geboten wird. Durch Durchschnitts-, Vereinigungs- oder Komplementäroperationen können Klassen miteinander verknüpft werden, um daraus neue zu erzeugen.

Das Ziel von OWL ist es, das Problem zu lösen [6, S. 83],

- Terminologien für einen bestimmten Zusammenhang zu erstellen,
- Eigenschaften besser einschränken zu können,
- logische Charakteristiken von Eigenschaften und
- die Äquivalenz von Begriffen zu definieren.

OWL stellt drei zunehmend ausdrucksstarke Teilsprachen bereit: *OWL Lite*, *OWL DL* und *OWL Full*. Diese drei Teilsprachen sind entwickelt worden, um den Anwendern je nach Anforderungen die Wahl zwischen verschiedenen Ausdrucksstärken zu ermöglichen [6, 59].

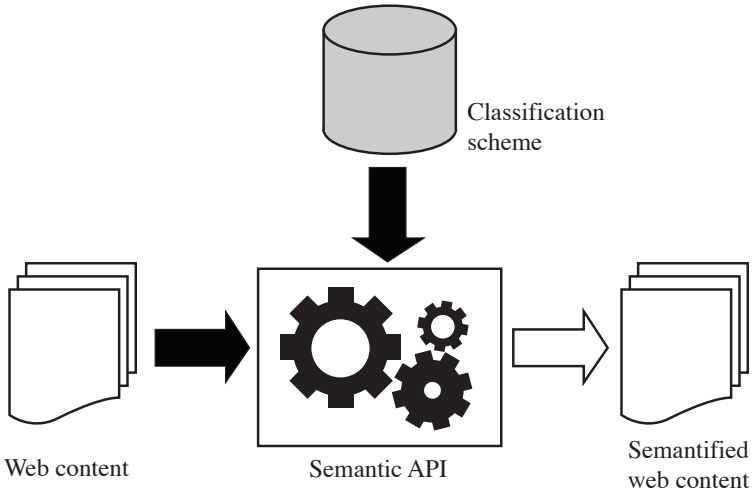

Abbildung 2.3: Semantic APIs [13].

2.3 Semantic APIs

Ohne eine semantische Auszeichnungssprache und die Unterstützung von RDF und OWL kann der Zugriff auf Informationen über Inhalten auch über *Applications Programmers Interfaces* (APIs) ermöglicht werden. Semantic APIs werden von Dotsika [13, S. 337] folgendermaßen beschrieben:

> APIs that take unstructured text (including web pages) as input and return the content's contextual framework are termed semantic APIs.

Abbildung 2.3 zeigt die Funktion von Semantic APIs. Texte werden dabei mit semantischen Metadaten versehen und als XML-Code oder RDF zurückgeliefert [13].

In den letzten Jahren stieg die Zahl der Anbieter von Semantic APIs im WWW. Diese Dienste bieten eine Textanalyse und arbeiten mit Methoden des *Natural Language Processing* (NLP). NLP erforscht die Verarbeitung natürlicher Sprache mit Hilfe von Computersystemen. Ein Teilbereich von NLP ist die *Informationsextraktion*. Bei einer Informationsextraktion werden bestimmte Fakten aus unstrukturierten Daten gewonnen. Dieser Prozess bietet eine Möglichkeit wertvolle Metadaten zu generieren [21].

Ein wichtiger Bereich der Semantic APIs ist die *Entitätenextraktion*, welche auch als *Eigennamenerkennung* bekannt ist. Sie ist ein Teilbereich der Informationsextraktion. Dieser Prozess hat das Ziel, aus unstrukturierten Texten Eigennamen zu identifizieren. Spezielle Ausdrücke wie Personen-, Firmen- und Produktnamen, aber auch komplexe Datums- oder Zeitausdrücke werden dabei gefunden. Besonders wichtig ist die Behandlung von

Referenzen zwischen den Eigennamen. So kann sichergestellt werden, dass die Person „Barack Obama“, „Präsident Obama“ oder „er“ in einem zusammengehörigen Text nicht als mehrere Personen erkannt wird [36].

2.3.1 Anbieter

Beispiele für Anbieter von semantischen Textanalysen sind *Alchemy API*[2], *SemantLink* [17], *Zemanta*[3] oder *OpenCalais*[4], welche unterschiedliche Arten von Metadaten anbieten und im Folgenden vorgestellt werden.

AlchemyAPI

AlchemyAPI setzt Natural Language Processing und Machine Learning Algorithmen ein, um semantische Metadaten zu extrahieren. So können Informationen über Personen, Orte, Unternehmen, Themen, Fakten, Beziehungen, Autoren und Sprachen gefiltert werden. Durch die Textkategorisierung können Webinhalten Themenkategorien wie zum Beispiel Sport oder Wirtschaft zugewiesen werden.

SemantLink

Das Projekt SemantLink [17] bietet eine Entitätenextraktion sowie eine Relevanzbewertung dieser. Dabei wird sowohl eine Globalrelevanz als auch eine Kontextrelevanz berechnet. Die dafür benötigten Informationen basieren auf *DBpedia, Facebook Open Graph* und der *Google AdWords API*. SemantLink liefert Entitäten wie Personen, Orte, Organisationen, Berufe, Geldbeträge und Datumsangaben.

Zemanta

Auch Zemanta bietet eine Textanalyse. Es analysiert von Nutzern generierte Texte mittels Natural Language Processing und semantischen Suchtechnologien. Die Zemanta API liefert Entitäten, Bilder, verwandte Artikel, Links und Tags für die Verlinkung. Wikipedia und Flickr bilden dabei die Quelle für die vorgeschlagenen Bilder [62].

OpenCalais

OpenCalais bietet einen Web Service, welcher automatisch semantische Metadaten für Webinhalten generiert. Der Service basiert auf Natural Language Processing, Machine Learning und anderen Methoden. Neben der Extraktion von Entitäten werden auch noch Fakten und Events aus unstrukturierten Dokumenten extrahiert [57].

[2] Alechemy API, http://www.alchemyapi.de
[3] Zemanta, http://www.zemanta.com/
[4] OpenCalais, http://www.opencalais.com/about

Kapitel 3

Usability

Beim Thema Usability, von dem im Deutschen auch von Gebrauchstauglichkeit oder Benutzerfreundlichkeit gesprochen wird, geht es nicht um eine einzelne Eigenschaft einer Benutzeroberfläche. Es steht dafür, wie gut ein Benutzer ein Produkt in seinem Umfeld zur Bewältigung seiner Aufgabe einsetzen kann. Die Benutzbarkeit eines Systems muss also im Kontext seiner Verwendung beurteilt werden, wie in Abbildung 3.1 grafisch dargestellt wird [44].

Von Jakob Nielsen und Hoa Loranger wird Usability wie folgt definiert [39, S. xvi]:

> Usability ist ein Qualitätsmerkmal, wie einfach etwas zu benutzen ist. Es geht genauer gesagt darum, wie schnell Menschen

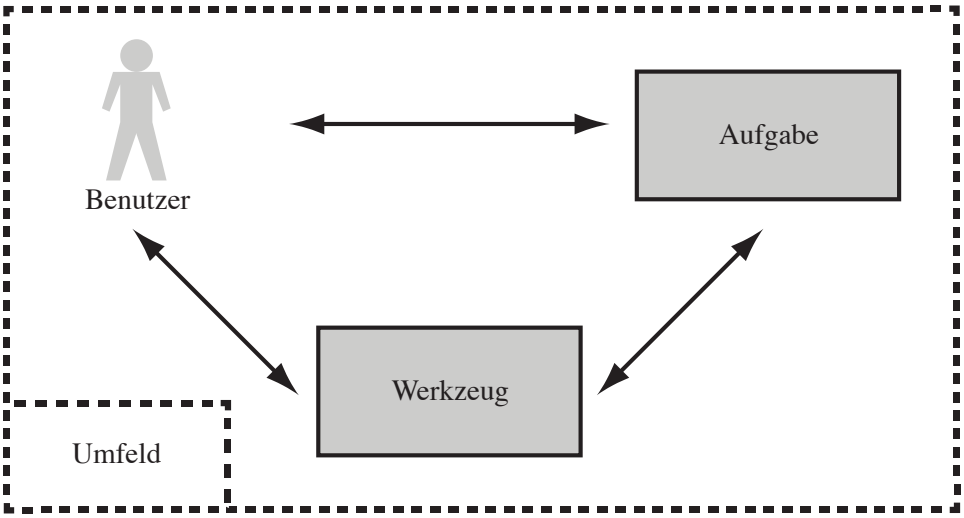

Abbildung 3.1: Vier Komponenten eines Mensch-Maschine-Systems nach [44].

> die Benutzung eines Gegenstands erlernen können, wie effizient sie während seiner Benutzung sind, wie leicht sie sich diese merken können, wie fehleranfällig der Gegenstand ist und wie er den Nutzern gefällt. Wenn die Nutzer einen Gegenstand weder nutzen möchten noch können, bräuchte er eigentlich gar nicht zu existieren.

Usability beschäftigt sich damit, Systeme möglichst einfach erlernbar und benutzbar zu gestalten. Nach Jakob Nielsen [38] ist Usability durch 5 Qualitätskomponenten definiert:

- *Erlernbarkeit:* Das System soll einfach zu erlernen sein, damit der Nutzer schnell grundlegende Aufgabenstellungen bewältigen kann.
- *Effizienz:* Das System soll effizient genutzt werden können. Der Benutzer sollte nach dem Erlernen des Systems möglichst produktiv damit arbeiten können.
- *Einprägsamkeit:* Benutzer sollen nach einem bestimmten Zeitraum der Nichtbenutzung ihre Kenntnisse über das System wiederherstellen können und das System nicht von Grund auf wiedererlernen müssen.
- *Fehlerrate:* Das System soll eine geringe Fehlerrate aufweisen und Aktionen, die nicht zum Ziel führen, sollen minimiert und mit geeigneten Fehlermeldungen ausgestattet werden.
- *Zufriedenheit:* Die subjektive Zufriedenheit des Anwenders ist wesentlich, da nur so gute Ausgangsbedingungen für ein erfolgreiches Lernen geschaffen werden können.

Die Definition des Begriffs Usability ist auch in einer ISO-Norm[1] festgehalten. Die DIN EN ISO 9241-11 definiert Gebrauchstauglichkeit als das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Kontext verwendet werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.

3.1 Usability-Guidelines

Das Wissen zur Erreichung einer guten Usability wird in unterschiedlichen Formen von Usability-Guidelines zusammengefasst. Diese zahlreichen Guidelines können aufgrund ihrer Art und ihres Ursprungs klassifiziert werden. Nach Mariage et al. [29] werden diese – wie in Abbildung 3.2 dargestellt – in *Principles*, *Guidelines* und *Recommendations* eingeteilt, wobei Principles eher allgemein gehaltene Guidelines darstellen und bis hin zu den Recommendations immer spezifischer werden.

[1] International Organization for Standardization, http://www.iso.org/

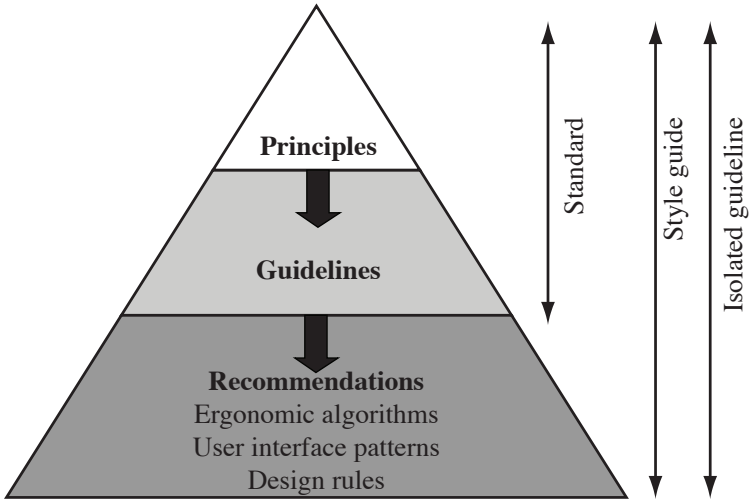

Abbildung 3.2: Klassifizierung von Guidelines nach Art und Ursprung [29].

3.1.1 Design-Grundsätze

Principles sind Leitsätze oder Grundsätze, welche bei der konzeptionellen Designentscheidung helfen. Sie sind sehr allgemein gehalten, damit sie in einer Vielzahl von Fällen angewendet werden können [29].

Ein Beispiel für Design-Principles sind die 10 Heuristiken von Jakob Nielsen zur Erreichung eines guten User Interface-Designs [60]. Diese beinhalten allgemeine Regeln, wie beispielsweise die Information über den Systemstatus:

> The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

Eine weitere Heuristik betrifft die Konsistenz und das Einhalten von Standards:

> Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

3.1.2 Gestaltungsrichtlinien

Basierend auf den Principles existieren *Guidelines*, welche sich mit einem bestimmten Bereich des Webdesigns beschäftigen.

3.1.3 Konventionen

Recommendations oder Konventionen sind eindeutig und lassen keinen Spielraum für Interpretationen. Sie werden von bestehenden Principles und Gui-

delines unter Berücksichtigung der Anforderungen und des Nutzungskontextes abgeleitet. Sie beziehen sich ebenfalls auf einen bestimmten Bereich des Designs. Zu den Recommendations zählen auch ergonomische Algorithmen und Gestaltungsregeln [29].

3.1.4 Standards

Standards sorgen für eine Vereinheitlichung von User Interfaces. Sie entsprechen den Anforderungen der von verschiedenen Organisationen herausgegebenen Normen.

Diese Normen definieren Usability präzise und stellen einen Konsens über gebündeltes Fachwissen von Usability-Experten dar. Außerdem lassen sich daraus Anforderungen and die Systemgestaltung als auch Prüfkriterien für Usability-Evaluationen ableiten [47].

Bekannte Organisationen sind dabei das W3C oder die ISO. Diese bietet den Standard DIN EN ISO 9241, welcher Richtlinien der Ergonomie der Mensch-System-Interaktion beinhaltet.

Des Weiteren gibt es Richtlinien für barrierefreie Webinhalte in den *Web Content Accessibility Guidelines (WCAG) 2.0* [56], welche Hinweise für eine barrierefreie Gestaltung für Webinhalte für Menschen mit Behinderungen geben. Die Richtlinien gliedern sich dabei in wahrnehmbar, bedienbar, verständlich und robust. Beispiele dafür sind, dass für alle Nicht-Text-Inhalte Text-Alternativen zur Verfügung gestellt werden oder, dass alle Funktionalitäten von der Tastatur aus verfügbar sein sollen.

3.2 Methoden der Usability-Evaluation

Unter Usability-Evaluation werden alle Analysen oder empirischen Studien der Usability von Produkten oder Prototypen verstanden. Dies hat zum Ziel, mögliche Usability-Probleme aufzudecken, also Aspekte des Produkts, die den Benutzer vor ein Problem stellen könnten. Im Zuge einer Usability-Evaluation kann somit überprüft werden, ob bzw. inwiefern ein Produkt den Ansprüchen und Erwartungen eines Benutzers gerecht wird [46].

Basierend auf Scriven [50] werden von Rosson und Carroll [46] Evaluationsmethoden in eine formative und eine summative Evaluation unterteilt.

Eine *formative Evaluation* bewertet das Produkt während der Entwicklung. Es sollen bereits in der Entwicklungsphase frühzeitig Fehler oder Mängel aufgedeckt und eliminiert werden.

Die *summative Evaluation* hingegen bewertet das Produkt nach seiner Fertigstellung. Ziel dabei ist die Messung der Qualität. Es beantwortet Fragen wie „Hat das Produkt die spezifizierten Ziele erreicht?“ oder „Ist das Produkt besser als seine Vorgänger oder Mitbewerber?“.

Des Weiteren kann bei Evaluationsmethoden zwischen analytischen und empirischen Evaluierungen unterschieden werden, welche im Folgenden näher

beschrieben werden.

3.2.1 Analytische Evaluationsverfahren

Die expertenorientierte Methode wird in der Usability-Literatur auch *Usability Inspection* genannt, während sie in der Psychologie-Literatur als analytische Methode bezeichnet wird [49].

Bei dieser Methode werden Experten als Gutachter herangezogen. Sie haben den Vorteil, dass man sie relativ schnell und ohne großen Aufwand durchführen kann. Außerdem können die als Ersatz-Anwender eingesetzten Experten ihr Fachwissen und ihre Erfahrung einbringen. Diese können sich jedoch nicht immer vollständig in die Situation der Benutzer hineindenken, was einen Nachteil dieser Methoden darstellt [49].

Beispiele für analytische Evaluationsverfahren sind der *Cognitive Walkthrough* und die *heuristische Evaluation.*

Cognitive Walkthrough

Bei einem Cognitive Walkthrough handelt es sich um eine aufgabenorientierte Methode, welche die Erlernbarkeit eines Produktes ermittelt. Die Idee dazu beruht auf der Beobachtung, dass viele Benutzer den Umgang mit einem Produkt durch Ausprobieren erlernen. Beim Cognitive Walkthrough wird das problemlösende Explorieren eines Produktes durch den Benutzer beobachtet. Dafür versetzt sich ein Usability-Experte in die Rolle eines potentiellen Benutzers und analysiert im Vorfeld definierte Aufgaben. Es wird festgestellt, ob diese Aufgaben für den Benutzer lösbar wären oder nicht. Mit den resultierenden Aufzeichnungen können dem Designer oder dem Entwickler aufgezeigt werden, wo und warum Beeinträchtigungen auftreten können bzw. wo das Produktdesign negative Auswirkungen auf die Interaktion mit den Benutzern haben könnte [49, 53].

Heuristische Evaluation

Die heuristische Evaluation begutachtet das Produkt umfassender und systematischer als die Methode des Cognitive Walkthrough. Dabei wird die Benutzerschnittstelle eines Produkts von einer Gruppe von Gutachtern auf die Einhaltung sogenannter *Heuristiken* überprüft.

Heuristiken sind Richtlinien und Prinzipien, die auf Grundlage empirischer Erkenntnisse für eine benutzerfreundliche Gestaltung einer Bedienoberfläche entwickelt wurden (siehe Abschnitt 3.1) [49].

Die heuristische Evaluation ist einfach und schnell durchführbar. Studien bzw. Projekte haben gezeigt, dass verschiedene Personen unterschiedliche Usability-Probleme aufdecken, weshalb es für einen einzelnen Gutachter schwierig ist, alle Usability-Probleme zu erkennen. So kann durch das Hinzu-

ziehen mehrerer Gutachter die Effektivität der Methode signifikant gesteigert werden [37].

Gutachter sollen dabei die Benutzerschnittstelle in zwei Durchgängen analysieren. Im ersten Durchgang sollen sie sich auf den Ablauf und im zweiten auf die einzelnen Elemente konzentrieren. Bei der Durchführung der Methode muss jeder Gutachter zunächst die Benutzerschnittstelle individuell analysieren, erst nach deren Abschluss dürfen sich die Gutachter darüber austauschen.

Heuristische Evaluierungen erzeugen eine Liste von möglichen Problemen, welche lediglich als Vorschläge anzusehen sind. Was bei einer Usability Evaluation essentiell ist, ist herauszufinden, was passiert, wenn Benutzer das Produkt in einer realen Situation verwenden [46].

3.2.2 Empirische Evaluationsverfahren

Anders als die analytischen Methoden werden die empirischen Methoden, welche in der Usability-Literatur auch als benutzerorientierte Methoden oder Usability Testing bezeichnet werden, mit tatsächlichen Benutzern durchgeführt. Deshalb gehören sie zu den aufwändigsten, aber auch wertvollsten Methoden, da man Feedback direkt vom Benutzer und nicht von Experten als Ersatz-Benutzer erhält [49].

Die verschiedenen Methoden des benutzerorientierten Verfahrens unterscheiden sich darin, was sie dokumentieren bzw. messen. Mit Hilfe von Interviews oder Fragebögen kann das Verhalten über potenzielles oder tatsächliches Verhalten dokumentiert werden. Mittels Beobachtungen kann hingegen tatsächliches Verhalten in einer konkreten Situation festgehalten werden. Befragungen sollten daher immer in Kombination mit anderen Methoden eingesetzt werden [49].

Benutzerbefragungen

Befragungen können in verschiedenen Phasen des Designprozesses erfolgen. Diese können mit Hilfe von Fragebögen oder Interviews durchgeführt werden.

Da das Aufbereiten eines Fragebogen sehr komplex ist, kann auf erprobte Fragebögen, wie den Benutzerfragebogen *ISONORM 9241/10*, zurückgegriffen werden. Dieser Benutzerfragebogen baut auf einen Normentwurf der ISO auf und enthält sieben Grundsätze: Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlerrobustheit, Individualisierbarkeit und Erlernbarkeit [41, 42].

Usability-Test

Ein Usability-Test wird eingesetzt, um herauszufinden, ob das entwickelte Produkt für die Zielgruppe gebrauchstauglich ist und den vorgesehenen Zweck erfüllt. Die Durchführung findet dabei häufig in Usability-Laboren

statt. Methoden sind beispielsweise *Lautes Denken* oder Fragebögen. Bei der Methode des Lauten Denkens werden die Testpersonen gebeten, alles was sie während des Tests denken und tun, laut auszusprechen. Dadurch lassen sie den Versuchsleiter verstehen, wie sie mit dem Produkt interagieren und wo sie auf Probleme stoßen [49].

Jakob Nielsen bezeichnet Lautes Denken als wertvollste Methode des Usability Enginieering [38, S. 195]:

> By verbalizing their thoughts, the test users enable us to understand how they view the computer system, and this again makes it easy to identify the users' major misconceptions. One gets a very direct understanding of what parts of the dialogue cause the most problems, because the thinking-aloud method shows how users interpret each individual interface item.

Fokusgruppen-Interview mit Benutzern

Fokusgruppen-Interviews sind eine qualitative Forschungsmethode um die Gedanken, Meinungen und Wahrnehmungen der Teilnehmer über ein bestimmtes Produkt, einen bestimmten Service oder eine Idee zu erörtern. Dies geschieht in einer Diskussionsgruppe, welche durch einen Moderator betreut und angeregt wird und sich meistens an einem Leitfaden orientiert. Der Forscher sorgt für eine bequeme und permissive Umgebung, um die Teilnehmer zu ermutigen und anzuregen, ihre Vorstellungen bzw. Auffassungen und Ansichten zu teilen, ohne sie dabei zu einer Entscheidung oder zu einen Konsens mit der Gruppe zu drängen. Für ein Projekt sollten immer drei bis vier Fokusgruppen gehalten werden, um mögliche Trends oder Muster identifizieren zu können. Die Gruppe sollte dabei aus zirka sechs bis acht Personen bestehen. Sie sollte also klein genug sein, damit jeder Teilnehmer die Möglichkeit hat, sich einzubringen. Die Fokusgruppe wird von einem qualifizierten Moderator abgehalten und geleitet. Die Diskussionen sind eher entspannt. Den Teilnehmern sollte es Vergnügen bereiten, ihre Ansichten preiszugeben. Sie beeinflussen damit die anderen Teilnehmer und werden wiederum von denen beeinflusst und angeregt [23, 35].

Kapitel 4

Ansätze einer Redakteursunterstützung in CMS durch das Semantic Web

Setzt man Technologien des Semantic Web bzw. Semantic APIs in CMS ein, können Redakteure bei ihren Aufgaben bereits auf verschiedenste Weise unterstützt werden. Einige dieser Bereiche werden im Folgenden vorgestellt. Begonnen wird mit der automatischen Erstellung einer Tag-Cloud, welche nicht nur Redakteuren, sondern auch Benutzern einer Website Unterstützung bzw. einen Mehrwert bietet. Außerdem werden Methoden einer automatischen Textzusammenfassung dargestellt. Des Weiteren werden Ansätze einer automatischen Eingliederung von Artikel in die bestehende Navigation aufgezeigt. Der letzte Punkt bildet die Textanreicherung mit Hilfe einer semantischen Textanalyse, auf welche auch das Anwendungsbeispiel in Kapitel 5 aufbaut.

4.1 Erstellung einer Tag-Cloud

Eine Schlagwortwolke bzw. *Tag-Cloud* ist eine Art der Informationsvisualisierung. Dabei werden Schlagwörter meist alphabetisch sortiert aufgelistet und variieren je nach Gewichtung in ihrer Schriftgröße [18].

4.1.1 Tagging

Tagging beschreibt das Auszeichnen von Artikeln mit frei gewählten Schlagwörtern, sogenannten Tags. Es handelt sich dabei um eine gemeinschaftliche, nicht hierarchische, freie Verschlagwortung von Begriffen. Das wichtigste dabei ist, dass die Inhalte später durch den Tag-Namen einfach wiedergefunden werden können [30]. Ein Beispiel für eine Social Tagging-Seite bietet Flickr[1],

[1] Flickr http://www.flickr.com/

auf der die Benutzer Fotos mit verschiedenen Tags versehen können.

4.1.2 Folksonomien und Tag-Clouds

Ergebnis des Social Taggings ist eine sogenannte *Folksonomie*. Die Visualisierung einer Folksonomie kann beispielsweise durch eine gewichtete Liste – eine Tag-Cloud – erfolgen. Eine Tag-Cloud visualisiert also die Tags einer Folksonomie. Je größer ein Tag angezeigt wird, desto häufiger wurde dieser verwendet [18, 52].

4.1.3 Eigenschaften und Usability von Tag-Clouds

Laut Rivadeneira et al. [45] gibt es verschiedene Aufgaben, bei denen Tag-Clouds Anwender unterstützen können:

- *Suche:* Mit Hilfe einer Tag-Cloud kann dem Nutzer die Navigation zu einem bestimmten Inhalt erleichtert werden, indem auch Tags mit in die Suche eingebunden werden, die das gesuchte Konzept abbilden.
- *Browsing:* Nutzern, die kein spezifisches Suchergebnis ansteuern, bietet die Tag-Cloud einen ersten Einstiegspunkt zum Stöbern.
- *Überblick (Impression Formation or Gisting):* Die Tag-Cloud kann dem Nutzer einen Eindruck über die zugrunde liegenden Daten bieten. Somit bildet diese einen Überblick über häufig und weniger häufig auftretende Themen.
- *Erkennung/Übereinstimmung:* Der Nutzer kann durch die Tag-Cloud erkennen, welche verschiedenen Aspekte von Informationen repräsentiert werden.

Sinclair und Cardew-Hall [52] bestätigten durch ihre Studie die Erfüllung dieser Funktionen. Als weiteren Vorteil von Tag-Clouds ermitteln sie, dass das Klicken auf einen Tag den Nutzern leichter fällt, als selbst einen Suchanfrage zu formulieren.

Neben diesen positiven Eigenschaften gibt es allerdings auch negative Aspekte. Hearst und Rosner [19] beschreiben, dass die Angabe der Bedeutung eines Wortes durch die Größe zu Problemen führen kann. Die Länge des Wortes ist mit seiner Größe verschmolzen wodurch die Wörter mit mehr Zeichen bedeutender erscheinen. Des Weiteren bemängeln sie, dass neue Themenbereiche eine gewisse Zeit benötigen, bis sie in der Tag-Cloud abgebildet werden.

In den folgenden Beispielen werden Ansätze erläutert, die sich besonders mit der Herausforderung der fehlenden Semantik zwischen den Tags beschäftigen und versuchen, diesem Problem entgegenzuwirken.

4.1.4 Beispiele

Kim et al. [22] beschreiben 2008 sogenannte „Tag Ontologien“. Foksonomien haben das Problem, dass die Tags sprachlichen Unklarheiten unterliegen, wenig oder keine Semantik besitzen und es viele Variationen gibt. Tags können frei und ohne Regeln erstellt werden und haben somit keine exakte Bedeutung und keinen Bezug zu einer strukturierten Ontologie. Innerhalb der Tags gibt es sowohl linguistische und grammatikalische Unterschiede als auch menschliche Fehler, wie beispielsweise Tippfehler. Kim et al. untersuchen genau diese Probleme, indem sie verschiedene Ansätze kollaborativer Tagging-Aktivitäten auf einer semantischen Ebene diskutieren. Sie sehen dabei Semantic Web-Technologien als Ergänzung zu Folksonomien.

2009 evaluieren Schrammel et al. [48] die Effekte von semantischer versus alphabetischer und zufälliger Anordnung von Tags in einer Tag-Cloud. Die Ergebnisse zeigen, dass semantisch geclusterte Tag-Clouds eine Verbesserung gegenüber zufälliger Anordnung bei gezielten Suchaufgaben bieten. In ihrer Studie wollten sie herausfinden, wie die Anordnung der Tags die Suchdauer, die Wahrnehmung der Tag-Cloud, die subjektive Zufriedenheit der Anwender – sowohl beim Suchen eines bestimmten Tags als auch beim Recherchieren allgemeiner Themen – und die Merkfähigkeit von Tags beeinflusst.

Mirizzi et al. [34] beschreiben in ihrer Arbeit „Semantic tag cloud generation via DBpedia“ ein neues hybrides Ranking-System. Dafür verwenden sie semantische Informationen aus DBpedia, um so die Anfragen der Benutzer zu erweitern und Schlüsselwörter zu bewerten. Inputs dieses Systems bilden zum einen die DBpedia Datensätze und zum anderen externe Informationsquellen wie klassische Suchmaschinen-Ergebnisse und Social Tagging-Systeme. Viele Empfehlungssysteme sind text-basiert und stoßen durch das Fehlen von semantischen Beziehungen zwischen den Schlüsselwörtern schnell an ihre Grenzen. Sie können unterschiedliche Schlüsselwörter mit derselben Bedeutung (Synonymie), ebenso wie Wörter mit mehreren Bedeutungen (Polysemie) nicht erkennen bzw. zuordnen. Weitere Probleme stellen Homonymie (Wörter, die für verschiedene Begiffe stehen), Kontextanalyse oder das Erkennen von Beziehungen zwischen Schlüsselwörtern als Hyponym (=Unterbegriff) oder Hyperonym (=Oberbegriff) dar. Um diesen Problemen entgegenzuwirken werden Semantik-basierte Systeme eingesetzt. Mirizzi et al. sehen bei vielen bestehenden Ansätzen in diesem Bereich die Hauptherausforderung darin, die eingesetzte Ontologie aktuell zu halten. Dieser Prozess ist mühsam und aufwändig. Ihrer Arbeit wollen sie das Problem mit Hilfe von DBpedia lösen.

4.2 Automatische Textzusammenfassung

Aufgrund der stetig wachsenden Informationsflut werden Zusammenfassungen immer wichtiger. Das Durchsuchen der bestehenden Menge an Daten

stellt einen enormen Zeitaufwand dar und lässt Zusammenfassungen daher nicht mehr aus unserem Alltag wegdenken. Sie sind in vielen unterschiedlichen Bereichen in verschiedenen Formen aufzufinden. Sowohl Klappentexte bei Büchern als auch Überschriften in Zeitungen oder Programmvorschauen zählen dazu.

Auch im redaktionellen Bereich können automatische Zusammenfassungen gut eingesetzt werden. Redakteure sind mit dieser Aufgabe häufig beschäftigt und fassen beispielsweise Dokumente zusammen, um dann auf diese zu verweisen.

Mit Hilfe von Zusammenfassungen kann eine übersichtliche Darstellung umfangreicher Dokumente erzielt werden. Große Informationsmengen können auf deren wichtigsten Inhalte reduziert werden. So entsteht die Möglichkeit, Informationen rasch zu erfassen. Effektive Entscheidungen können so in kürzerer Zeit getroffen werden. Eine Textzusammenfassung sollte für den Benutzer interessante Informationen enthalten. Redundante und belanglose Informationen sollten dabei ausgeschlossen werden. Außerdem sollte sie kohärent und verständlich formuliert sein. Somit kann eine Zusammenfassung des selben Quelldokuments je nach Ausmaß der Anpassung für den Nutzer und den Kondensationsanforderungen zu verschiedenen Ergebnissen führen [16, 27].

Mani beschreibt in seinem Buch „Automatic Summarization“ das Ziel einer Zusammenfassung wie folgt [27, S. 1]:

> The goal of automatic summarization is to take an information source, extract content from it, and present the most important content to the user in a condensed form and in a manner sensitive to the users's or applications's needs.

Die automatische Textzusammenfassung stellt dabei ein interdisziplinäres Thema dar. Um diesen Aufwand unter anderen Content-Redakteuren abzunehmen, ist eine Automatisierung nötig.

4.2.1 Arten der Zusammenfassung

Es gibt verschiedene Arten von Zusammenfassungen. Im Folgenden werden einige Ausprägungen bzw. Unterscheidungsmerkmale näher betrachtet:

Extract vs. Abstract

Ein Extract ist eine Aneinanderreihung von Schlüsselwörtern oder Sätzen des Quelltextes. Es werden also Teile der Datenquelle kopiert und zu einem kürzeren Text verknüpft. Bei einem Abstract wird hingegen ein neuer, zusammenhängender Text generiert. Extraction-Ansätze sind einfacher auf größere Datenmengen adaptierbar, da sie in ihrer Länge oder Anzahl an Wörtern oder Sätzen begrenzt sind. Aufgrund der geringen Berücksichtigung der Ko-

härenz kann die resultierende Zusammenfassung jedoch schwieriger zu lesen sein [16].

Indikativ vs. Informativ

Weiters kann zwischen indikativer und informativer Zusammenfassung unterschieden werden. Indikative Zusammenfassungen beinhalten den behandelten Sachverhalt und die Art der Behandlung. Ergebnisse werden dabei ausgespart. Sie sollen auf den relevanten Inhalt aufmerksam machen und zum Lesen des gesamten Dokuments anregen, welches genauere Informationen und Ergebnisse bietet. Im Gegensatz dazu umfassen informative Zusammenfassungen möglichst alle relevanten Informationen des Quelltextes. Dazu zählen das Themengebiet, die Zielsetzung, Hypothesen, Methoden und auch Ergebnisse. Neben den indikativen und informativen Zusammenfassungen wird häufig noch eine dritte Art – die kritische Zusammenfassung – unterschieden [16].

Generisch vs. Benutzerorientiert

Generische Zusammenfassungen sind eher allgemein gehalten und sprechen so ein breiteres Publikum an als benutzerorientierte. Benutzerorientierte Zusammenfassungen passen das Ergebnis an den jeweiligen Nutzer an. Obwohl generische Zusammenfassungen populärer sind, gewinnen aufgrund der personalisierten Filterung von Informationen benutzerorientierte Zusammenfassungen immer mehr an Bedeutung [16].

Single Document vs. Multi Document Summary

Informationen über den Inhalt eines einzelnen Dokuments werden als *Single Document Summary* bezeichnet. Eine *Multi Document Summary* hingegen gibt einen Überblick über eine Kollektion von Dokumenten oder einzelne Dokumente im Kontext von vorher generierten Zusammenfassungen [27].

4.2.2 Ansätze der Textzusammenfassung

Begonnen hat die Entwicklung der automatischen Textzusammenfassung in den späten 1950er-Jahren. Die erste Implementierung eines Algorithmus zur Satzextraktion beschrieb Luhn im Jahr 1958 in seiner Arbeit „Automatic Creation of Literature Abstracts“ [25]. Ziel dabei war die automatische Auswahl von Sätzen zur Erstellung einer Zusammenfassung. Er berechnete dabei die Wichtigkeit der Wörter aufgrund ihrer Häufigkeit. Dieser Faktor basiert auf der Annahme, dass ein Autor bestimmte Wörter bzw. Begriffe, die mit dem Thema verbunden sind, in seinen Argumenten wiederholt. Dadurch wird das eigentliche Thema in den Vordergrund gestellt. Außerdem wertete er die

Position von relevanten Wörtern innerhalb eines Satzes. In einer Kombination dieser beiden Faktoren wollte Luhn die Relevanz der einzelnen Sätze bestimmen.

Auf Grundlage von Luhns Algorithmus entwickelte im Jahr 1969 Edmundson eine neue Methode zur Textextraktion. Seine Arbeit wurde unter dem Titel „New Methods in Automatic Extraction“ [14] veröffentlicht. Wie Luhn war auch er der Meinung, dass die richtige Auswahl signifikanter Sätze zu einer erfolgreichen Zusammenfassung führt. Dabei beachtete er jedoch noch zusätzliche Faktoren:

Anders als Luhn verwendete Edmundson keine Zeitungsartikel zur Entwicklung seiner Methode, sondern naturwissenschaftliche Dokumente, welche eine bestimmte Struktur aufweisen. Deshalb sind für ihn neben den *key words* (häufig vorkommende Wörter aus dem Text) auch *title words* und *heading words* (Wörter aus dem Titel und aus Überschriften eines Textabschnittes) und *sentence location* (Position eines Satzes im Text) von großer Wichtigkeit. Bei seiner Methode sind die Erstellung eines *Dictionary* und eines *Glossars* elementar. Dictionaries sind Wortlisten, die allgemeine Wörter enthalten und unabhängig von dem zu abstrahierenden Dokument sind. Glossare beziehen sich hingegen auf ein bestimmtes Dokument. Aufbauend auf diese Faktoren schlug Edmundson verschiedene Methoden für eine Textzusammenfassung vor:

- *Cue Method (Stichwort-Methode):* Die Relevanz eines Satzes hängt vom Vorkommen spezieller Wörter ab. Bonuswörter wie beispielsweise „wichtig“ oder „signifikant“ haben einen größeren Gewichtungsfaktor, während Stigmaworte wie „unklar“ oder „vielleicht“ zu einer geringen Gewichtung führen. Sogenannte Null-Worte sind beispielsweise Artikel, Pronomen oder Hilfsverben und sind bei der Gewichtung irrelevant.
- *Key Method (Schlüsselwort-Methode):* Dieser Faktor ist äquivalent zu Luhns Methode. Schlüsselwörter werden aufgrund der Häufigkeit extrahiert.
- *Title Method (Titel-Methode):* Schlüsselwörter werden aus Überschriften oder Titeln extrahiert, in einem Glossar gesammelt und erhalten eine höhere Relevanz.
- *Location Method (Methode zur Ermittlung der Satzposition):* Edmundson nimmt des Weiteren an, dass bestimmte Sätze unter einer Überschrift oder einem Titel eine höhere Relevanz besitzen und wichtige Sätze häufig sehr bald oder sehr spät in einem Dokument und dessen Absätzen auftauchen.

Aufbauend auf diese beiden Algorithmen entstanden weitere, modernere Variationen von Extraktionsmethoden. Heutige Extraktionsansätze verwenden Methoden basierend auf *Machine Learning* und *Natural Language Analysis*.

Kupiec et al. [24] beschrieb 1995 eine von Edmundson abgeleitete Methode, welche anhand einer Menge von Daten und den dazugehörigen Zusam-

menfassungen lernen kann. Dabei entscheiden fünf Merkmale, welche teilweise auf die von Edmundson beruhen, über die Wichtigkeit der Sätze. Das erste Merkmal ist die Satzlänge. Kurze Sätze werden eher selten in Zusammenfassungen inkludiert. Des Weiteren werden Sätze, die bestimmte Phrasen wie „zusammenfassend“ oder „Ergebnisse“ enthalten, als wichtig eingestuft. Außerdem betrachten sie die ersten zehn und die letzten fünf Absätze eines Dokuments und beurteilen die darin enthaltenen Sätze danach, ob sie den Absatz einleiten, abschließen oder ob sie sich in der Mitte des Absatzes befinden. Kupiec et al. definieren ebenfalls die am häufigsten vorkommenden Inhaltswörter als Schlüsselwörter und bewerten Sätze mit Schlüsselwörtern als wichtig. Auch Eigennamen werden als potenziell wichtig betrachtet.

Im Jahr 2008 stellten Ercan und Cicekli [15] eine Methode zur automatischen Textzusammenfassung vor, welche *lexikalische Ketten* verwendet. Eine lexikalische Kette ist ein Graph, dessen Knoten die Bedeutung der Wörter repräsentieren und dessen Kanten die semantische Relation dieser Wörter darstellen. Hierfür werden Ontologien, wie beispielsweise WordNet[2], benötigt, um semantische Beziehungen zu ermitteln.

Auch die Autoren Bawakid und Oussalah stellten 2008 ein System vor, welches WordNet für eine Evaluierung der semantischen Ähnlichkeit einsetzt [3]. Des Weiteren gibt es noch Anätze mit *Hidden Markov Models*, beispielsweise von Conroy and O'leary [10]. Praktische Anwendungen von Zusammenfassungstools sind das *AutoAbstract* von Apache OpenOffice[3] oder *AutoSummarize* von Microsoft Office Word[4]. Im Webbereich gibt es beispielsweise den Online-Service *getDIGEST*[5], welcher lange Texte sinngemäß zusammenfasst, indem semantische und morphologische Technologien eingesetzt werden.

4.2.3 Evaluierung

Zu den Qualitätsmerkmalen von Zusammenfassungen zählen neben Kohäsion und Kohärenz die Kompression und der Informationsgehalt.

Die Evaluierung von Zusammenfassungen ist ein schwieriges Verfahren, da es keine ideale Zusammenfassung gibt, die als Vorlage verwendet werden kann. Grundsätzlich können die Methoden der Evaluierung in zwei Kategorien klassifiziert werden: die intrinsischen und die extrinsischen Methoden [28].

Intrinsische Methode

Den Vergleich der Zusammenfassung mit einer oder mehreren Referenz-Zusammenfassungen gehört zu den intrinsischen Methoden. In einer weiteren

[2] WordNet, http://wordnet.princeton.edu/
[3] OpenOffice, http://www.openoffice.org/
[4] Microsoft Office Word, http://office.microsoft.com/en-us/word/
[5] getDIGEST http://getdigest.de/

Methode wird der Text mit dem Quelldokument verglichen. Ein solcher Vergleich ist dann sinnvoll, wenn die vom System erzeugte Zusammenfassung mit von Menschen extrahierten semantischen Informationen des Quelltextes verglichen wird [28].

Extrinsische Methode

Bei den extrinsischen Methoden wird die Zusammenfassung im Hinblick auf eine Aufgabe, wie beispielsweise eine externe Fragestellung, evaluiert. Dabei wird gemessen, wie gut der generierte Text zur Beantwortung der Frage beiträgt. Eine automatisch generierte Zusammenfassung kann ebenfalls im Hinblick auf den Aufwand, der für die Nachbearbeitung des Textes aufgebracht werden muss, bewertet werden [28].

Herausforderungen

Die Hauptschwierigkeit beim Evaluieren von Zusammenfassungen ist das Finden einer fairen Lösung, mit dem die Ergebnisse der Systeme verglichen werden können. Eine generierte Zusammenfassung kann oft von der Vorlage abweichen und trotzdem gut sein. Des Weiteren müssen Zusammenfassungen in verschiedenen Kompressionsraten bewertet werden können. Das erhöht wiederum den Umfang und die Komplexität der Auswertung. Auch benutzerorientierte Zusammenfassungen erschweren die Evaluierung zusätzlich [28].

4.3 Eingliederung von Artikeln in die Navigation

Um Texte in einem CMS oder anderen Systemen automatisch in die Navigation einordnen zu können, benötigt man ein Verfahren, um dessen Themenschwerpunkt zu extrahieren. Im Folgenden werden zwei verschiedene Ansätze – Textklassifikation und Text Clustering – vorgestellt. Mit Hilfe dieser könnte man Anwender in ihrem redaktionellen Prozess unterstützen, indem zum Beispiel Artikel einer Nachrichtenseite automatisch Kategorien wie Sport oder Politik zugeordnet werden bzw. indem neue Kategorien gebildet werden.

4.3.1 Textklassifikation

Automatische Textklassifikation oder Textkategorisierung findet seine Anwendung beispielsweise in Spamfiltern, in der Zuordnung von Zeitungsartikeln zu verschiedenen Themen, in der Organisation von Webseiten in hierarchischen Kategorien oder in der Organisation von Dokumenten.

Bei einer Textklassifikation handelt es sich um den Prozess, Textdokumente zwei oder mehreren Kategorien aufgrund deren Inhalt und bestimmten

Merkmalen zuzuordnen. Die gängigste Form ist die binäre Klassifikation, bei der die Dokumente einer aus zwei Kategorien zugewiesen werden. E-Mails können so beispielsweise als „Spam“ oder „kein Spam“ kategorisiert werden [33].

Prozess der Klassifizierung

Häufig wird der Text für die Klassifikation vor-verarbeitet, indem Schritte wie die Eliminierung von *Stopp-Wörtern* oder *Stemming* vorgenommen werden. Stopp-Wörter sind Wörter, die häufig in der Sprache vorkommen, wie Artikel, Pronomen, Präpositionen oder Hilfsverben. Stemming bezeichnet das Verfahren Varianten eines Wortes in ihren Wortstamm zurückzuführen. Aus dem bereinigten Text werden dann Merkmale mittels statistischer oder semantischer Ansätze gesammelt. Danach wird eine entsprechende Technik des maschinellen Lernens für die Klassifizierung eingesetzt. Zum Extrahieren der Merkmale bzw. der relevanten Wörter gibt es verschiedene Methoden des *Information Retrieval*, wie *TF-IDF* (term frequency und inverse document frequency, zu deutsch „Vorkommenshäufigkeit“ und „inverse Dokumentenhäufigkeit“), *LSI* (latent semantic indexing) oder *Multi-word*. Im Anschluss daran wird eine Technik des maschinellen Lernens eingesetzt, um den Klassifikator zu trainieren. Dieser wird anhand einer Menge an Textdokumenten getestet. Beispiele dafür sind *Naive Bayes*, *Neuronale Netze*, *Entscheidungsbäume* oder eine Kombination verschiedener Ansätze [11].

Die automatische Textklassifizierung hat einige Herausforderungen zu bewältigen. Darunter fällt der Umgang mit unstrukturierten Texten, das Auswählen von Merkmalen, das Abrufen von Metadaten für die Klassifizierung und die Auswahl einer Technik des maschinellen Lernens. LSI und Multiword-Techniken sind semantisch-orientierte Techniken, mit welchen versucht wird, Probleme in der Klassifikation wie Polysemie und Synonymie zu überwinden. Bei der Multi-word-Technik handelt es sich um eine Sequenz von aufeinanderfolgenden Wörtern, die gemeinsam eine semantische Bedeutung haben (zum Beispiel „University of Applied Sciences Upper Austria“) [11].

Beispiele

Viele existierende Systeme verwenden die *„Bag-of-words“-Methode* für eine Textklassifizierung. Nach dieser Methode wird jedes Dokument nur durch die Worte dargestellt, die dieses auch beinhaltet. Die Semantik wird in diesem Fall ignoriert. Deshalb stellen dabei Synonyme, Polyseme sowie Ausdrücke, die aus mehreren Wörtern bestehen, eine Herausforderung auf lexikalischer Ebene dar. Auch auf konzeptioneller Ebene gibt es Schwierigkeiten, da eine Generalisierung der Wörter fehlt. Somit kann beispielsweise bei „Auto“ und „Motorrad“ nicht auf den Oberbegriff „Fahrzeug“ geschlossen werden [7].

Bloehdorn und Hotho [7] beschreiben, wie mit Hilfe von Semantic Web-

Technologien – genauer gesagt mit dem Einsatz von Hintergrundwissen in Form einer einfachen Ontologie – die Textklassifizierung verbessert werden kann, indem genau diese beschriebenen Problembereiche behandelt werden.

Auch Bawakid und Oussalah [2] beschreiben ein System, welches eine automatische, Semantik-basierte Textkategorisierung durchführt. Zusätzlich zur „Bag-of-words"-Methode setzen sie WordNet ein, um die Ähnlichkeit der Dokumente zu den verschiedenen Themen zu ermitteln. Sie vergleichen dabei verschiedene WordNet-basierende Ansätze, wie die Erweiterung der Wörter mit deren Synonymen oder das Konvertieren von Verben in Nomen.

Ein konkretes Anwendungsbeispiel in diesem Bereich bietet OpenText[6] mit einer *Semantic Navigation*[7]. Dabei werden Inhalte analysiert und mit Metadaten annotiert und für den Nutzer automatisch zusammengestellt.

4.3.2 Text Clustering

Anders als bei der Textklassifizierung sind bei dem Text Clustering (auch *Document Clustering*) die Kategorien bzw. Cluster vorher nicht bekannt. Ziel ist es also, neue Gruppen in den Daten zu identifizieren. Mit dieser Methode könnten Artikel nicht nur in eine bestehende Navigation eingeordnet werden, sondern können diese auch erweitern.

Prozess des Text Clustering

Das Clustern von Dokumenten ist ein unüberwachter Prozess, bei dem Dokumente automatisch in Gruppen basierend auf der Ähnlichkeit derer Inhalte eingeteilt werden. Der Prozess ist unüberwacht, da er keine Muster verwendet, um das Clustering zu leiten bzw. auch keine Vorgaben über das Ergebnis macht. Die Herausforderung beim Clustering ist es also, Dokumente bzw. Webseiten ohne Vorwissen in aussagekräftige Cluster zu gruppieren [33].

Beispiele

2002 beschreiben Choudhary und Bhattacharyya [9] eine Text Clustering-Methode, in der sie auch semantische Informationen in Form von Beziehungen zwischen Wörtern in Sätzen beachten. Sie setzten dafür die *Universal Networking Language* (UNL) ein, welche das Dokument in Form eines semantischen Graphen darstellt.

Auch Shaban entwickelt 2009 in seiner Arbeit „A Semantic Approach for Document Clustering" [51] einen neuen Ansatz, um Dokumente zu Clustern, indem er semantische Informationen der Inhalte nutzt. Er will damit die Genauigkeit der Messung von Ähnlichkeiten erhöhen und setzt dafür auf die

[6] OpenText, http://www.opentext.com/

[7] OpenText Semantic Navigation, http://www.opentext.com/global/products/web-content-management/web-experience-management/opentext-semantic-navigation.htm

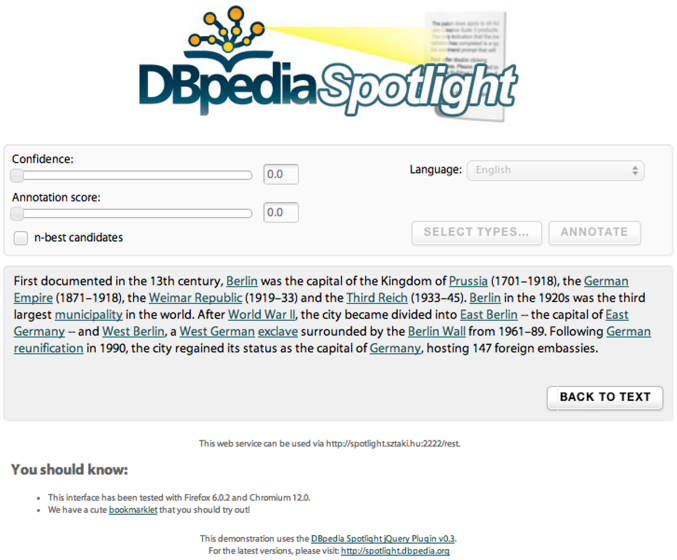

Abbildung 4.1: Demo einer Anreicherung des Textes mit DBpedia URIs (http://dbpedia-spotlight.github.io/demo/).

Analyse des Textes. Die drei Hauptteile dieses Ansatzes sind *Text Parser*, *Similarity Estimator* und *Mining Process*. Der Text Parser ist verantwortlich, den Text einzulesen und zu konvertieren. Der Similarity Estimator nimmt zwei Texte und bestimmt die semantische Ähnlichkeit zwischen ihnen. Den dritte Teil bildet der Document Mining Process, welcher in seiner Arbeit das Semantic Document Clustering bildet. Diese drei Hauptteile des semantischen Ansatzes von Shaban werden beispielsweise auch von Mahalle und Shah [26] aufgegriffen.

4.4 Semantische Textanreicherung

Das Anreichern von Texten ist eines der Gebiete, in denen Semantic APIs häufig zum Einsatz kommen. Von Benutzern bzw. Redakteuren erstellte Texte in CMS oder Blogs können einfach und schnell mit Zusatzinformationen erweitert werden.

In der Arbeit „DBpedia Spotlight: Shedding Light on the Web of Documents“ [31] beschreiben Mendes et al eine Entitätenextraktion, welche ausschließlich auf DBpedia basiert. *DBpedia Spotlight* ist eine Open Source-Webapplikation zur automatischen Annotation von Texten mit DBpedia URIs (siehe Abbildung 4.1). Es erlaubt Benutzern die Konfiguration dieser Annotationen auf ihre spezifischen Bedürfnisse.

4.4.1 Möglichkeiten einer Anreicherung

Zu den möglichen Anwendungsfällen einer Anreicherung zählen beispielsweise das Bebildern von Texten, das Erkennen von Orten im Text oder das Erstellen von Querverweisen.

Mihalcea und Csomai [32] beschreiben eine automatische Anreicherung von Texten mit Links zu Wikipedia. Das dabei entwickelte System „Wikify!“ identifiziert die wichtigsten Konzepte aus dem Dokument und verlinkt diese Konzepte zu den entsprechenden Wikipedia-Seiten. Es besteht also aus zwei Schritten: der Extraktion von Schlüsselwörtern und der anschließenden Disambiguierung, also der Auflösung sprachlicher Mehrdeutigkeiten. Die von dem System produzierten Annotierungen können verwendet werden, um Online-Dokumente zu semantisch relevanten Informationen zu verweisen. Die Evaluierung dieser Annotationen ergab dabei, dass sie sehr zuverlässig und kaum von der manuellen Annotation zu unterscheiden sind.

4.4.2 Praktische Beispiele

Das Blogsystem WordPress bietet mit dem Plugin *Tagaroo*[8] mit Hilfe des Services Open Calais ein automatisches Tagging von Blog-Einträgen an. Das Plugin analysiert die geschriebenen Einträge und schlägt dafür Tags vor. Diese Tags können einerseits im Eintrag verwendet und andererseits zum automatischen Inkludieren von Bildern von Flickr eingesetzt werden.

Des Weiteren existiert für WordPress das Plugin *Zemanta*[9], welches während des Schreibens unter anderem verwandte Artikel und Bilder sowohl aus dem eigenen Blog als auch aus dem Netzwerk vorschlägt.

Da die semantische Textanreicherung großes Potenzial für eine gute Unterstützung von Redakteuren in ihrer täglichen Arbeit bietet, wurde im Zuge dieser Arbeit für das CMS TYPO3 eine Erweiterung entwickelt, welche in Kapitel 5 genauer beschrieben wird.

[8] Tagaroo, http://tagaroo.opencalais.com/
[9] Zemanta WordPress Plugin, http://wordpress.org/plugins/zemanta/

Kapitel 5

Anwendungsbeispiel *SemantLink*

In diesem Kapitel wird ein konkretes Beispiel für die Unterstützung von Redakteuren in CMS mit Hilfe des Semantic Web geboten. Die Anwendung *SemantLink* ist eine Erweiterung des CMS TYPO3 und ermöglicht – ähnlich wie das in Abschnitt 2.3.1 beschriebene Plugin von Zemanta – eine Anreicherung von verfassten Artikeln mit Zusatzinformationen wie Bildern, weiterführenden Artikeln und Kartenausschnitten.

Dabei wird im ersten Schritt der verfasste Text an eine Semantic API gesendet und analysiert, welche passende Schlüsselwörter liefert. Im zweiten Schritt werden anhand dieser Schlüsselwörter – die vom Benutzer bearbeitet werden können – verwandte Artikel, Bilder und Kartenausschnitte aus definierten internen und externen Quellen vorgeschlagen. Die vom Redakteur erstellten Texte können so schnell und einfach ansprechend gestaltet bzw. angereichert werden.

Der Fokus dieses Anwendungsbeispiels liegt auf der direkten Einbettung der Unterstützung in den Arbeitsbereich der Redakteure, um somit die Usability für sie zu erhöhen, ohne sie zusätzlich zu belasten oder sie in ihrem Arbeitsfluss zu stören.

In den folgenden Abschnitten wird erläutert, was, warum und wie die Anwendung umgesetzt wurde und wie das Ergebnis aussieht. Abschnitt 5.1 beschreibt dabei das *was* und *warum*, Abschnitt 5.2 das *wie* und 5.3 das Resultat der Implementierung.

5.1 Konzeption

SemantLink wurde für das TYPO3 CMS entwickelt und verwendet verschiedene APIs. Hauptkomponente dieser Anwendung bildet ein Plugin, welches das TYPO3 CMS erweitert und sowohl das Benutzerinterface als auch die API-Zugriffe verwaltet. Wie in Abbildung 5.1 ersichtlich, bietet das Plugin

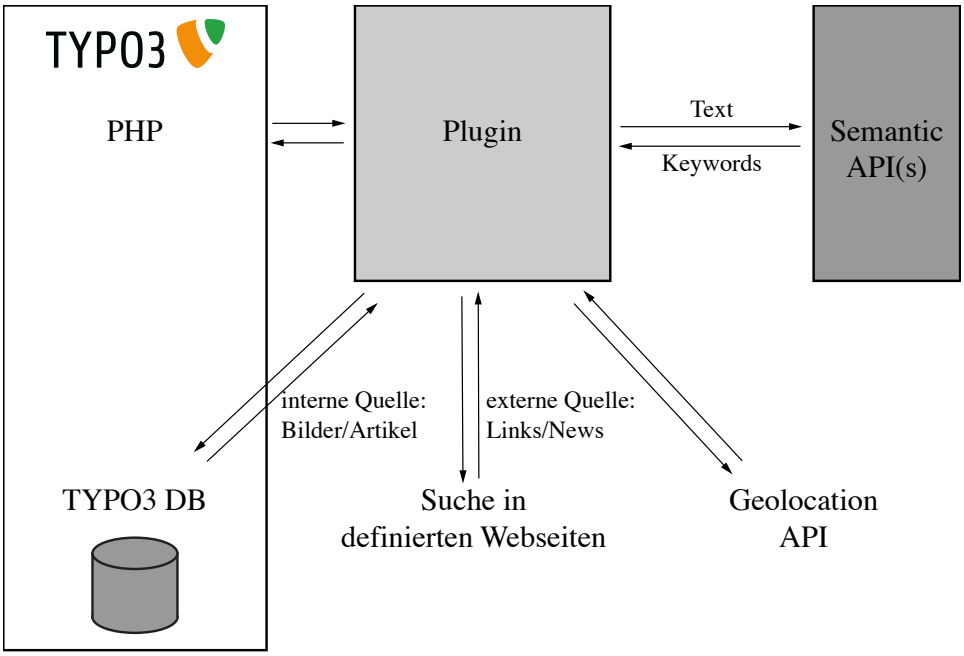

Abbildung 5.1: Architektur der Anwendung *SemantLink*.

eine Anbindung an eine Semantic API. Der vom Redakteur verfasste Text wird per Klick vom Plugin an eine Semantic API übermittelt, welche eine semantische Textanalyse durchführt und die extrahierten Schlüsselwörter mit einer Relevanzbewertung zurückliefert. Im Plugin werden daraus in einem PHP-Skript die relevanten Informationen gefiltert und zur Ausgabe vorbereitet. Anschließend werden die Daten dem für die Ausgabe zuständigen JavaScript übergeben und von diesem dem Redakteur angezeigt.

Der Redakteur hat nun die Möglichkeit, die gelieferte Liste mit relevanten Schlüsselwörtern zu bearbeiten. Er kann auch selbst zusätzliche Schlüsselwörter definieren, mit denen das Plugin weiterarbeiten soll. Optional werden alle gelieferten Schlüsselwörter für die weitere Suche verwendet.

Die Suche erfolgt sowohl intern – in der Datenbank von TYPO3 – als auch extern mit Hilfe verschiedener APIs. Diese liefern externe, weiterführende Artikel, verwandte interne Artikel, Bilder und einen passenden Kartenausschnitt. Per Klick können diese Komponenten in den Text eingefügt werden.

5.1.1 TYPO3 CMS

Für die Umsetzung dieses Projekts wurde das freie Enterprise CMS TYPO3 in der Version 6.0[1] eingesetzt. Das System ist *Open Source* und wurde unter

[1] http://wiki.typo3.org/TYPO3_6.0

der GPL-Lizenz veröffentlicht und ist somit lizenzkostenfrei erhältlich. Es wurde ursprünglich vom Dänen Kasper Skårhøj im Jahr 1997 entwickelt und wird heute von einer großen TYPO3-Gemeinschaft vorangetrieben. TYPO3 ist ein frei konfigurierbars CMS, basiert auf der Skriptsprache PHP und verwendet MySQL als Datenbank, wobei auch andere Datenbanksysteme, wie Oracle oder Postgres, genutzt werden können.

Das TYPO3 CMS kann mit sogenannten Erweiterungen von Anwendern mit zusätzlichen Funktionen angereichert werden. Beispiele dafür sind News, Shop-Systeme oder Foren. Es wir vor allem im deutschen Sprachraum häufig eingesetzt und gehört zu den bekanntesten CMS aus dem Bereich der freien Software.

Zwischen der im November 2012 neu veröffentlichten Version 6.0 und der vorhergehenden Version wurden einige Neuerungen eingeführt. Bereiche des Backends, wie beispielsweise der Extension-Manager, wurden grundlegend überarbeitet. Es wurde außerdem eine Optimierung der Usability des gesamten Backends durchgeführt. Dazu zählt beispielsweise die Drag & Drop-Funktionalität im Page-Modul oder die Einführung von Namespaces [58].

Die weite Verbreitung dieses CMS, die Erweiterbarkeit und die Tatsache, dass es Open Source ist, machten dieses CMS besonders interessant für dieses Projekt.

5.1.2 Eingesetzte Technologien

Zusätzlich zu TYPO3 wurden noch weitere Technologien zur Umsetzung der Erweiterung *SemantLink* eingesetzt.

Extbase

Die Erweiterung *SemantLink* wurde mit dem MVC Framework *Extbase*[2] umgesetzt. Extbase wird seit TYPO3 Version 4.3 als Systemerweiterung ausgeliefert und ist ab Version 6.0 ein verbindlicher Bestandteil.

Extbase wird in TYPO3 unter anderen eingesetzt, um die Varianz zwischen den Erweiterungen zu reduzieren. Es setzt dabei auf *objektorientierte Programmierung*, *Domain Driven Design*, *Model-View-Controller* und *Test-Driven Development* [43].

TinyMCE RTE

Als Texteditor wird der auf JavaScript basierende Open Source Rich-Text-Editor (RTE) *TinyMCE RTE*[3] eingesetzt, welcher in Form einer Erweiterung in TYPO3 integriert werden kann[4]. Dieser hat gewisse Vorteile gegen-

[2] Extbase MVC Framework http://forge.typo3.org/projects/typo3v4-mvc

[3] TinyMCE RTE http://www.tinymce.com/

[4] TYPO3 Erweiterung tinymce_rte http://typo3.org/extensions/repository/view/tinymce_rte

über dem *htmlArea RTE*[5], welcher standardmäßig in TYPO3 installiert ist. Gründe für den Einsatz dieses Open Source-Projekts sind die gute Dokumentation und die konsistente API. Mit diesem Editor kann der vom Redakteur eingegebene Text ohne Speicherung über JavaScript bearbeitet werden, was grundlegend für diese Anwendung ist.

5.1.3 Verwendete APIs

Sowohl für die externe Suche von weiterführenden Artikeln und Kartenausschnitten, als auch für die semantische Textanalyse wurden verschiedenen APIs eingesetzt.

Externe Suche

Für die externe Suche wurde die Google Custom Search API[6] verwendet. Es können dabei die gewünschten Webseiten festgelegt werden und die API ermöglicht eine Suche in diesen. Somit können Redakteure bzw. die Administratoren einer Webseite genau definieren, welche Seiten durchsucht und den Redakteuren als weiterführende Artikel vorgeschlagen werden sollen.

Über einen *Representational State Transfer* (REST) kann diese API aufgerufen werden. Die notwendigen Parameter für die Abfrage sind der API-Schlüssel, der eindeutige Schlüssel für die erstellte Suchmaschine sowie die Suchabfrage. Bei einer erfolgreichen Anfrage liefert der Server den Status Code „200 OK HTTP“ sowie ein JSON-File mit den Suchresultaten.

Semantic API

Grundlegend für die Suche nach weiterführenden Medien ist eine semantische Analyse des vom Redakteur erstellten Textes. Dafür wird in der Anwendung *SemantLink* die gleichnamige Applikation SemantLink eingesetzt (siehe 2.3.1), welche im Jahr 2012 entwickelt wurde. In den Konfigurationsmöglichkeiten der TYPO3-Erweiterung kann diese auch durch die Zemanta API ersetzt werden. SemantLink zeichnet sich dabei durch die Relevanzbewertung aus. Die Applikation führt nach der Vorverarbeitung eine Entitätenextraktion durch und bewertet im Anschluss deren Relevanz. Dazu werden verschiedene Quellen aus dem Semantic Web und dem Social Web herangezogen. In Form eines Webservices über eine REST-Schnittstelle wird diese Applikation verfügbar gemacht. Zemanta liefert laut [1] gute und relevante Entitäten bzw. Links als Ergebnisse für die Texte und wurde deshalb als alternative API eingesetzt.

[5] htmlArea RTE http://typo3.org/extensions/repository/view/rtehtmlarea_manual
[6] Google Custom Search API https://developers.google.com/custom-search/

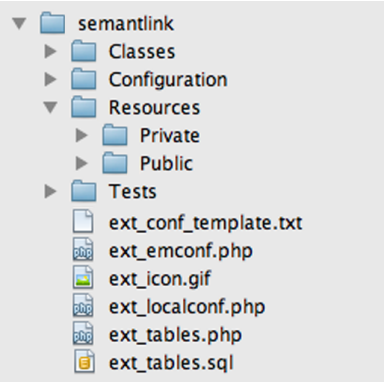

Abbildung 5.2: Aufbau einer Extbase-Erweiterung.

Geolocation API

Das Anzeigen von im Text vorkommenden Orten wird von der Google Maps API[7] übernommen. Für die Anzeige des Kartenausschnitts wird dafür die Google Maps JavaScript API v3 eingesetzt.

5.2 Implementierung

TYPO3 basiert auf der Skriptsprache PHP. Als Datenbank wird für die Anwendung MySQL eingesetzt.

5.2.1 Struktur

Die implementierte Erweiterung *SemantLink* wurde mit Hilfe des MVC-Frameworks Extbase entwickelt, welche auch die Struktur der Erweiterungen fest vorgibt (siehe Abbildung 5.2). Im Erweiterungsverzeichnis selbst liegen die grundlegenden Konfigurationsdateien, welche im Folgenden kurz beschrieben werden:

- *ext_emconf.php* beinhaltet allgemeine Informationen über die Erweiterung wie die Versionsnummer, Angaben über den Autor oder die möglichen Abhängigkeiten.
- *ext_localconf.php* beinhaltet grundlegende Konfigurationseinstellungen. Für die Erweiterung *SemantLink* werden hier unter anderem die Backend-Funktionen als Ajax-Funktionen registriert.
- *ext_tables.php* enthält weiterführende Konfigurationseinstellungen.
- *ext_tables.sql* enthält die Struktur für die benötigten Datenbanktabellen.

[7] https://developers.google.com/maps/

- *ext_ conf_ template.txt* stellt ein optionales File dar, welches die Konfigurationsmaske für den Extension Manager erzeugt. Für die *SemantLink*-Erweiterung werden hier die Schlüssel und die verschiedenen APIs verwaltet.
- Im Ordner *Classes* ist die komplette Programmlogik enthalten.
- Im Ordner *Configuration* befinden sich unter anderem das Table Configuration Array (TCA), welches das Verhalten der Datenbanktabellen beschreibt.
- Das Verzeichnis *Ressources* enthält alle statischen Ressourcen. Hier sind auch die CSS-Dateien und JavaScript-Files gespeichert.

5.2.2 Ablauf

Erstellt man im TYPO3 Backend einen neuen Eintrag vom Typ *SemantLink*, bietet diese dem Redakteur die gewohnten Felder, die er auch sonst von TYPO3 bei der Erstellung eines Content-Elements vom Typ „Text“ erhält. Neben dem Titel gibt es den Texteditor. Hier wurde jedoch nicht der von TYPO3 standardmäßige Editor, sondern der TinyMCE RTE verwendet, wie in Abschnitt 5.1.2 erläutert wurde.

Neben diesen Eingabefeldern ist die Hauptkomponente von *SemantLink* die Anreicherungsfunktion. Die einzelnen Felder werden im TCA festgelegt. Das TCA ist ein globales Array im TYPO3, welches – über die Möglichkeiten von SQL hinaus – zusätzliche Definitionen von Datenbanktabellen bietet. Es kann beispielsweise festgelegt werden, welche Felder im TYPO3 Backend bearbeitet werden können oder welche Felder sichtbar sind. Des Weiteren definiert es auch die Beziehungen zwischen der Tabelle und anderen Tabellen und wie die einzelnen Felder validiert werden sollen [61].

Für die *SemantLink*-Anreicherungsfunktion wird dabei kein normaler Feldtyp benötigt, sondern eine sogenannte *userFunc*, welche folgendermaßen eingebunden wird:

```
'semantlink' => array(
    'exclude' => 0,
    'label' => 'LLL:EXT:semantlink/Resources/Private/Language/
    locallang_db.xlf:tx_semantlink_domain_model_semantlinktext.
    semantlink',
    'config' => array(
        'type' => 'user',
        'userFunc' => 'SemantLink\Semantlink\TCA\user_OutputService->
    render',
        'parameters' => array(
            'title' => 'title',
            'text' => 'text',
            'keywords' => 'keywords',
            'map' => 'map'
),    ),    ),
```

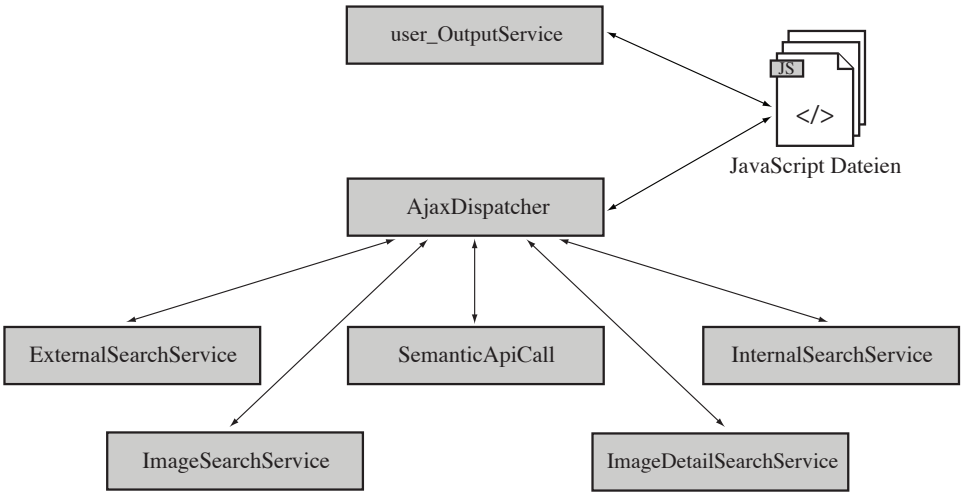

Abbildung 5.3: Ablauf der Anwendung *SemantLink*.

Hier wird also die Funktion `render` der PHP-Klasse `user_OutputService` aufgerufen, ausgeführt und dem Redakteur im TYPO3 Backend als zusätzliches Feld ausgegeben. Die angegebenen Parameter werden der Funktion als Array mit übergeben.

Abbildung 5.3 zeigt den Ablauf der *SemantLink*-Erweiterung. In der Klasse `user_OutputService` wird nun die Ausgabe für das Feld zusammengefügt. Es wird das HTML erzeugt, die CSS-Dateien eingebunden und die JavaScript-Files aufgerufen. In den JavaScript-Dateien werden die verschiedenen Ajax-Aufrufe getätigt und die Benutzerinteraktion verwaltet. Diese werden alle vom `AjaxDispatcher` entgegengenommen und an die jeweiligen Service-Klassen weitergeleitet. Hier werden auch die verschiedenen Konfigurationen und API-Schlüssel aus der Konfigurationsdatei ausgelesen und weitergegeben. Im Ordner „Classes/Service/“ befinden sich die wichtigsten Funktionalitäten der Erweiterung. Von hier aus wird auf die APIs zugegriffen und es werden deren Antworten verwaltet. Diese werden anschließend wieder über den `AjaxDispatcher` zurück an das JavaScript gegeben und über den `user_OutputService` im TYPO3 Backend ausgegeben. Die Klasse `SemanticApiCall` verwaltet den Zugriff auf die Semantic APIs (SemantLink und Zemanta) und der `ExternalSearchService` übernimmt die Kommunikation mit der Google Custom Search API.

Die Erweiterung wurde als eigenständiges Element entwickelt, das im TYPO3 Backend im Listen-Modul angelegt wird. Im Seitenbaum kann anschließend das Plugin, welches eine Listen- und Detailansicht aller Artikel erzeugt, eingefügt werden. Somit werden alle erzeugten Artikel auf einer Seite im Frontend gelistet. Je nach Bedarf könnte die Erweiterung *SemantLink* aber auch als eigenständiges Content Element erzeugt werden. Eine weitere

Möglichkeit wäre, die Tabelle `tt_content`, und somit die einzelnen Content Elemente, um dieses eine Feld zu erweitern.

Zum Testen des Protoypes wurde die Erweiterung in das bestehenden System der Testpersonen integriert. Für die nahtlose Eingliederung in ihr System wurde die letzte Variante – die Einbindung als Erweiterung von `tt_content` und somit als zusätzliches Feld der Content Elemente – gewählt. Die Testpersonen mussten so kein neues Content Element kennenlernen, sondern die Anreicherungs-Funktion wurde ihnen in ihren gewohnten Prozess integriert. Beim Erstellen eines Content-Elements vom Typ „Text“, wurde die Anreicherungs-Funktion einfach unterhalb des Text-Editor eingebunden.

5.3 User Interface

Im Folgenden wird die grafische Benutzeroberfläche der entwickelten TYPO3 Erweiterung vorgestellt. Zu Beginn wird Schritt für Schritt der Ablauf der Textanreicherung beschrieben. Anschließend werden die Konfigurationsmöglichkeiten aufgezeigt.

5.3.1 Ablauf der Textanreicherung

Nach der Installation von *SemantLink* und der abhängigen Erweiterung TinyMCE RTE über den TYPO3 Extension Manager kann die Anwendung genutzt werden. Abbildung 5.4 zeigt, dass sich neben dem Texteditor der Bereich für die Anreicherung präsent aber unaufdringlich einfügt. Dieser wurde so platziert, dass er während des Verfassens eines Textes jederzeit erreichbar ist und nicht gescrollt werden muss.

Klickt man auf die Schaltfläche „Suche nach passenden Keywords“ wird der Text aus dem Editor bereinigt und an eine Semantic API gesendet. Die API liefert bei einer erfolgreichen Extraktion von Schlüsselwörtern diese zurück. Die Schlüsselwörter werden nach ihrer Relevanz absteigend aufgelistet. Schlägt die Suche fehl, wird dies dem Nutzer mit einem Lösungsvorschlag zur Fehlerbehebung mitgeteilt. Werden beispielsweise keine Schlüsselwörter zum Text gefunden, wird dem Nutzer vorgeschlagen, die Suche mit einem längeren Text erneut zu starten.

Die gelisteten Schlüsselwörter, zu sehen in Abbildung 5.5, können vom Redakteur nun bearbeitet werden. Gelistete Schlüsselwörter können deaktiviert und neue hinzugefügt werden.

Sind die Schlüsselwörter an die Wünsche des Redakteurs angepasst, kann dieser mit Klick auf die Schaltfläche „Suche passende Medien“ die interne und externe Suche nach Anreicherungsmöglichkeiten starten. Abbildung 5.6 zeigt das Ergebnis, welches in weiterführende Links, Bilder aus der Google Bilder-Suche, verwandte Artikel und einen Kartenausschnitt gegliedert ist.

Die weiterführenden Links können per Mausklick in den Text an die aktuelle Cursorposition eingefügt werden und ebenso wieder entfernt werden

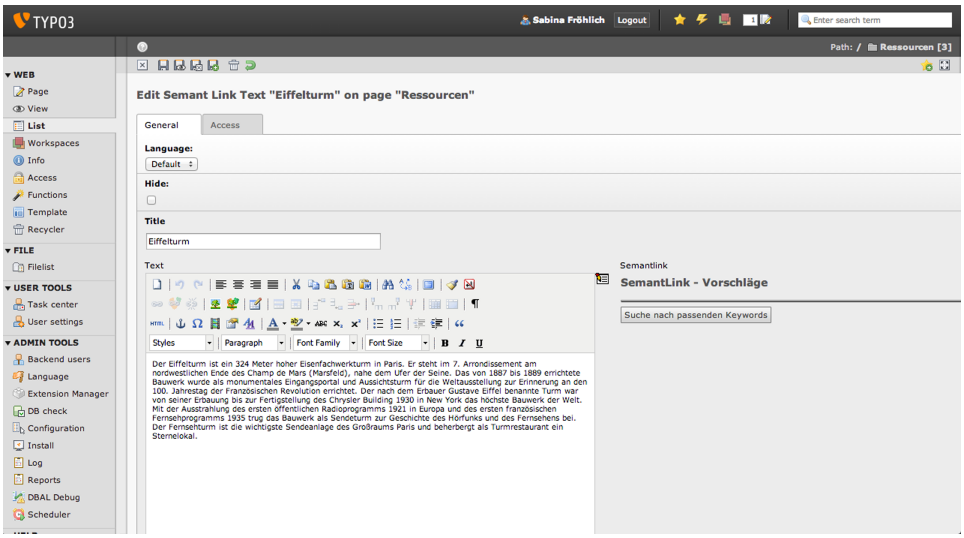

Abbildung 5.4: Einfügen eines Textes im TYPO3 Backend mit integrierter *SemantLink*-Anwendung (Artikel aus http://de.wikipedia.org/wiki/Eiffelturm).

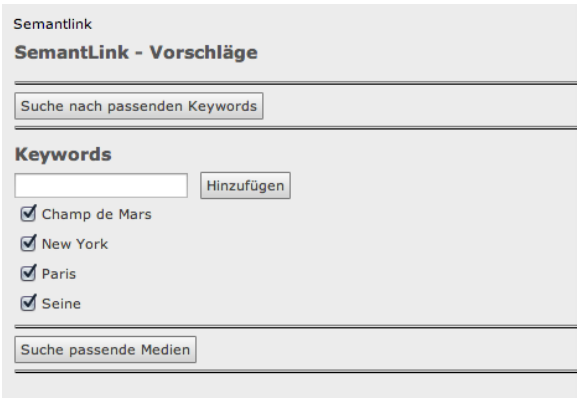

Abbildung 5.5: Extrahierte Schlüsselwörter von der Semantic API.

(siehe Abbildung 5.7). Vor dem Einfügen können die Artikel auch genauer gelesen werden. In einem Pop-up erscheint der verlinkte Inhalt des Suchergebnisses.

Auch die Bilder können einfach per Klick in den Text integriert werden. Diese werden automatisch mit dem Titel und einem Link zur Quelle beschriftet und übernehmen so die Aufgabe der Quellenangabe, die andernfalls durch den Redakteur ausgeführt werden müsste.

Beinhaltet die eigene Datenbank Artikel, die die Schlüsselwörter der semantischen Textanalyse geliefert hat, werden auch diese vorgeschlagen. Re-

Abbildung 5.6: Ergebnis der Anwendung *SemantLink*.

Abbildung 5.7: Ausgewählter weiterführender Link.

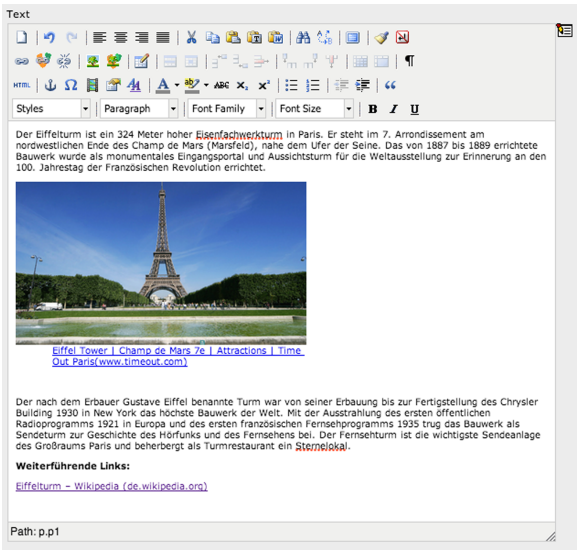

Abbildung 5.8: Beispiel eines angereicherten Textes im Editor.

dakteure haben so einen guten Überblick darüber, welche Texte zu dem jeweiligen Thema schon verfasst worden sind und können gegebenenfalls auf diese verweisen.

Beinhalten die Schlüsselwörter des Weiteren einen Ort oder eine andere Form von Location, wird diese in einem Google Maps-Kartenausschnitt dargestellt. Diese kann vom Redakteur wiederum verändert oder angepasst und per Klick in den Text an die aktuelle Cursorposition eingefügt werden. Besonders interessant ist dieser Bereich für Artikel über Veranstaltungen oder Nachrichten, um den Nutzern bei der Ausgabe der Artikel einen Mehrwert bieten zu können.

In Abbildung 5.8 ist ein Beispiel eines angereicherten Textes im Editor im TYPO3 Backend zu sehen. Der angereicherte Text kann nun gespeichert und im Frontend angezeigt werden.

5.3.2 Konfiguration

Die Anwendung wurde so umgesetzt, dass der Nutzer die Möglichkeit hat, die einzelnen Schnittstellen selbstständig zu konfigurieren. Im Extension Mana-

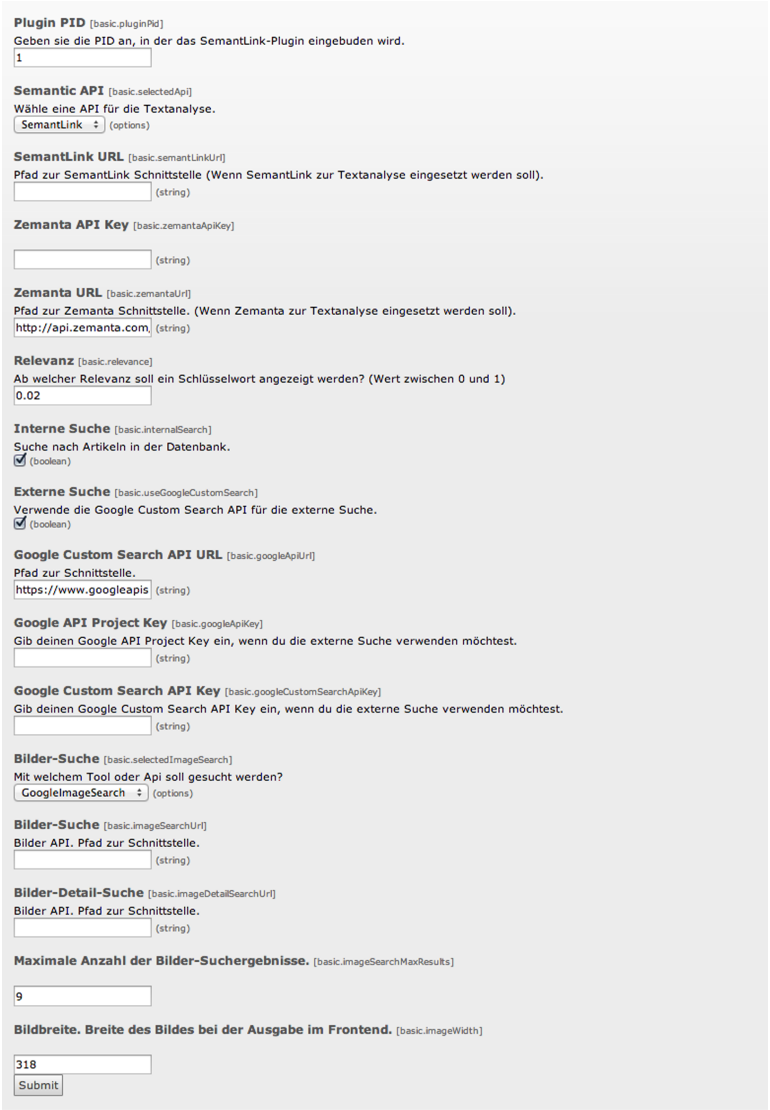

Abbildung 5.9: Konfigurationsmöglichkeiten der Erweiterung *SemantLink*.

ger im TYPO3 Backend erscheint bei der Erweiterung *SemantLink* ein Formular für diese Konfigurationsmöglichkeit. Wie in Abbildung 5.9 zu sehen ist, kann beispielsweise angegeben werden, welche Semantic API eingesetzt werden soll. Eine weitere Möglichkeit der Konfiguration ist das Aktivieren bzw. Deaktivieren der internen oder externen Suche. Außerdem können hier die Keys für die verschiedenen APIs verwaltet werden. Auch die Anzahl der Bilder-Suchergebnisse und die Bildgröße können hier festgelegt werden.

Kapitel 6

Evaluierung des Anwendungsbeispiels

Nach der Beschreibung der Konzeption und Entwicklung des implementierten Prototypen beschreibt dieses Kapitel dessen Evaluierung. Zur Ermittlung der Gebrauchstauglichkeit und der Zufriedenheit der Anwender wurde eine empirische Studie durchgeführt. Im Folgenden wird die durchgeführte Untersuchung zur Evaluierung des Prototypen aufgezeigt. Dabei werden neben der Testumgebung auch die Methoden und der Ablauf der Studie beschrieben. Die aus der Untersuchung resultierenden Ergebnisse werden abschließend dargestellt und interpretiert.

6.1 Ausgangssituation und Planung der Studie

Ziel dieser Studie war die Evaluierung der Usability und vor allem der Zufriedenheit der Anwender mit dem Prototypen. Der wichtigste Aspekt dabei war die nahtlose Eingliederung des Prototypen in den alltäglichen Arbeitsablauf der Anwender. Die Studie sollte zeigen, ob das entwickelte Produkt von der Zielgruppe verwendet werden würde und ob es diese bei ihrer Arbeit unterstützen kann. Ein weiteres wichtiges Ziel der Studie war das Sammeln von Verbesserungsvorschlägen und zusätzlichen Funktionalitäten für den Prototypen, um in einer weiterentwickelten Version die Zielgruppe bestmöglich unterstützen zu können.

6.1.1 Testpersonen

Zur Durchführung der Studie wurden sieben Testpersonen aus einem Unternehmen gewählt, welche einerseits regelmäßig Beiträge verfassen und andererseits bereits längere Erfahrung mit dem Content Management System TYPO3 haben. Die Testpersonen – vier weibliche und drei männliche – sind zum Zeitpunk der Studie zwischen 20 und 49 Jahre alt. Die Mehrheit ver-

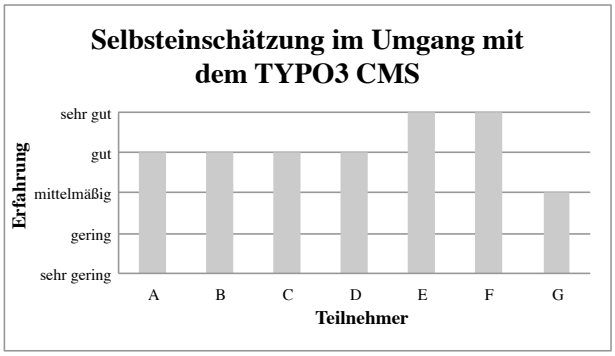

Abbildung 6.1: Eigene Einschätzung über die Erfahrung der Testpersonen ihrer Erfahrung mit dem TYPO3 CMS.

bringt an Werktagen durchschnittlich 6 oder mehr Stunden am Computer und verwendet dabei das TYPO3 CMS mehrmals pro Woche. Zum Arbeiten mit dem CMS setzen dabei alle den Webbrowser *Mozilla Firefox*[1] ein. Die Testpersonen arbeiten alle im Bildungsbereich und unter ihnen sind sowohl Content-Redakteure als auch Betreuer verschiedener Themengebiete. Diese wurden im Vorfeld gebeten, ihre Erfahrungen im Umgang mit dem TYPO3 CMS einzuschätzen. In Abbildung 6.1 sind die Angaben der Probanden aufgeschlüsselt.

6.1.2 Testbedingungen

Die Studie wurde im gewohnten Arbeitsumfeld der Teilnehmer durchgeführt. Bereits im Vorfeld wurde die Anwendung in das TYPO3 CMS der Testpersonen implementiert und erforderte daher keine Installation vor der Evaluierung. Die Probanden nutzten für die Studie ihre eigenen (Arbeits-) Notebooks mit ihrem favorisierten Browser. Da die Studie im Arbeitsumfeld der Probanden stattfand, stand uns ein Besprechungsraum mit Internetzugang zur Verfügung, um mit dem Prototypen arbeiten zu können. Die Testpersonen wurden des Weiteren in zwei Gruppen geteilt, damit ein ruhiges und entspanntes Arbeiten möglich war.

6.1.3 Aufgaben

Während der Studie sollten von den Teilnehmern verschiedene Aufgaben mit dem entwickelten Prototypen gelöst werden. Diese waren genau definiert und gewannen mit jeder weiteren Aufgabe an Komplexität. Die Probanden sollten so einen guten Einstieg im Umgang mit dem Tool erfahren. Die verschiedenen Aufgaben sollten Aufschluss über die einzelnen Funktionalitäten geben und werden im Abschnitt 6.2.5 genauer beschrieben.

[1] Mozilla Firefox, http://www.mozilla.org/ (abgerufen am 05.08.2013)

6.2 Methoden und Ablauf der Nutzertests

Im Folgenden werden die gewählten Methoden der Studie genauer beschrieben und mit alternativen Ansätzen verglichen.

6.2.1 Benutzerorientierte vs. Expertenorientierte Methoden

Bei den benutzerorientierten Methoden wird vom Benutzer direktes Feedback eingeholt, wie bereits in Abschnitt 3.2 ausführlich beschrieben wurde. Die expertenorientierten Methoden beziehen ihr Urteil von Gutachtern. Hier ist es oft sinnvoll, gemischte Experten-Teams einzusetzen, um das Problem von mehreren Perspektiven zu evaluieren.

Die beiden Methoden unterscheiden sich je nach Anwendung in der Anzahl der gefundenen Probleme sowie in der Art dieser Probleme. Durch die expertenorientierte Methode kann mit Unterstützung von Usability-Experten eine Vielzahl von Usability-Problemen identifiziert werden. In einer Studie von Desurvire [12] wurde aufgezeigt, dass ein Evaluator mit der besten Methode trotzdem 56 Prozent der Probleme, die in einem Produkttest im Labor gefunden wurden, übersehen hat. Auch beim Einsatz einer Gruppe von Evaluatoren können diese die benutzerorientierten Methoden nicht ersetzen, sondern würden nur als Ergänzung dienen.

Benutzerdefinierte Methoden sind jedoch sehr kosten- und zeitintensiv und häufig auch sehr komplex. Die Benutzer sind es aber schlussendlich auch, die das Produkt verwenden müssen und darüber entscheiden, ob es leicht zu bedienen ist und deshalb akzeptiert wird oder nicht [49].

Doch beide Methoden haben ihre Berechtigung. Empfehlenswert ist deshalb auch – abhängig vom jeweiligen Projekt – eine Kombination von verschiedenen Methoden [38].

6.2.2 Quantitative vs. Qualitative Methoden

Bei einer quantitativen Untersuchungsmethode wird in der Regel mit einer großen Anzahl von Teilnehmern und statistischen Auswertungen gearbeitet. Die qualitative Methode setzt hingegen auf eine verhältnismäßig geringe Anzahl von Testpersonen und es werden interpretativen Analysen herangezogen. Sämtliche expertenorientierte Methoden sind qualitativer Natur, wobei ein Experte eine subjektive, aber begründete Einschätzung abgibt. Bei einer Usability-Evaluation kommt es darauf an, konkrete Probleme möglichst effizient zu identifizieren.

Für die Studie wurden benutzerorientierte Methoden gewählt. Dabei wurde weniger auf Benutzerbeobachtung als auf Benutzerbeteiligung gesetzt. Dafür wurde unter anderem eine *Benutzerbefragung mit Fragebögen* sowie ein *Fokusgruppen-Interview* vorbereitet und durchgeführt.

6.2.3 Fragebogen

Diese Form der Evaluierung kann in verschiedenen Phasen des Designprozesses erfolgen. Grundlage für den entwickelten Benutzerfragebogen bildete der ISONORM-Fragebogen, welcher in Abschnitt 3.2.2 erläutert wurde. Fragen bezüglich der Aufgabenangemessenheit wurden dabei übernommen, erweitert und angepasst. Dieser Abschnitt ist besonders geeignet, um herauszufinden, ob die Software die Erledigung der Arbeitsaufgaben ermöglicht, ohne die Anwender als Benutzer unnötig zu belasten. Jede Frage kann mit Hilfe einer sechsstufigen Skala von sehr negativ bis sehr positiv beantwortet werden.Der ISONORM-Fragebogen arbeitet ursprünglich mit einer siebenstufigen Skala.

Eine ungerade Zahl von Stufen bietet den Befragten eine mittlere Kategorie, also eine neutrale Antwortmöglichkeit. Dieser mittlere Wert wird nicht immer eindeutig interpretiert und teilweise als „Fluchtkategorie“ angesehen. Mit einer geraden Anzahl nimmt man den Befragten jedoch die Chance, bewusst bzw. gezielt eine neutrale Antwort abzugeben [40].

Die Fragen in der Studie beziehen sich alle auf die Zufriedenheit der Teilnehmer. Um eine gewissen Tendenz zu einer positiven oder negativen Einstellung bezüglich des Prototypen zu erlangen, wurde eine gerade Anzahl von Antwortmöglichkeiten gewählt.

Der Fragebogen eignet sich gut, um eine generelle Einschätzung und Bewertung des Prototyps durch die Teilnehmer zu bekommen, ist jedoch zu allgemein gehalten, um konkrete Probleme identifizieren zu können und sollte daher mit weiteren Instrumenten kombiniert werden [54].

6.2.4 Fokusgruppen-Interview

Wie in Abschnitt 3.2.2 beschrieben, besteht eine Fokusgruppe aus einer Gruppe von Personen, die gemeinsam über ein Produkt oder einen Teil eines Produkts diskutieren. Der Versuchsleiter hat hier die Aufgabe, der Moderation und somit die Aufgabe die Diskussion aufrechtzuerhalten oder in die richtige Richtung zu lenken. Mit Hilfe des Fokusgruppen-Interviews kann nach den gesammelten quantitativen Daten der Fragebögen noch qualitatives Material gesammelt werden, um die eventuell aufgetretenen Probleme besser verstehen zu können.

6.2.5 Vorgehensweise

Die Studie wurde an einem Tag durchgeführt, wobei die Testpersonen in zwei Gruppen aufgeteilt wurden. Der Arbeitsplatz jedes Teilnehmers bestand aus dem eigenen Arbeitslaptop, auf dem in einem selbst gewählten Browser der entwickelte Prototyp aufgerufen wurde. Die Projektleitung bestand dabei aus drei Personen – dem Leiter und zwei Assistenten zur Dokumentation von möglichen Problemen, Fragen oder wichtigen Aussagen der Probanden.

Einführung:

Gestartet wurde mit einer Einführungsphase. Diese diente vor allem dazu, die Testpersonen zu begrüßen und sie auf die Studie vorzubereiten bzw. das Ziel sowie den Ablauf dieser näher zu bringen. Sie wurden darauf hingewiesen, dass alle erhobenen Daten vertraulich behandelt werden und sie die Studie jederzeit abbrechen können. Außerdem wurde verdeutlicht, dass es bei der Studie nicht darum geht, ihre Kompetenzen, Fähigkeiten oder die Teilnehmer selbst zu testen, sondern vielmehr darum, das Tool zu testen. Im Anschluss daran wurden die Testpersonen gebeten, eine Einverständniserklärung zu unterzeichnen. Damit bestätigen die Teilnehmer, dass sie freiwillig an der Untersuchung teilnahmen, alle Untersuchungsergebnisse in anonymisierter Form ausgewertet und dargestellt werden. Im Zuge der Einführung wurde ein Fragebogen mit demografischen Daten, der Computernutzung und der Nutzung des TYPO3 CMS gestellt (siehe Anhang A.3). Anschließend wurde der entwickelte Prototyp vorgestellt und das Ziel der Studie noch einmal verdeutlicht. Außerdem wurden sie darüber informiert, dass von den Projektleitern während der Studie Notizen über möglicherweise auftretende Probleme gemacht werden. Vor Beginn der Studie wurden die Teilnehmer noch auf die anschließende Fokusgruppe hingewiesen. Dafür wurden Notiz-Zettel ausgehändigt, damit sich die Probanden etwaige Fragen oder Probleme für später notieren konnten.

Aufgabenstellung:

Um den Prototypen testen zu können, wurden verschiedene Szenarien bzw. Aufgabenstellungen definiert (siehe Anhang A.3). Diese beinhalteten alle Aufgaben zum Thema Textanreicherung und sollten von den Testpersonen umgesetzt werden. Für jede Aufgabe wurde den Probanden ein fertiger Text zur Verfügung gestellt. Dieser konnte in das TYPO3 CMS eingefügt werden, um anschließend verschiedene Aufgaben damit zu erledigen. Nach Beendigung jeder Aufgabe mussten ein paar kurze Fragen beantwortet werden. Die Probanden mussten so die vorgegebenen Texte je nach Aufgabe mit Bildern, weiterführenden Links oder Kartenausschnitten anreichern. Die Fragen konnten sie mit einer fünfstelligen Skala von „trifft vollständig zu“ bis zu „trifft gar nicht zu“ bewerten. Hier wurden außerdem die gefundenen Schlüsselwörter auf ihre Aussagekraft getestet. So konnten die Teilnehmer die Anzahl der gefundenen Schlüsselwörter, die sie auch für eine Suche verwenden würden, dokumentieren. Die Ausführung aller Aufgaben dauerte im Durchschnitt 1,5 Stunden. Die Probanden bearbeiteten die Aufgaben selbstständig. Während der Durchführung konnten jederzeit Fragen gestellt werden. Von der Projektleitung wurden auftretende Probleme oder andere Zwischenfälle notiert.

Fragebogen:

Anschließend an die Bearbeitung der einzelnen Aufgabenstellungen wurde der Fragebogen zur Bewertung der Zufriedenheit der Probanden ausgehändigt. Dieser beinhaltete folgende Fragen:

- Ich bin mit der Bedienung des *SemantLink*-Tools gut zurechtgekommen.
- Das *SemantLink*-Tool hat mich bei meinen Aufgaben gut unterstützt.
- Die Verwendung des *SemantLink*-Tools war ein zeitlicher Zusatzaufwand im Vergleich zum gewohnten Ablauf.
- Das *SemantLink*-Tool werde ich in Zukunft beim Erstellen von News und anderen Beiträgen einsetzen.
- Das *SemantLink*-Tool bietet mir alle Möglichkeiten, die ich für die Bearbeitung meiner Aufgaben benötige.
- Die für die Aufgabenbearbeitung notwendigen Informationen befinden sich immer am richtigen Platz auf dem Bildschirm.
- Das *SemantLink*-Tool ist auf die von mir zu bearbeitenden Aufgaben zugeschnitten.
- Ich bin mit dem TYPO3 CMS inklusive *SemantLink*-Tool allgemein zufrieden.
- Ich würde das *SemantLink*-Tool auch anderen Kollegen zur Erstellung von News und anderen Beiträgen empfehlen.

Das Ausfüllen dieses Fragebogens dauerte zirka fünf bis zehn Minuten.

Fokusgruppen-Interview:

Nachdem beide Gruppen die Studie vollzogen hatten, wurde eine Fokusgruppe mit allen Teilnehmern abgehalten. Die Fokusgruppe beinhaltete zwei Themengebiete. Einerseits wurde der allgemeine, übliche Prozess der Erstellung eines Beitrags abgefragt. Die Probanden sollten also ihren üblichen Arbeitsablauf beschreiben und dabei besonders auf Herausforderungen und Probleme, aber auch auf verwendete Werkzeuge und den Prozess der Recherche für neue Beitragsthemen eingehen. Einleitende Fragen waren beispielsweise folgende:

- Wie sieht bei Ihnen der übliche Prozess der Erstellung eines Beitrags aus?
- Was ist dabei zu tun? Wie gehen Sie mit bestehenden Werkzeugen um?
- Woher bekommen Sie Anregungen für einen neuen Beitrag?
- Auf welche Probleme oder Herausforderungen stoßen Sie dabei?

Andererseits wurden einleitende Fragen zum Prototypen gestellt. Ziel hierbei war es, Verbesserungsvorschläge und ein allgemeines Gefühl der Zufrieden-

heit der Teilnehmer mit dem Tool zu bekommen. Es wurde darüber diskutiert, was die Probanden als positiv und was als negativ empfunden hatten:

- Wo kann die *SemantLink*-Anwendung ihrer Meinung nach unterstützend wirken?
- Kann es Ihnen in alltäglichen Aufgaben helfen? Wenn ja, inwiefern hilft es Ihnen?
- Sind Sie auf Probleme oder Unklarheiten gestoßen?
- Was könnte aus Ihrer Sicht verbessert werden?
- Wünschen Sie sich irgendwelche Zusatzfunktionalitäten?

Pilotphase:

Nach Beendigung der Usability-Studie wurde diese ausgewertet. Aufgrund der Ergebnisse wurde der Prototyp noch einmal überarbeitet. Dabei wurden zwar keine grundlegenden Funktionalitäten verändert, aber besonders die Ausgabe der Schlüsselwörter und die der Bilder mussten angepasst werden. Es wurden beispielsweise doppelte Schlüsselwörter entfernt, genauso wie Bilder, die aufgrund ähnlicher Suchwörter doppelt auftauchten, entfernt und übersichtlicher aufgelistet. Nach dieser Überarbeitung wurde der Prototyp erneut in das System der Testpersonen integriert. Die Anwendung sollte anschließend über einen längeren Zeitraum intensiv in der Praxis getestet werden. Die Probanden bekamen dafür keine vordefinierten Aufgaben, sondern konnten ihre täglichen Arbeiten damit erfüllen.

Diese Pilotphase sollte noch genauer aufschließen, ob die Anwendung hilfreich für die Probanden ist und ob sie sich gut in ihre Prozesse integriert. Durch die längere, intensivere Auseinandersetzung sollten außerdem Vorschläge für Zusatzfunktionalitäten resultieren.

Nach Abschluss dieser Pilotphase wurden die Testpersonen wiederum gebeten, den Fragebogen zur Zufriedenheit in Form eines Online-Fragebogens auszufüllen. Dies bietet in der Auswertungsphase einen guten Vergleich, wie sich die Zufriedenheit über einen längeren Zeitraum entwickelt hat. Vorteil der Online-Abwicklung ist jener, dass die Testpersonen anonym sind und auch keine Scheu davor haben müssen, möglicherweise auch ein schlechtes Urteil abzugeben, da sie nicht in direktem Kontakt mit den Testleiter stehen.

6.3 Ergebnisse und Interpretation

Im nachfolgenden Abschnitt werden die Ergebnisse des Fragebogens sowie wichtige Punkte des Fokusgruppen-Interviews aufgezeigt und diskutiert.

Die Ergebnisse der Studie bilden im Wesentlichen die drei Fragebögen – zwei im Rahmen der Usability-Studie und einer nach der Pilotphase – und die Aufnahmen des Fokusgruppen-Interviews. Außerdem sind noch die Notizen der Beobachtung während der Usablility-Studie vorhanden.

6.3.1 Auswertung der Fragebögen zur Zufriedenheit

Mit den Ergebnissen des Fragebogens können noch keine konkreten Probleme identifiziert werden. Die Aussagen sind zu allgemein gehalten. Sie sollten vielmehr die Zufriedenheit und die generelle Einschätzung der Probanden bezüglich der Anwendung aufzeigen.

Interpretation der Ergebnisse

Die Auswertung des Fragebogens zur Zufriedenheit nach der Usability Studie und nach der Pilotphase ist in Abbildung 6.2 zu sehen. Laut Analyse der Fragebögen sind alle Testpersonen mit der Bedienung des *SemantLink*-Tools sowohl während der Usability-Studie als auch während der Pilotphase sehr gut zurechtgekommen. Während der Usability-Studie wurden sie von dem Tool bei ihren Aufgaben auch gut unterstützt. Der etwas schlechtere Wert nach der Pilotphase lässt sich darauf zurückführen, dass die Probanden durch die längere und intensivere Beschäftigung in der Pilotphase bei verschiedenen Aufgaben mehr Unterstützung benötigt hätten. Angegeben wurde dabei beispielsweise eine erweiterte Bildersuche in zusätzlichen Portalen, das Einbinden von Videos und eine verbesserte externe Suche bei der je nach Gegenstand für verschiedene Redakteure auch verschiedene Webseiten und Portale durchsucht werden. Das Verwenden des Tools stellte keinen zusätzlichen Aufwand dar, da es nahtlos in das bekannte System integriert ist. Die Mehrheit der Teilnehmer würde das Tool auch in Zukunft für ihre Arbeiten einsetzen, wünschen sich jedoch noch einige Erweiterungen um effizient damit arbeiten zu können. Das *SemantLink*-Tool beinhaltet schon eine Reihe von Features, deckt jedoch noch nicht alle Möglichkeiten einer Hilfestellung der Probanden ab. Die meisten Probanden würden das Tool auch anderen Kollegen zur Erstellung von News und anderen Beiträgen empfehlen.

Fairness und Gültigkeit

Sowohl die Usability-Studie als auch die Pilotphase fand in der gewohnten Umgebung der Testpersonen statt. Die Anwendung wurde direkt in die gewohnte Arbeitsumgebung integriert. In der Pilotphase wurden außerdem, anders als bei der Usability-Studie, die alltäglichen Aufgaben mit dem Tool durchgeführt und keine vorgegebenen.

Zur Darstellung des Fragebogens muss angemerkt werden, dass dieser nach der Usability-Studie von sieben Personen, nach der Pilotphase jedoch nur von vier Personen ausgefüllt wurde.

6.3.2 Auswertung des Fokusgruppen-Interview

Ziel der Fokusgruppe war es, verschiedene Verbesserungsvorschläge und Erweiterungen zu erörtern. Im Folgenden werden wichtige Aussagen der Fokus-

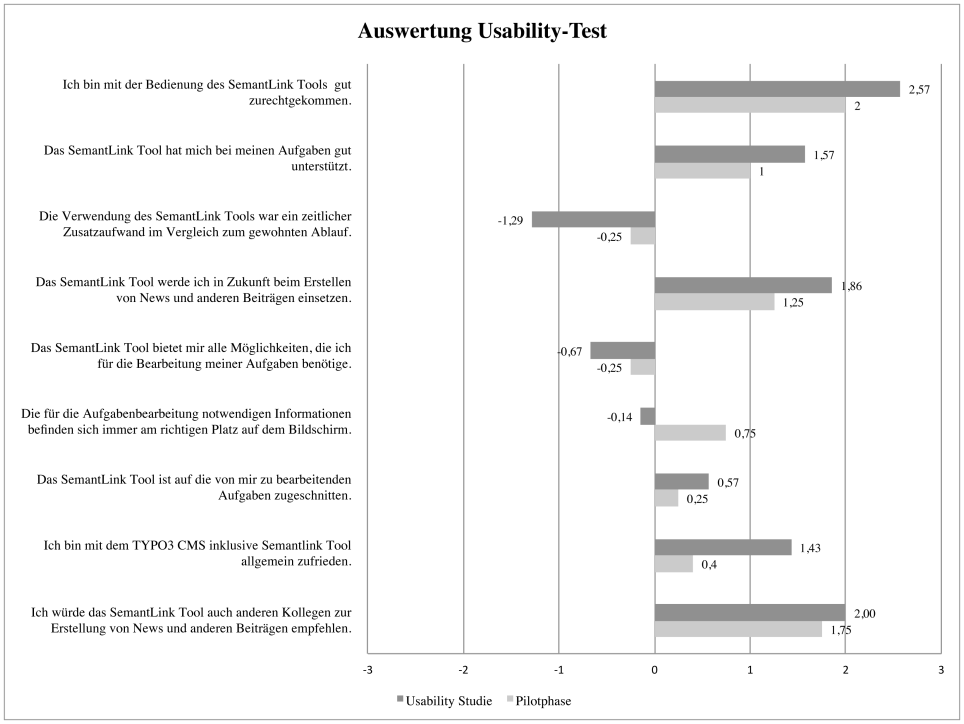

Abbildung 6.2: Auswertung der Fragebögen.

gruppe aufgelistet und interpretiert:

„Ich würde mir die Bilder gerne noch selber schön zusammenschneiden." Die Testpersonen würden es bevorzugen, den Bildausschnitt selbst zu wählen. So könnten sie selbst über das Format bestimmen. Wird ein Bild beispielsweise als Vorschau und im Text verwendet, könnten so unterschiedliche Ausschnitte gewählt werden und man hätte trotzdem einen Wiedererkennungswert.

„Ich schreibe selbst keine News, kann es mir aber sehr gut vorstellen, dass das Tool ein gutes Hilfsmittel ist. Wenn man es nicht nutzen will, zwingt einen ja keiner dazu." Diese Aussage belegt die Erreichung des Ziels, dass das Tool die Anwender in ihren Aufgaben nicht zusätzlich belastet, sondern sich nahtlos in den Prozess integriert. Das Tool ist jederzeit zugänglich, wird es nicht benötigt, behindert es einen aber auch nicht.

„Der Arbeitsfluss wird durch das Tool dynamischer und eleganter." „Es erleichtert gewisse Aufgaben." Ein Bild kann man auch selbst im Web suchen, doch der Prozess, bis das Bild in dem Text eingefügt ist, wird durch das Tool enorm verkürzt. Auch die Quellenangabe steht bereits unterhalb des Bildes und erfordert somit keinen zusätzlichen Aufwand. Außerdem erweitert das Tool das eigene Blickfeld. Häufig denken die Anwender nicht daran, ein gewisses Video oder einen weiterführenden Link in den Text zu integrieren.

Durch die Vorschläge des Tools wird man daran erinnert und erreicht so einen Mehrwert.

Die Möglichkeit des Einfügens eines Kartenausschnitts wurde auch als sehr praktisch angesehen. Besonders beim Veröffentlichen von Veranstaltungen kann dies für die User sehr hilfreich sein. Weitere Features, welche für das Tool interessant und erwünscht sind, wären die Möglichkeit einer Einbettung von YouTube Videos bzw. auch eine Vorschau der eigenen, in der Datenbank gespeicherten Videos. Des Weiteren wäre eine Personalisierung bzw. eine Erweiterung der Suche erwünscht. Für verschiedene Redakteure sind unterschiedliche Portale und Webseiten von Interesse. Es wäre gewünscht, die Suche im Vorfeld auf diese auszurichten.

Kapitel 7

Schlussbemerkung

In dieser Arbeit wurden verschiedene Ansätze einer Unterstützung von Redakteuren in CMS mit Hilfe des Semantic Web vorgestellt. Zu einem dieser Ansätze wurde ein konkreter Anwendungsfall vorgestellt und anschließend evaluiert. Ziel dabei war es, die Redakteure bestmöglich in ihren täglichen Aufgaben zu unterstützen, indem verschiedene Prozesse durch den Einsatz des Semantic Web vereinfacht wurden. Der Arbeitsfluss sollte dadurch aber nicht beeinträchtigt oder gestört werden.

7.1 Fazit

Der entwickelte Prototyp hat gezeigt, dass man mit Hilfe der Technologien des Semantic Web viele Prozessschritte in redaktionellen Arbeiten inhaltlich anreichern bzw. zeitlich verkürzen kann. Die Anwendung bietet bereits jetzt in vielen Bereichen gute Unterstützung und wurde von den Testpersonen gut angenommen.

Da es sich bei der umgesetzten Lösung nur um eine frühe Version eines Prototypen handelt, gibt es diesbezüglich noch viele Verbesserungspotenzial und Möglichkeiten der Erweiterung. Bereiche wie Personalisierung der einzelnen Suchportale für verschiedene Redakteure und auch detailliertere Bereiche wie das Zuschneiden von Bildern bieten Möglichkeiten einer Weiterentwicklung.

Zusätzlich zur Usability-Studie hätte vorher eine heuristische Evaluation mit einer kleinen Gruppe von Experten durchgeführt werden können. Somit hätten einzelne Probleme bereits vorzeitig aufgedeckt werden können und die Probanden wären davon nicht abgelenkt gewesen. Sie hätten ihre Konzentration noch mehr auf die Verbesserungsvorschläge lenken können, was wiederum zu detaillierteren Ergebnissen geführt hätte.

Ein großes Potenzial an Verbesserung bieten auch die Semantic APIs, besonders im Bereich der Sprachvielfalt. Viele bestehende APIs sind aktuell noch auf eine oder wenige Sprachen beschränkt.

7.2 Ausblick

Semantic Web und insbesondere Semantic APIs können vermehrt eingesetzt werden, um nicht nur Benutzer, sondern auch Redakteure zu unterstützen und bei verschiedenen Aufgaben Hilfestellung zu bieten. Wie die vorgestellten Ansätze gezeigt haben, gibt es noch viele Möglichkeiten, Redakteuren ihren Arbeitsprozess zu erleichtern und somit für sie auch die Usabilty in CMS zu verbessern. Bei dem erstellten Anwendungsbeispiel handelt es sich um einen ersten Prototypen, bei dem es noch viel Optimierungspotenzial gibt. Unter anderem können die aus der Usability Studie hervorgegangen Verbesserungs-, Anpassungs- und Erweiterungsmöglichkeiten eingearbeitet werden. Mit Hilfe weiterer Evaluierungen kann mit dieser Anwendung eine optimale Lösung zur Textanreicherung entwickelt werden. Besondere Beachtung sollte auch der Interaktion der Redakteure mit dem Tool gegeben werden. Hier besteht noch viel Verbesserungspotenzial, um die verschiedenen Anreicherungsverfahren möglichst einfach und für sich selbst sprechend zu gestalten. Um den Prototypen für verschiedene Betriebe einsatzfähig zu machen, sollte des Weiteren die Konfiguration einfacher gestaltet und eventuell erweitert werden.

Anhang A

Inhalt der CD-ROM

A.1 Masterarbeit

Pfad: /

Masterarbeit.pdf	Einsatz einer semantischen Textanalyse zur Unterstützung von Redakteuren in Content Management Systemen

A.2 Bilder

Pfad: /Bilder

*.pdf, *.png	Alle verwendeten Grafiken dieser Masterarbeit

A.3 Evaluierung

Pfad: /Evaluierung

Ergebnisse.xlsx	Ergebnisse aller Fragebögen
FokusgruppenLeitfaden.pdf	Leitfaden für das Fokusgruppen-Interview
texte.txt	Verwendete Texte der Usability Studie
UsabilityStudie.pdf . . .	Unterlagen inkl. Fragebögen der Usability Studie

A.4 Online-Quellen

Pfad: /Onlinequellen

BawakidAndOussalah2008.pdf	A Semantic Summarization System: University of Birmingham at TAC 2008

ChoudharyAndBhattacharyya2002.pdf	Text Clustering using Semantics
Hassan-MonteroAnd Herrero-Solana2006.pdf	Improving Tag-Clouds as Visual Information Retrieval Interfaces
IsonormFragebogen.pdf	Beurteilung der Software-Ergonomie anhand des ISONORM-Fragebogens
OpenCalais.pdf	Open Calais – How Does Calais Work?
SemanticWebStack.png	W3C Semantic Web Stack
TCAReference.pdf . . .	TCA Reference Release 6.1.0
TYPO3ReleaseNotes.pdf	TYPO3 6.0 Release Notes
UsabilityHeuristics.pdf .	10 Usability Heuristics for User Interface Design
W3C_OWL.pdf	W3C – OWL Web Ontology Language Overview
W3C_RDFS.pdf	W3C – RDF Vocabulary Description Language 1.0: RDF Schema
W3C_WCAG.pdf . . .	W3C – Richtlinien für barrierefreie Webinhalte (WCAG) 2.0
ZemantaService.pdf . .	Zemanta service: Everything you need to know about Zemanta API beside the specification

A.5 SemantLink

Pfad: /SemantLink

README.rtf	README mit Installationsanleitung und Abhängigkeiten
semantlink.zip	TYPO3 Erweiterung SemantLink
tinymce_rte_0.8.1.t3x	TinyMCE RTE, abhängiger Editor für die TYPO3 Erweiterung SemantLink

Quellenverzeichnis

Literatur

[1] Andrea Bauer. „Analyse von Semantic Web-APIs und deren Methoden“. Masterarbeit. Hagenberg, Austria: Interactive Media; FH Oberösterreich – Fakultät für Informatik, Kommunikation und Medien, 2011.

[2] Abdullah Bawakid und Mourad Oussalah. „A Semantic-Based Text Classification System“. In: *Cybernetic Intelligent Systems (CIS), 2010 IEEE 9th International Conference on*. 2010, S. 1–6.

[3] Abdullah Bawakid und Mourad Oussalah. *A Semantic Summarization System: University of Birmingham at TAC 2008*. 2008. URL: http://www.nist.gov/tac/publications/2008/participant.papers/abawakid.proceedings.pdf.

[4] Tim Berners-Lee und Mark Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. 1. Aufl. Harper San Francisco, 1999.

[5] Tim Berners-Lee, James Hendler und Ora Lassila. „The Semantic Web“. In: *Scientific American* 284.5 (2001), S. 34–43.

[6] Klaus Birkenbihl. „Standards für das Semantic Web“. In: *Semantic Web: Webe zur vernetzten Wissensgesellschaft*. Hrsg. von Tassilo Pellegrini und Andreas Blumauer. Berlin, Heidelberg: Springer, 2006, S. 9–25.

[7] Stephan Bloehdorn und Andreas Hotho. „Boosting for Text Classification with Semantic Features“. In: *Proceedings of the 6th international conference on Knowledge Discovery on the Web: Advances in Web Mining and Web Usage Analysis*. WebKDD'04. Berlin, Heidelberg: Springer-Verlag, 2006, S. 149–166.

[8] Andreas Blumauer und Tassilo Pellegrini. „Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen“. In: *Semantic Web: Webe zur vernetzten Wissensgesellschaft*. Berlin, Heidelberg: Springer-Verlag, 2006, S. 9–25.

[9] Bhoopesh Choudhary und Pushpak Bhattacharyya. *Text Clustering using Semantics*. 2002. URL: http://www2002.org/CDROM/poster/79.pdf.

[10] John M. Conroy und Dianne P. O'leary. „Text summarization via hidden Markov models". In: *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. (New Orleans, Louisiana, USA). SIGIR '01. New York, NY, USA: ACM, 2001, S. 406–407.

[11] Mita K. Dalal und Mukesh A. Zaveri. „Automatic Text Classification: A Technical Review". In: *International Journal of Computer Applications* 28.2 (2011), S. 37–40.

[12] Heather W. Desurvire. „Faster, Cheaper!! Are Usability Inspection Methods as Effective as Empirical Testing?" In: *Usability Inspection Methods*. Hrsg. von Jakob Nielsen und Robert L. Mack. New York, NY, USA: John Wiley & Sons, 1994, S. 173–202.

[13] Fefie Dotsika. „Semantic APIs: Scaling up towards the Semantic Web". In: *International Journal of Information Management* 30.4 (2010), S. 335–342.

[14] H. P. Edmundson. „New Methods in Automatic Extracting". In: *Journal of The ACM* 16.2 (1969), S. 264–285.

[15] Gonenc Ercan und Ilyas Cicekli. „Lexical Cohesion Based Topic Modeling for Summarization". In: *Proceedings of the 9th international conference on Computational linguistics and intelligent text processing*. CICLing'08. Berlin, Heidelberg: Springer-Verlag, 2008, S. 582–592.

[16] Udo Hahn und Inderjeet Mani. „The Challenges of Automatic Summarization". In: *IEEE Computer* 11 (2000), S. 29–36.

[17] Stephan Hamberger. „Entitätenextraktion und Relevanzbewertung mit Hilfe von semantischen Datenquellen und APIs". Masterarbeit. Hagenberg, Austria: Interactive Media; FH Oberösterreich – Fakultät für Informatik, Kommunikation und Medien, 2012.

[18] Yusef Hassan-Montero und Víctor Herrero-Solana. *Improving Tag-Clouds as Visual Information Retrieval Interfaces*. Okt. 2006. URL: http://www.yusef.es/improving_tagclouds.pdf.

[19] Marti A. Hearst und Daniela Rosner. „Tag Clouds: Data Analysis Tool or Social Signaller?" In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*. 2008, S. 160.

[20] Pascal Hitzler u. a. *Semantic Web: Grundlagen*. Berlin: Springer, 2008.

[21] Peter Jackson und Isabelle Moulinier. *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*. Bd. 5. Natural Language Processing. Benjamins, 2002.

[22] Hak Lae Kim u. a. „The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies“. In: *Proceedings of the 2008 International Conference on Dublin Core and Metadata Applications*. Berlin, Deutschland: Dublin Core Metadata Initiative, 2008, S. 128–137.

[23] Richard A. Krueger und Mary Anne Casey. *Focus Group: A Practical Guide for Applied Research*. 3. Aufl. Sage Publications, Inc., 2000.

[24] Julian Kupiec, Jan Pedersen und Francine Chen. „A trainable document summarizer“. In: *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*. 1995, S. 68–73.

[25] Hans P. Luhn. „The Automatic Creation of Literature Abstracts“. In: *IBM Journal* 2.2 (1958), S. 159–165.

[26] Sumit D. Mahalle und Ketan Shah. „Document Clustering by using Semantics“. In: *International Journal of Scientific & Engineering Research* 3.9 (2012), S. 1–5.

[27] Inderjeet Mani. *Automatic Summarization*. Natural Language Processing 3. Amsterdam: John Benjamins, 2001.

[28] Inderjeet Mani. „Summarization Evaluation: An Overview“. In: *Proceedings of the Second NTCIR Workshop on Research in Chinese & Japanese Text Retrieval and Text Summarization* (2001), S. 77–85.

[29] Céline Mariage, Jean Vanderdonckt und Costin Pribeanu. „State of the Art of Web Usability Guidelines“. In: *The Handbook of Human Factors in Web Design*. Hrsg. von Kim-Phuong L. Vu und Robert W. Proctor. CRC Press, 2005, S. 688–700.

[30] Ivo Marinchev. „Practical Semantic Web – Tagging and Tag Clouds“. In: *Cybernetics and Information Technologies* 6.3 (2006), S. 33–39.

[31] Pablo N. Mendes u. a. „DBpedia Spotlight : Shedding Light on the Web of Documents“. In: *Proceedings of the 7th International Conference on Semantic Systems*. I-Semantics '11. ACM, 2011, S. 1–8.

[32] Rada Mihalcea und Andras Csomai. „Wikify! Linking Documents to Encyclopedic Knowledge“. In: *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. CIKM '07. New York, NY, USA: ACM, 2007, S. 233–242.

[33] Gary Miner u. a. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Academic Press, 2012.

[34] Roberto Mirizzi u. a. „Semantic Tag Cloud Generation via DBpedia“. In: *Electronic Commerce and Web Technologies (EC-Web)*. Hrsg. von Francesco Buccafurri und Giovanni Semeraro. Bd. 61. Lecture Notes in Business Information Processing. Springer-Verlag, 2010, S. 36–48.

[35] David L. Morgan. *Focus Groups as Qualitative Research*. Sage Publications, Inc., 1997.

[36] Günter Neumann. „Informationsextraktion“. In: *Computerlinguistik und Sprachtechnologie – Eine Einführung*. Hrsg. von K. U. Carstensen u. a. Heidelberg, Berlin: Spektrum Akademischer Verlag, 2001, S. 448–456.

[37] Jakob Nielsen. „Heuristic Evaluation“. In: *Usability Inspection Methods*. Mack, Robert L. und Nielsen, Jakob, 1994. Kap. 2, S. 25–62.

[38] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, 1993.

[39] Jakob Nielsen und Hoa Loranger. *Web Usability*. Addison-Wesley, 2006.

[40] Rolf Porst. *Fragebogen: Ein Arbeitsbuch*. 3. Aufl. Verlag für Sozialwissenschaften, 2011.

[41] Jochen Prümper. „Der Benutzungsfragebogen ISONORM 9241/10: Ergebnisse Zur Reliabilität und Validität.“ In: *Software-Ergonomie '97: Usability Engineering: Integration von Mensch-Computer-Interaktion und Software-Entwicklung*. Hrsg. von Rüdiger Liskowsky, Boris M. Velichkowsky und Wolfgang Wünschmann. Bd. 49. Berichte des German Chapter of the ACM. Stuttgart: Teubner, 1997, S. 253–262.

[42] Jochen Prümper und Michael Anft. „Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung – ein Fallbeispiel“. In: *Software-Ergonomie '93: Von der Benutzeroberfläche zur Arbeitsgestaltung*. Hrsg. von Karl-Heinz Rödiger. Bd. 39. Berichte des German Chapter of the ACM. Stuttgart: Teubner, 1993, S. 145–156.

[43] Jochen Rau und Sebastian Kurfürst. *Zukunftssichere TYPO3-Extensions mit Extbase & Fluid*. O'Reilly Verlag, 2010.

[44] Markus Richter und Flückiger Michael D. *Usability Engineering kompakt: Benutzbare Software gezielt entwickeln*. 2. Aufl. Spektrum, 2010.

[45] A. W. Rivadeneira u. a. „Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. New York, NY, USA: ACM, 2007, S. 995–998.

[46] Mary Beth Rosson und John M. Carroll. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Morgan Kaufmann, 2002.

[47] Florian Sarodnick und Henning Brau. *Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung*. Huber, Bern, 2006.

[48] Johann Schrammel, Michael Leitner und Manfred Tscheligi. „Semantically Structured Tag Clouds: An Empirical Evaluation of Clustered Presentation Approaches“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. New York, NY, USA: ACM, 2009, S. 2037–2040.

[49] Werner Schweibenz und Frank Thissen. *Qualität im Web: Benutzerfreundliche Webseiten durch Usability-Evaluation*. Berlin, Heidelberg: Springer-Verlag, 2003.

[50] Michael Scriven. „The Methodology of Evaluation“. In: *Perspectives of Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation*. Hrsg. von R. Tyler, R. Gagné und M. Scriven. Bd. 1. Chicago, USA: Rand McNally, 1967, S. 39–83.

[51] Khaled B. Shaban. „A Semantic Approach for Document Clustering“. In: *Journal of Software* 4.5 (2009), S. 391–404.

[52] James Sinclair und Michael Cardew-Hall. „The folksonomy tag cloud: when is it useful?“ In: *Journal of Information Science* 34.1 (2008), S. 15–29.

[53] Cathleen Wharton u. a. „The Cognitive Walkthrough Method: A Practitioner's Guide“. In: *Usability Inspection Methods*. Hrsg. von Jakob Nielsen und Robert L. Mack. New York: John Wiley & Sons, 1994, S. 105–140.

Online-Quellen

[54] Lothar Bräutigam. *Beurteilung der Software-Ergonomie anhand des ISONORM-Fragebogens*. 2008. URL: http://www.ergo-online.de/site.aspx?url=html/software/verfahren_zur_beurteilung_der/beurteilung_der_software_ergo.htm (besucht am 29.09.2013).

[55] Dan Brickley und R. V. Guha. *W3C. RDF Vocabulary Description Language 1.0: RDF Schema*. 2004. URL: http://www.w3.org/TR/rdf-schema/ (besucht am 27.09.2013).

[56] Ben Caldwell u. a. *W3C. Richtlinien für barrierefreie Webinhalte (WCAG) 2.0*. 2008. URL: http://www.w3.org/Translations/WCAG20-de/ (besucht am 27.09.2013).

[57] *How Does Calais Work?* URL: http://www.opencalais.com/about (besucht am 29.09.2013).

[58] Helmut Hummel. *TYPO3 6.0 Release Notes*. URL: http://typo3.org/download/release-notes/typo3-60-release-notes/ (besucht am 27.09.2013).

[59] Deborah L. McGuinness und Frank van Harmelen. *W3C. OWL Web Ontology Language Overview*. 2004. URL: http://www.w3.org/TR/owl-features/ (besucht am 27.09.2013).

[60] Jakob Nielsen. *10 Usability Heuristics for User Interface Design*. 1995. URL: http://www.nngroup.com/articles/ten-usability-heuristics/ (besucht am 27.09.2013).

[61] *TCA Reference Release 6.1.0*. 2013. URL: http://docs.typo3.org/typo3cms/TCAReference/Index.html (besucht am 27.09.2013).

[62] Andraž Tori und Tomaž Šolc. *Zemanta service: Everything you need to know about Zemanta API beside the specification*. 2011. URL: http://developer.zemanta.com/media/files/docs/zemanta_api_companion.pdf (besucht am 27.09.2013).

[63] *W3C Semantic Web Stack*. 2007. URL: http://www.w3.org/2007/03/layerCake.png (besucht am 27.09.2013).