

# Digitale Kuratierungstechnologien

## Verfahren für die effiziente Verarbeitung, Erstellung und Verteilung qualitativ hochwertiger Medieninhalte

**Georg Rehm**

DFKI GmbH

Forschungsbereich Sprachtechnologie

Alt-Moabit 91c

10559 Berlin

georg.rehm@dfki.de

**Felix Sasaki**

DFKI GmbH

Forschungsbereich Sprachtechnologie

Alt-Moabit 91c

10559 Berlin

felix.sasaki@dfki.de

### 1 Einleitung: Digitale Kuratierung

Das Kuratieren digitaler Informationen, Daten, Meldungen und Medieninhalte hat sich in den vergangenen Jahren als eine grundlegende Tätigkeit mit neuen Anforderungen herauskristallisiert, die von handelsüblichen Content-Management-Systemen schon längst nicht mehr abgedeckt werden. Kuratieren ist ein komplexer wissens- und zeitintensiver Prozess, in dem Redakteure oder interdisziplinäre Teams aus heterogenen Quellen ein neues, abgestimmtes Gesamtwerk entwickeln, das auf einen spezifischen Fokus ausgerichtet ist. Die hierzu erforderlichen Arbeiten umfassen das Auswählen, Zusammenfassen, zeitliche Einordnen, Internationalisieren, Anreichern, Visualisieren und Erklären der verschiedenen Inhalte, wobei insbesondere zu berücksichtigen ist, dass Geschwindigkeit, Volumen und Anzahl der Quellen (Online-Zeitungen, Nachrichtenportale, Twitter, Facebook, Instagram etc.) sowie der zu verarbeitenden Information stetig anwachsen. Ein Beispiel ist die Entwicklung eines interaktiven Exponats für ein Besucherzentrum, das bei Ausgrabungen gefundene Objekte mit Fotos, Beschreibungen und Zeitangaben auf einer Karte visualisiert und die Auswahl geeigneter Objekte, Erstellung entsprechender Inhalte, Gestaltung der Karte und Festlegung thematischer Perspektiven erfordert.

### 2 Projektüberblick

Dieser Beitrag gibt einen kurzen Überblick über das Verbundprojekt „Digitale Kuratierungstechnologien“, an dem die vier in Berlin ansässigen Unternehmen art+com AG, Condat AG, 3pc GmbH und kreuzwerker GmbH sowie das DFKI

als Forschungspartner teilnehmen und das voraussichtlich ab dem 1. September 2015 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wird.<sup>1</sup> Das Ziel des zweijährigen Vorhabens ist es, die komplexen, von Redakteuren und Wissensarbeitern durchgeführten digitalen Kuratierungsprozesse durch Sprach- und Wissenstechnologien zu unterstützen.

Das DFKI wird Komponenten aus diesem Bereich einbringen und weiterentwickeln und gemeinsam mit den vier KMU-Partnern zu einer Plattform für digitale Kuratierungstechnologien ausbauen, die Funktionen zur Recherche, Anreicherung, Analyse, Kombination (z.B. thematisch, chronologisch, räumlich), Zusammenfassung und Internationalisierung von Inhalten umfasst. Branchen- und Plattformtechnologien werden die Realisierung branchenspezifischer Workflows und skalierbarer Anwendungen in den jeweiligen Branchen vereinfachen. Die Plattform ermöglicht den Industriepartnern, innovative und effizienz- sowie qualitätssteigernde Lösungen für vier unterschiedliche Branchen (Museen und Showrooms; TV-/Radio und Web-TV; Verlage und Medienhäuser; Archive und Bibliotheken) effizienter zu entwickeln, zu betreiben, zu integrieren und zu verwerten.

Die vom Forschungspartner DFKI eingebrachten Technologien umfassen Methoden, Komponenten und Ansätze aus dem Gebiet der Sprach- und Wissenstechnologien, die im Rahmen zahlreicher Projekte wie z.B. ATLAS, COLLATE, LT Web, META-NET, QTLaunchPad, EuroMatrix, EuroMatrixPlus und Trendminer (BMBF, BMWi,

---

<sup>1</sup> Siehe <http://artcom.de>, <http://condat.de>, <http://3pc.de>, <http://kreuzwerker.de> sowie <http://dfki.de/lt>.

EU/EC etc.) entwickelt wurden. Diese Methoden können den folgenden drei Bereichen zugeordnet werden:

1. *Semantische Analyse*: Tiefe Analyse mit hoher Präzision und der Möglichkeit zur Adaption an verschiedene Domänen am Beispiel von Informationsextraktion (Zeiten, Orte, Themen, generische benannte Entitäten), automatische Textzusammenfassung, Sentiment-Analyse sowie Klassifikation und Clustering von Informationen.
2. *Semantische Generierung*: Unterstützung des Storytellings durch Text-, Hypertext- und Reportgenerierung für ausgewählte Typen von Dokumenten auf Basis von Verfahren zur Informationsextraktion und generischen Textschemata, die z.B. als thematische oder textsortenspezifische Strukturgrammatiken repräsentiert werden.
3. *Mehrsprachige Technologien*: Robuste und adaptierbare Komponenten für maschinelle Übersetzung sowie Integration verschiedener Wissensquellen unter Berücksichtigung der Kuratierungs-Workflows bei den Industriepartnern für eingehende (Inbound-Translation) und zu publizierende Dokumente (Outbound-Translation) sowie Integration mono- und multilingualer Linked-Open-Data-Quellen (LOD).

### 3 Technologieplattform

An dieser Stelle soll die Bedeutung der Technologieplattform im Rahmen der Wertschöpfungskette hervorgehoben werden. Je nach Anwendungsfall und Branche fällt die Wertschöpfungskette zur Kuratierung von Inhalten unterschiedlich aus. Es sind drei Arten von Akteuren zu unterscheiden:

- die kuratierende Institution, z.B. Museum, Fernsehsender, Verlag oder Archiv;
- Dienstleister/Agenturen, die für die kuratierende Institution Inhalte und Technologien bereit stellen bzw. Komplettlösungen entwickeln (z.B. die vier KMU-Partner);
- an der Kuratierung beteiligte Redakteure und Wissensarbeiter, z.B. interne Mitarbeiter oder Dienstleister, aber auch externe Wissenschaftler, Experten oder Freiberufler.

Die Plattform für digitale Kuratierungstechnologien soll diesen unterschiedlichen Akteuren eine umfassende Menge von Funktionalitäten bieten, die den Kuratierungsprozess unterstützen. Durch den Einsatz von Sprach- und Wissenstechnologien können einzelne, bisher noch rein manuell

bzw. intellektuell durchgeführte Kuratierungstätigkeiten (teil-)automatisiert werden. Die Akteure können durch die Nutzung der Plattform größere Mengen an Inhalten schneller sichten und weiterverarbeiten. Mit der Technologieplattform wird eine deutliche Effizienzsteigerung und Kostensenkung des Kuratierungsprozesses angestrebt – bei gleichbleibender oder sogar verbesserter Qualität. Dabei können, wie in Abb. 1 dargestellt, vier Schichten unterschieden werden, nämlich Kernkomponenten der Sprach- und Wissenstechnologie, allgemeine Kuratierungskomponenten, allgemeine Plattformkomponenten sowie Branchentechnologien.

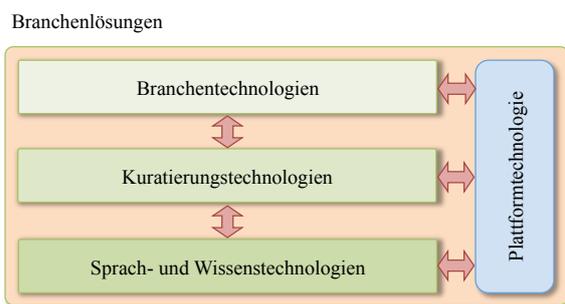


Abbildung 1: Plattform für Kuratierungstechnologien

Das DFKI wird die bereits vorhandenen Komponenten so weiter entwickeln, dass sie den Anforderungen der anderen Schichten genügen und von den Industriepartnern in ihre jeweiligen Branchenlösungen integriert werden können. Die Industriepartner wiederum konzipieren und entwickeln generische Technologiekomponenten, die für ihre Branchenanwendungen benötigt werden, aber auch in anderen Lösungen eingesetzt werden können. Wir fokussieren insbesondere die folgenden Zielmerkmale der Plattform:

- Vollintegrierte robuste, performante und skalierbare Komponenten mit offenen APIs für eine effiziente Einbettung in branchenspezifische Kuratierungs-Workflows;
- Einfache Nutzbarkeit der Cloud-Plattform durch browserbasierte SaaS-Webarchitektur;
- Anwendungsorientierte Branchenlösungen mit hoher Usability (User Interfaces, Interaktionsdesign, Informationsvisualisierung).

### 4 Schlussfolgerungen

Unsere grundlegende Arbeitshypothese ist, dass der gezielte Einsatz sprachtechnologischer Verfahren digitale Kuratierungsprozesse deutlich effizienter und produktiver gestalten kann und sind überzeugt, diese Hypothese gemeinsam mit den KMU-Partnern im Rahmen des hier knapp skizzierten Vorhabens verifizieren zu können.