

CMS-Kennzahlen mit der REx-Methode

Content Intelligence für Redaktionssysteme

Von Claudia Oberle und Wolfgang Ziegler

Wie nutzen wir unser Redaktionssystem? Ist unsere Arbeitsweise effizient, die Wiederverwendung die erhoffte? Antworten auf Fragen und Wünsche nach quantitativen Kennzahlen waren für Content-Management-Anwendungen bislang selten. Mit dem Report Exchange Format könnte sich das ändern.

Content-Management-Systeme (CMS) gehören für viele Unternehmen zum gewohnten Werkzeug der Dokumentationserstellung. Komplexe Erstellungsprozesse wie das multilinguale Informationsmanagement und die Informationserfassung über verteilte Standorte lassen sich schwer ohne Systemunterstützung bewältigen.

Zu den methodischen Grundlagen des Einsatzes von CMS gehören dabei die Modularisierung und die systematische, das heißt kontrollierte Wiederverwendung von Inhalten. Für die Planung einer Systemeinführung und entsprechender Dokumentationsprozesse wird im Rahmen eines „Content Engineering“ idealerweise ein Modulkonzept entwickelt, das sich zum Beispiel mit Hilfe von Modularisierungsmatrizen darstellen lässt [1]. Aus diesen Vorarbeiten können Abschätzungen über die prinzipiellen Wiederverwendungsmöglichkeiten der modularen Inhalte getroffen und somit grundlegende Argumentationen für die Effizienz und den Nutzwert eines Systemeinsatzes entwickelt werden [2].

In der Regel wird der methodische und quantitative Effizienznachweis einer Systemnutzung im Nachhinein aber selten geführt. (System-) Kennzahlen werden aus Mangel an einer entsprechenden auswertbaren internen Datenbasis kaum erhoben, sowohl vor als auch nach der Systemeinführung. Ausnahmen bilden die quantifizierbaren externen Kosten wie Übersetzungsaufwände oder die Nutzung von weiteren Dienstleistungen bei der Erstellung und Publikation von Inhalten.

Grundlegende Kennzahldefinitionen für Dokumentationsprozesse sind zwar vorhanden [3], lassen

sich aber bisher nicht auf einfache Weise in Redaktionssystemen nutzen. Die Systeme bieten im Rahmen der Prozessunterstützung Funktionen wie den modularen (Wieder-) Verwendungsnachweis und ein kontrolliertes Änderungsmanagement. Auswertungen zum Beispiel von modularen Verwendungshäufigkeiten, Projektkennzahlen oder regelmäßigen Statistiken werden aber lediglich in Einzelfällen erhoben.

Vom Prozessmanagement zu Content Intelligence

Was bisher fehlt, ist eine standardisierte Möglichkeit, die reale Effizienz und generell die Arbeitsweise mit einem Redaktionssystem im Sinne eines „Monitoring“ transparent zu machen. Dazu wäre eine stärkere systemtechnische Unterstützung in den folgenden Phasen hilfreich:

- Pre-Documentation („Plan“): Planung von modulbasierten Dokumentationsprojekten für Arbeitsgruppen und die verteilte Inhaltserstellung
- Documentation Process („Do“): Dokumentationserstellung mit Projektmonitoring und -management; aktive (System-)Unterstützung der modularen Wiederverwendung und der Inhaltserstellung
- Post-Documentation („Check“): Kennzahlen und Nutzungsstatisti-



Claudia Oberle ist Lehrbeauftragte an der Hochschule Karlsruhe und Doktorandin an der Universität Mainz. Sie hat einen Bachelor-Abschluss im Studiengang „Translation Studies for Information Technologies“ der Universität Heidelberg und der Hochschule Mannheim. Den Master-Abschluss in Technischer Redaktion hat sie in Karlsruhe absolviert. Seit 2010 ist sie zudem für die icms GmbH tätig.



Dr. Wolfgang Ziegler ist Diplom-Physiker und seit 2003 Professor für Informations- und Content-Management an der Hochschule Karlsruhe. Prof. Ziegler leitet den Master-Studiengang Kommunikation und Medienmanagement. Unternehmen, die sich für die REx-Auswertungen interessieren, können sich direkt an ihn wenden.

ken für die Effizienzkontrolle und Charakterisierung

Aus Sicht des Qualitäts- und Prozessmanagements entspricht dies den Anforderungen aus dem klassischen PDCA-Zyklus („Plan-Do-Check-Act“) nach Deming [4]. Nach der Post-Documentation-Phase („Check“) sollte dabei die Optimierung („Act“) der Content-Management-Prozesse und der Arbeitsweisen folgen – Abbildung 1. In unserem Beitrag wird vorrangig die Post-Documentation-Phase betrachtet, um zu CMS-Kennzahlen zu gelangen, die eine Optimierung erst ermöglichen.

Auch aus wissenschaftlicher Sicht ist es wünschenswert, für CMS eine standardisierte Definition von Kennzahlen im Sinne einer quantitativen Metrik einzuführen. Durch systematische Untersuchungen einer möglichst großen Zahl an Implementierungen sollten statistische Erkenntnisse zur langfristigen CMS-Nutzung zum Beispiel für unterschiedliche Branchen und Unternehmensgrößen gewonnen werden können. Um dies zu realisieren, wurde der in den nächsten Abschnitten dargestellte Report-Exchange (REx)-Mechanismus eingeführt. Mit diesem sollten dann auch empirische, aber statistisch bisher nicht verifizierte Aussagen belegt werden, zum Beispiel zur Wiederverwendung im Verhältnis zur Modulgröße oder zu den Einsatzmöglichkeiten eines feinmodularen Variantenmanagements [1].

Ausgangspunkt der eingesetzten CMS-Metrik sind einfache Wiederverwendungs- und Systemkennzahlen, die ein statisches Bild des CMS zu einem festen Zeitpunkt (Post-Documentation) vermitteln [5]. Die Auswertung dieser Kennzahlen ist bereits technisch realisiert. Als Erweiterung sollen dynamische Kennzahlen als Planungskennzahlen (Pre-Documentation), Verlaufs- und Projektmanagementgrößen (Documentation Process) dargestellt werden. Die (system-)technischen Analysen sind konzeptionell ebenfalls mit semantischen oder linguistischen Methoden zu kombinieren, wie sie bereits im Bereich der Controlled-Language-Checker und der Authoring-Memory-Funktionen genutzt werden. Die Gesamtheit der linguistischen und technischen Analyse- beziehungs-

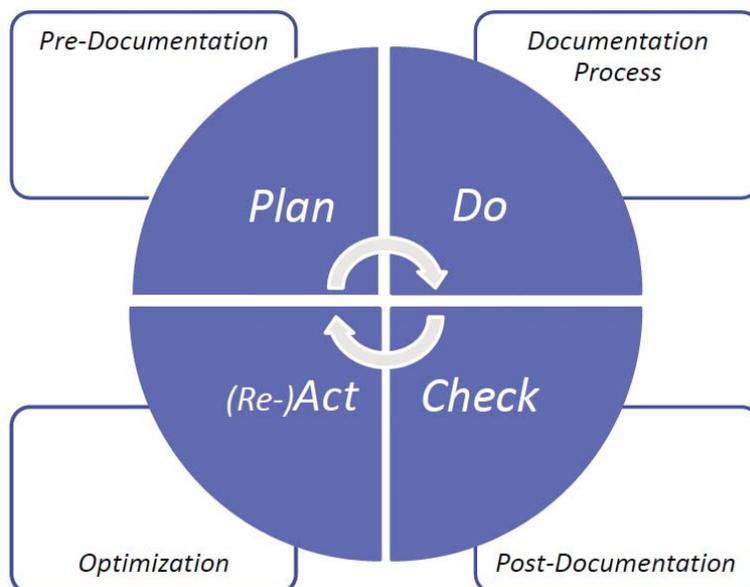


Abb. 1: PDCA-Zyklus nach Deming angewandt auf die Dokumentationsphasen; Optimierung mit Hilfe von „Content Intelligence“-Methoden

weise Reporting-Verfahren lässt sich – in Analogie zu Business Intelligence zum Beispiel für kaufmännische Unternehmensdaten – daher als „Content Intelligence“-Methode bezeichnen.

CMS-Daten im REx-Format

Die technologische Grundlage der Kennzahlenerhebung ist derzeit eine XML-Exportdatei, die eine Reihe von CMS-Basisdaten beinhaltet (Abbil-

dung 2, Seite 50). Im Wesentlichen sind dies die Kennungen der CMS-Objekte (Dokumente, Module und Medien). Mit diesen werden die Informationen geliefert, welche anderen Unterobjekte jeweils referenziert werden. Zusätzlich können Informationen über Sprachkennungen, Versionen, Objektgrößen (z. B. Wortzahl) oder Zeitstempel mitgeliefert werden. Die formale Strukturdefinition einer derartigen Report-Exchange (REx-) Datei wurde als XML-Schema ent-

Spezialist für technische Übersetzungen

Welche Sprachen spricht Ihr Produkt?

...mit unserer Hilfe jede!

ÜBERSETZUNGSBÜRO
Terber & Partner

Friedrich-Ebert-Straße 7 · D-48153 Münster
E-Mail: info@terberundpartner.de · www.terberundpartner.de
Telefon: +49 (0) 251 / 5 20 90 - 0 · Fax: +49 (0) 251 / 5 20 90 - 40

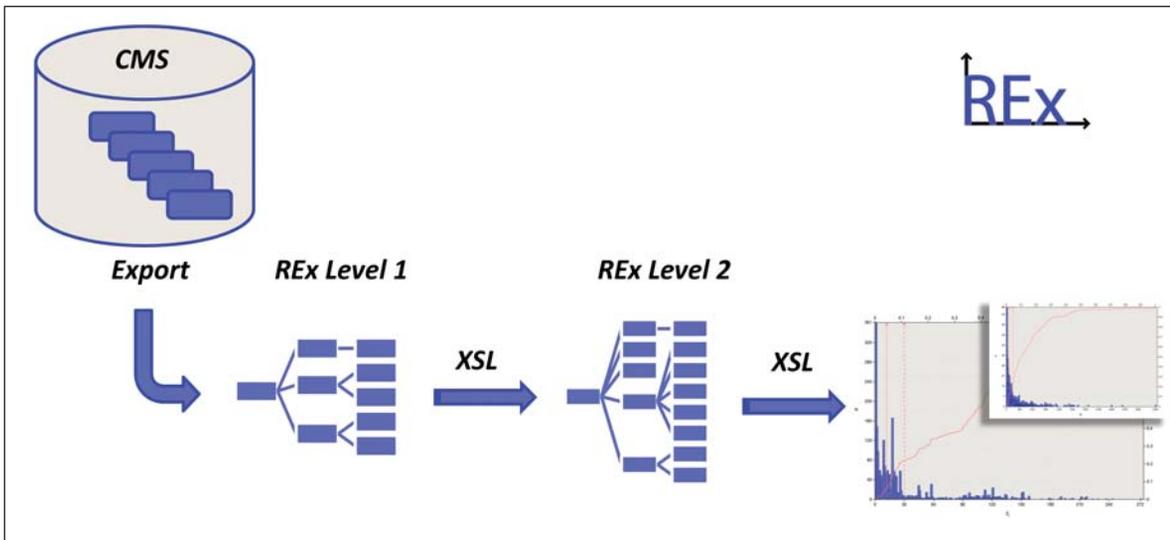


Abb. 2: Logik des Exports von CMS-Daten in das REx-Format (Level 1) und Auswertung (Level 2) mit Visualisierung im REx-Report

wickelt [6, 7] und öffentlich zur Verfügung gestellt [8]. Eine Reihe von CMS-Anbietern hat bereits eine REX-Schnittstelle implementiert und deren Kunden nutzen einen „RExport“ (REx-Export) für erste Kennzahlenermittlungen.

Die eigentliche statistische und grafische Auswertung wird mit Hilfe von XSL-Skripten außerhalb des CMS durchgeführt. Die Skripte entstanden ursprünglich im Rahmen einer Master-These an der HS Karlsruhe [7]. Sowohl der Umfang der REX-Definition als auch die Auswertemöglichkeiten werden in Projekten kontinuierlich weiterentwickelt. Bei der Erweiterung auf die aktuell entstehende REX-Version 1.2 werden zum Beispiel alle weiteren vorhandenen Objektmetadaten auswertbar gemacht. Als Ergebnis erhalten die Systemnutzer eine Übersicht über absolute und gemittelte Systemgrößen, zum Beispiel Modul-, Dokumentzahl, Wiederverwendungsgrad von Dokumenten oder Zahl von Modulverwendungen. Zusätzlich wird jeweils eine grafische Verteilung der gemittelten und weiteren Größen für eine genauere Analyse erstellt. Der Gesamtumfang der ausgewerteten Größen, das heißt die CMS-Metrik, ist jeweils in der REX-Spezifikation definiert [8].

Erstprojekte mit Kooperationspartnern (CMS-Anbieter und CMS-Nutzer) liefern bereits interessante Ergebnisse. Es handelt sich dabei jeweils um anonymisierte Exporte oder auch anonymisierte Auswertungen. Dies ist insofern wichtig, als auch kei-

ne personenbezogenen Daten in REX erfasst oder ausgewertet werden sollen. Es werden prinzipiell nur solche Daten, die im CMS enthalten sind, wie zum Beispiel der Verwendungsnachweis von Objekten, statistisch kumuliert ausgewertet und visualisiert. Für die Diskussion mit dem Betriebsrat ist dies unter Umständen eine wichtige Tatsache.

Typische Ergebnisse

Wie bereits beschrieben, erhalten REX-Anwender unter anderem eine Übersicht über einige grundlegende Systemkennzahlen. Abbildung 3 zeigt ein typisches Analyseergebnis für ein reales CMS mit insgesamt 492 erstellten Dokumenten und einer mittleren Wiederverwendungsrate von

Analyse			Drucken	
Zusammenfassung			Angaben zum CMS	
Dokumente	D	492	Unternehmen	none
Module	M	2275	CMS	none
Verwendete Module	N_{used}	1914	Zeitpunkt	none
Mittelwert: Anzahl Module pro Dokument	\bar{N}_i	150,73	Anwendungszweck	none
Mittelwert: Inhaltsbezogene Wiederverwendungsrate von Dokumenten	\bar{r}_i^c	0,9904	Angaben zur Analyse	
Mittelwert: Modulbezogene Wiederverwendungsrate von Dokumenten	\bar{r}_i^m	0,9927	Typ der Analyse	Kanalauswertung
Mittelwert: Segmente pro Modul	\bar{W}_i	155,26	Sprachen	de
Mittelwert: Gewichtete Verwendungszahl von Modulen	\bar{Z}_i^c	33,94	Dokumenttypen	all
Mittelwert: Verwendungszahl von Modulen	\bar{Z}_i	30,19	Ausgabemedien	all
Gesamtheit in Segmenten	$\bar{W}_i \cdot N$	353000	Nicht verwendete Objekte	enthalten
Wiederverwendeter Inhalt in Segmenten	$Z^{*1} \cdot W_i$	1,17E7	Ausgewertete Einheit für die Segmentgrößen	word
Wiederverwendeter Inhalt in Modulen	Z^{*1}	66771	Angaben zu den Segmentgrößen	
			Einheit "word"	gewicht für
			Dokumente	0%
			Module	100%
			Fragmente	-

99 Prozent. Derart hohe Raten zwischen 80 und 100 Prozent sind nach den bisherigen Daten nicht untypisch für CMS-Anwendungen im Maschinen- und Anlagenbau. Fast alle Dokumente bestehen dann aus bereits in anderen Dokumenten verwendete-

Abb. 3: Anonymisierte Zusammenfassung von gemittelten Basiskennzahlen einer CMS-Implementierung

SPRACHEN SIND
UNSERE WERKZEUGE.

ÜBERSETZUNGEN?
PLUSLINGUA!

ISO 9001:2008 und DIN EN 15038:2006 zertifiziert

+49 (0)89 741185-197

www.pluslingua.de

Abb. 4: Verteilung der Modulgrößen W in gezählten Worten; Balkenhöhe gibt die Zahl N von Modulen einer Größe an: Häufung kleiner Module und starker Abfall der Häufigkeit großer Module.

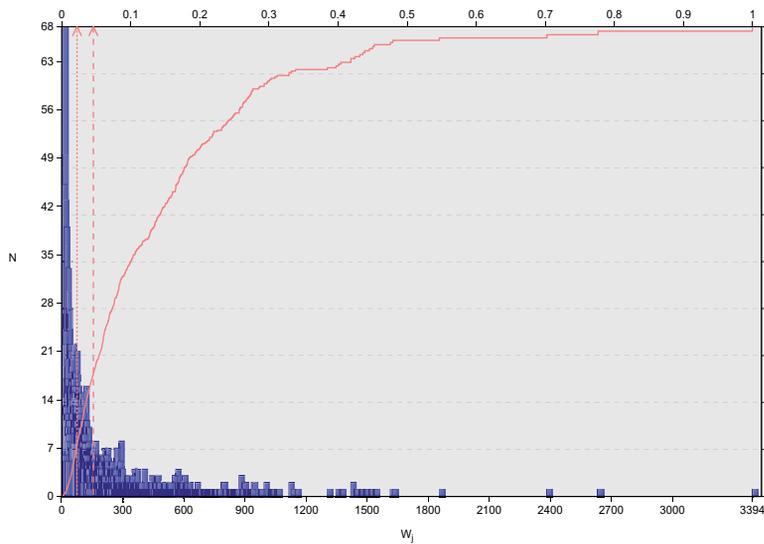


Abb. 5: Verteilung der Verwendungszahlen Z von Modulen im CMS; Balkenhöhe gibt die Zahl N von Modulen zu einer Wiederverwendungszahl an: Häufung von Modulen, die weniger als 30 Mal wiederverwendet werden; weitere Verteilung bis zu 272 Wiederverwendungen.

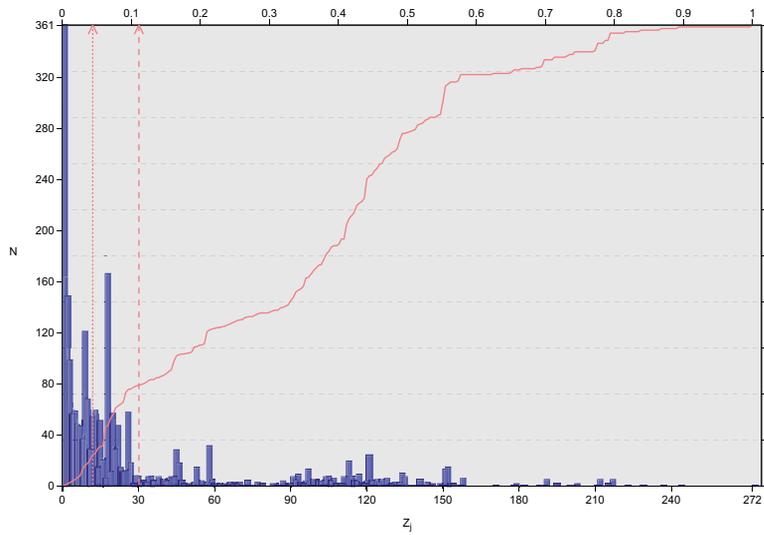
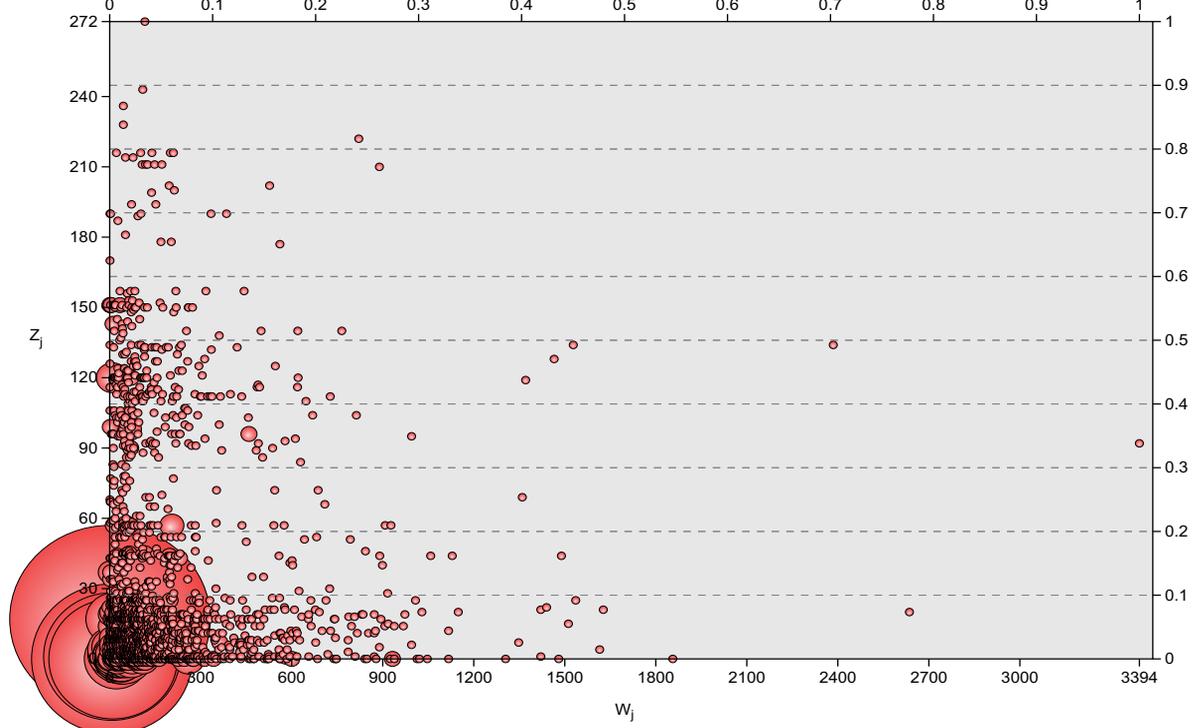


Abb. 6: Korrelation von Modulgröße W (horizontal) und Verwendungszahl N von Modulen in Dokumenten (vertikal); je größer die Kreise, desto mehr Module haben die gleiche Größe und Verwendungszahl.



Das betrachtete System besitzt insgesamt 66.771 modulare Wiederverwendungen (11 Mio. wiederverwendete Worte) im Vergleich zu einer relativ kleinen Zahl von 1914 verwendeten Ausgangsmodulen (insgesamt 2.275 Module und 353.000 Worte im System). Hierdurch wird schnell der Effizienzgewinn eines Systemeinsatzes klar: Alle kontrolliert wiederverwendeten Inhalte wären ohne CMS-Einsatz durch „unkontrolliertes copy&paste“ entstanden, inklusive einer fehlenden Übersetzungskontrolle. Die Ausgangsmodule im Beispiel werden dagegen im Mittel in 30 Dokumenten genutzt und bestehen aus durchschnittlich 155 Worten.

Es lassen sich darüber hinaus eine Vielzahl weiterer Größen im Detail analysieren. So kann die Verteilung der Wiederverwendung aller vorkommenden Module (Abbildung 5) oder die Verteilung der Modulgrößen (Abbildung 4) untersucht werden. Ähnliche Betrachtungen lassen sich für verwendete Medien und verschiedene dokumentspezifische Größen anstellen.

In den Daten der ersten Kooperationsprojekte ließ sich quantitativ ein typisches Verhalten der Verteilung von Modulgrößen erkennen. Die aus früheren Umfragen ermittelte Tendenz [2] zu kleinen Modulen wurde auch in der Praxis der REx-Daten bisher bestätigt (Abbildung 4). Zudem ist die Zahl der Wiederverwendungen der einzelnen Module stark verteilt. Grob formuliert: Eine Vielzahl der Module werden wenig verwendet, wenige werden häufig verwendet (Abbildung 5). Korreliert man beide Größen, erhält man Darstellungen wie in Abbildung 6. In dieser findet man für das untersuchte System eine Vielzahl (große Kreise und Häufung in der linken unteren Ecke) von kleinen Modulen, die nur in wenigen Dokumenten verwendet werden. Hintergrund könnte das komplexe Variantenmanagement auf feinmodularer Ebene sein. Große Module werden tendenziell seltener genutzt. Am häufigsten werden bestimmte, ebenfalls relativ kleine Module verwendet. Letztere sind zum Beispiel standardisierte Sicherheitshinweise oder andere allgemein einsetzbare Abschnitte. Mit Hilfe der REx-Daten lassen sich mit der Kopplung und Rückverfolgung in das CMS auch einzelne Objekte weiter untersuchen.

Somit scheinen die bisherigen Daten quantitativ zum Beispiel die bekannte Aussage zu stützen, dass die Wiederverwendung von Modulen mit zunehmender Modulgröße abnimmt [1, S. 313]. Eine breitere Datenbasis sollte zukünftig weitere statistische Studien und Rückschlüsse auf die CMS-Nutzung ermöglichen. Die Untersuchungen zeigen allerdings auch die teilweise schwierige Auswertbarkeit beziehungsweise die nicht eindeutige Interpretierbarkeit der Daten durch verschiedene Faktoren:

- unterschiedliche Systemphilosophien und technische Besonderheiten der einzelnen CMS (Bedeutung von „Versionen“; eingeschränkte Exportmöglichkeiten für Objektinformationen wie die Objektgröße)
- die Individualanpassung der Systeme nach jeweiligen Kundenanforderungen

- individuelle Prozessgestaltung und Systemnutzung in den Unternehmen (Versionierung wird nicht oder nicht nachvollziehbar durchgeführt)
- die Nutzung des Variantenmanagements (erstellte Dokumentvarianten werden bei RExport teilweise nicht berücksichtigt.)

Für standardisierte und vergleichende Auswertungen (Benchmarks) müssen diese Faktoren zukünftig genau berücksichtigt und falls nötig spezifisch in die RExporte oder die REx-Auswertungen implementiert werden.

Ausblick

Zusammengefasst verfolgt die Anwendung der vorgestellten CMS-Metrik zwei Ziele:

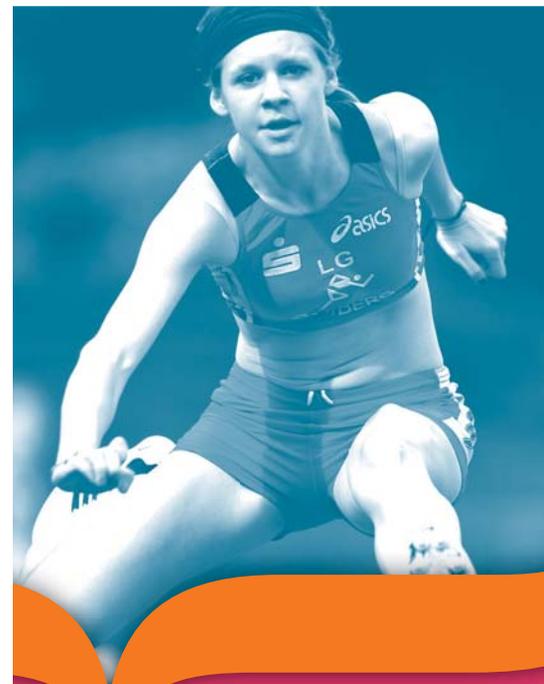
- Bereitstellung von Werkzeugen für das unternehmensinterne Prozessmanagement zur Optimierung der CMS-Nutzung
- Wissenschaftliches und methodisches Studium der CMS-Nutzung und Erarbeitung von branchenabhängigen Best-Practices und Benchmarks für CMS

Die verwendete Technologie der RExporte ist in den Systemen als Schnittstelle relativ einfach zu implementieren und lässt sich – neben einer Reihe von linguistischen Werkzeugen – als ein Weg zur Einführung von „Content Intelligence“-Verfahren in CMS verstehen. Die ersten Untersuchungen zeigen interessante Auswertungsmöglichkeiten mit Blick auf die Effizienz modularer Wiederverwendung und einer Vielzahl von weiteren Systemgrößen.

Anstehende Erweiterungen der Methodik sind die Auswertungen dynamischer, das heißt zeitlicher und projektbezogener Daten. Darüber hinaus sollen alle Objektinformationen berücksichtigt werden, um zum Beispiel die Nutzung von Klassifikationssystemen (Metadaten) oder der Semantik (Informationsstrukturen) zu ermitteln. Technisch könnten sich die REx-Methode und die metrische Analyse zukünftig als Applikation etablieren, die als integrierte beziehungsweise additive Komponente für CMS oder als externer Service realisiert werden kann.



Mit der richtigen Technik *schneller* zum Ziel



Mit den Sirius-Redaktionssystemen bringen wir Technologie und Anwender zusammen.

Konsequent medienneutral – flexibel – sicher.

Mit Sirius CMS erstellen und mehrsprachig publizieren: Anleitungen, Produktinformationen, Kataloge und mehr.

www.acolada.de



Magazin für internationales Informations- management

Jetzt Media-Kit
anfordern:
vertrieb@tekomp.de



magazine for international information management
tcworld

www.tcworld.info

Langfristig könnten durch „Content Intelligence“-Methoden eine aktive(re) Systemunterstützung zur leichteren Wiederverwendung („Reuse“) von Modulen entstehen. Vorbild auf sprachlicher Ebene wären die Sprachkontrolle und das Authoring-Memory, die im Erstellungsprozess Inhalte kontrollieren und aktive Vorschläge machen. In diesem Sinne wären Entwicklungen zu einem „Controlled Reuse“ (Überprüfung der Qualität und Effizienz der Wiederverwendungen) und einem „Reuse Memory“ (Vorschlag für passende Wiederverwendungen) ein analoger Ansatz.

Bereits jetzt können interessierte Unternehmen mit der REx-Methode relevante Kenngrößen des CMS-Einsatzes ermitteln und den technischen Effizienzgrad nach der Systemeinführung quantifizieren. Damit ist eine Basis gelegt für eine breitere Untersuchung und Erforschung von CMS-Implementierungen.

Literatur zum Beitrag

- [1] Drewer, Petra; Ziegler, Wolfgang (2011): Technische Dokumentation – Eine Einführung in die übersetzungsgerechte Texterstellung und in das Content-Management. Vogel Verlag.
- [2] Straub, Daniela; Ziegler, Wolfgang (2008): Effizientes Informationsmanagement durch spezielle Content-Management-Systeme. 2. Auflage, tekomp, Stuttgart.

- [3] Straub, Daniela; Grau, Michael; Fritz, Michael (2008): 101 Kennzahlen für die Technische Kommunikation. tekomp, Stuttgart.
- [4] Deming, W. Edwards (2000): Out of the Crisis, MIT Press.
- [5] Ziegler, Wolfgang (2008): Metrische Untersuchung der Wiederverwendung im Content Management, Hochschule Karlsruhe, www.home.hs-karlsruhe.de/~ziwo0001/PUB/CMS-Metrik_Ziegler.pdf
- [6] Knopf, Dominik (2009): Report Exchange Format 'Rex' – Metrische Untersuchung der Wiederverwendung im Content Management. Master-These Hochschule Karlsruhe.
- [7] Oberle, Claudia (2010): Kennzahlen für das Content Management – Auswertung und Visualisierung von Daten im Report Exchange Format (REX-Format). Master-These Hochschule Karlsruhe.
- [8] www.home.hs-karlsruhe.de/~ziwo0001/rex/intro.html

Autorenanschrift

Claudia Oberle
Hochschule Karlsruhe –
Technik und Wirtschaft
claudia.oberle@hs-karlsruhe.de

Prof. Dr. Wolfgang Ziegler
Hochschule Karlsruhe –
Technik und Wirtschaft
wolfgang.ziegler@hs-karlsruhe.de
www.hs-karlsruhe.de