

Kurzfassung der Dissertation

Rapid-Content-Integration

Ein Ansatz zur schnellen Integration
von Inhalten in Content-Management-Systeme

Dipl.-Ing. Kai Schmitz-Hofbauer

Lehrstuhl für Software-Technik, Ruhr-Universität Bochum

Sowohl Menge als auch Vielfalt der täglich in Unternehmen zu erfassenden und zu bearbeitenden Inhalte steigen ständig. Viele dieser Inhalte stammen aus den unterschiedlichsten Quellen und beinhalten wichtige Informationen. Dies trifft in hohem Maße auf Inhalte in Form von weniger stark strukturierten Informationen wie beispielsweise Texte, Abbildungen und multimediale Inhalte zu. Zur Erstellung digitaler Inhalte kommt in Abhängigkeit von dem jeweiligen Anwendungshintergrund eine Vielzahl von Anwendungssystemen zum Einsatz. Ein beträchtlicher Prozentsatz dieser Systeme speichert die Inhalte in einer dateigebundenen Dokumentenform in einem bestimmten Dokumentenformat. Derartige Inhalte sind nur schwierig zentral nutzbar und für weitere Anwendungsszenarios verwendbar. Insbesondere gestaltet sich eine dokumentenformatübergreifende konsistente Wiederverwendung der Inhalte schwierig.

In der vorliegenden Arbeit wird der Rapid-Content-Integration-Ansatz vorgestellt, welcher die Integration von Inhalten, die in Form von digitalen Dokumenten vorliegen, in ein zentrales Content-Repository ermöglicht. Als zentrales Forschungsergebnis wird ein flexibler und an verschiedene Ausgangssituationen anpassbarer **Integrationsprozess** vorgestellt. Dieser nimmt eine „Anpassung“ und „Überführung“ von Inhalten beliebiger Dokumentenformate in die Strukturen von Content-Management-Systemen vor, so dass die Inhalte dort entsprechend den technischen Möglichkeiten und betrieblichen Notwendigkeiten zielgerichtet weiterverwendet werden können.

Den Ausgangspunkt des Integrationsprozesses bilden digitale Dokumente, die mit unterschiedlichsten Anwendungssystemen erstellt wurden. Die Ausführung bestimmter Prozessschritte erfolgt in Abhängigkeit von einem konkreten Format. Unbekannte Dokumentenformate müssen zunächst in ein vereinheitlichtes, verarbeitbares Format überführt werden. Dies ist eine Aufgabe der **Formatharmonisierung**. Sollten in dem Ursprungs- bzw. harmonisierten Dokument die für eine Weiterverarbeitung notwendigen logischen Dokumentstrukturen fehlen, so werden diese im Rahmen einer **semi-automatischen Strukturerkennung** zunächst mit Unterstützung eines Anwenders aus den typografischen Merkmalen abgeleitet. Den anschließenden Prozessschritt bildet die **Dekomposition**, welche die flache Struktur eines Dokumentes in eine Struktur miteinander verknüpfter Content-Objekte überführt. Diese werden in einem Content-Repository gespeichert und können bei Bedarf auch einzeln wiederverwendet werden. Charakteristisch dabei ist, dass der Ansatz es ermöglicht, Inhalte von den Spezifika der jeweiligen Formate zu befreien, in eine **medienneutrale Darstellung** zu überführen und strukturiert zu speichern. Die Inhalte können – sofern das Zielsystem dies zulässt – auf vielfältige Art und Weise genutzt und medienübergreifend kombiniert, wiederverwendet sowie zur Erzeugung unterschiedlicher Ausgaben verwendet werden. Während bisherige Ansätze aus Forschung und Praxis jeweils nur ein Teilproblem erfassen und lösen, zeichnet sich dieser Ansatz durch eine ganzheitliche Betrachtungsweise aus.

Zur Umsetzung des Integrationsprozesses wird eine flexible, anpassungsfähige, erweiterbare und integrative Architektur entworfen. Die Kernfunktionalitäten des Prozesses werden auf unabhängige Komponenten abgebildet, die von einem **Orchestrierungssystem** flexibel gemäß den jeweiligen Notwendigkeiten verwendet werden. Besonders hervorzuheben ist das Konzept der **generischen Formatharmonisierung** in Form eines virtuellen Druckertreibers, der Inhalte in einer vereinheitlichten Form zur Verfügung stellt und Weiterverarbeitungsmaßnahmen initiiert. Dieser ermöglicht es so, Inhalte über die Druckschnittstelle werkzeug- und formatunabhängig strukturiert in ein Content-Management-System zu überführen.

Die technische Umsetzung des Systems, die Rapid-Content-Integration-Plattform, befindet sich seit mehreren Jahren im praktischen Einsatz. Als *Proof of Concept* werden ausgesuchte Praxisbeispiele vorgestellt. Den Schwerpunkt der Betrachtungen bildet dabei die W3L-Plattform, die sich durch eine medienneutrale Datenhaltung und eine leistungsstarke Publishing-Funktionalität auszeichnet. Mithilfe der W3L-Plattform wird die Praxistauglichkeit des Ansatzes in den Anwendungsbereichen E-Learning, Verlagswesen, *Print On Demand* sowie Erstellung von technischen Dokumentationen und Schulungsunterlagen verdeutlicht.

Zusammenfassend kann festgestellt werden, dass sich durch die Ergebnisse dieser Arbeit eine konsistente, formatübergreifende Wiederverwendbarkeit von Inhalten erreichen lässt und Medienbrüche aufgehoben werden können. Wie die Praxisbeispiele verdeutlichen, lassen sich mithilfe dieses Ansatzes durch Integration dokumentenbasierter Inhalte in Content-Management-Systeme Medien-Produkte in verschiedenen Anwendungsbereichen effizient, effektiv und wirtschaftlich rentabel erzeugen.