



JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
PROFESSUR BWL – WIRTSCHAFTSINFORMATIK
UNIV.-PROF. DR. AXEL C. SCHWICKERT

Volckmann, Jörn; Müller, Andreas; Schwickert, Axel C.

**Anforderungsanalyse zur Mehrsprachigkeit
eines Web-Content-Management-Systems**

ARBEITSPAPIERE WIRTSCHAFTSINFORMATIK

Nr. 4 / 2005
ISSN 1613-6667

Arbeitspapiere WI Nr. 4 / 2005

- Autoren:** Volckmann, Jörn; Müller, Andreas; Schwickert, Axel C.
- Titel:** Anforderungsanalyse zur Mehrsprachigkeit eines Web-Content-Management-Systems
- Zitation:** Volckmann, Jörn; Müller, Andreas; Schwickert, Axel C.: Anforderungsanalyse zur Mehrsprachigkeit eines Web-Content-Management-Systems, in: Arbeitspapiere WI, Nr. 4/2005, Hrsg.: Professur BWL – Wirtschaftsinformatik, Justus-Liebig-Universität Gießen 2005, 45 Seiten, ISSN 1613-6667.
- Kurzfassung:** „Think global – act local!“ Ein bekannter Spruch, der im World Wide Web seine Gültigkeit nicht verloren hat. Im Zuge der zunehmenden Globalisierung wächst die Notwendigkeit für einen internationalen mehrsprachigen Web-Auftritt, der auf die jeweilige Zielgruppe lokalisiert zugeschnitten wird. Für den Anbieter einer globalen Web Site stellen sich verschiedene Probleme und Aufgaben. Eine globale Web Site zu erstellen heißt unter anderem, kulturelle Unterschiede zu erkennen und entsprechend in der E-Business-Strategie zu berücksichtigen. Ziel des Arbeitspapiers ist es, grundlegende Anforderungen der Mehrsprachenfähigkeit einer Web Site und daraus resultierend an ein WCMS abzuleiten. Im zweiten Kapitel werden die Implikationen der Globalisierung auf eine Web Site dargestellt, um daraus Anforderungen und Vorgehensweisen für die Gestaltung einer Web Site abzuleiten. Darauf aufbauend werden die grundlegende Struktur von WCMS und die Unterstützungsmöglichkeiten bei der Gestaltung einer mehrsprachigen Web Site durch WCMS dargestellt. Im dritten Kapitel werden die grundlegenden Anforderungen an ein mehrsprachiges WCMS erarbeitet. Dazu werden die aufgabenspezifischen Anforderungen an eine mehrsprachige Web Site und daraus abgeleitet an ein WCMS beschrieben. Abschließend werden die technikspezifischen Anforderungen näher untersucht.
- Schlüsselwörter:** Mehrsprachigkeit, Multilingualität, Globalisierung, Internationalisierung, Lokalisierung, Web Site, Web Content Management, WCMS, Web-Content-Management System, ISO, ASCII, Unicode, UTF

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis	3
1 Problemstellung, Ziel und Aufbau	4
2 Globalisierung im Internet	6
2.1 Globalisierung und Web Sites	6
2.2 Globalisierung und Web-Content-Management-Systeme	12
3 Anforderungen an ein mehrsprachiges WCMS	16
3.1 Arten von Anforderungen	16
3.2 Aufgabenspezifische Anforderungen.....	17
3.2.1 Sprachliche Variablen und Konstanten	17
3.2.2 Anforderungen an eine sprachneutrale Bedienungsfläche	24
3.2.2.1 Zeichenexpansion.....	24
3.2.2.2 Platzhalter.....	26
3.2.2.3 Benutzerspezifische Spracheinstellung.....	27
3.2.3 Speicher-Struktur der Web Site.....	29
3.2.4 Anforderungen an die Dokumentation des WCMS	30
3.2.5 Rechtliche Aspekte.....	31
3.3 Technikspezifische Anforderungen	32
3.3.1 Sprachen- und Länder-Codierung	32
3.3.2 Zeichensatzunterstützung.....	35
4 Zusammenfassung	39
Anhang	41
Literaturverzeichnis	43

Abkürzungsverzeichnis

ASCII.....American Standard Code for Information Interchange

E-BusinessElectronic Business

E-Commerce.....Electronic Commerce

ERP.....Enterprise Resource Planning

HTML.....Hypertext Markup Language

ITInformationstechnologie

LAMP.....Linux, Apache, MySQL, PHP

PCPersonal Computer

PDF.....Portable Document Format

RFC.....Internet Requests for Comments

URLUniform Resource Locator

US.....United States

W3C.....World-Wide-Web-Konsortium

WCMWeb Content Management

WCMSWeb-Content-ManagementSystem

WWW.....World Wide Web

WYSIWYGWhat you see is what you get

XMLeXtensible Markup Language

1 Problemstellung, Ziel und Aufbau

Die rasante Entwicklung der Informations- und Kommunikationstechnologien hat eine technologische Basis geschaffen, die Länder und Kontinente näher zusammenrücken läßt. Das Internet ist das Kommunikationsmedium der Zukunft und revolutioniert zunehmend die Geschäftswelt. Durch die zunehmende Elektronisierung verändern sich die Geschäftsprozesse und das E-Business ermöglicht neue länderübergreifende Unternehmensformen.¹

Die steigenden weltweiten Nutzerzahlen des Internets zeigen, über welches Kundenpotenzial das Internet heute schon verfügt und in Zukunft noch verfügen wird. Nach einer Untersuchung von Global Reach waren im März 2004 bereits 729 Millionen Benutzer weltweit Online.² Die Prognose von Global Reach über die Anzahl der Internetbenutzer liegt im Jahre 2005 bei etwa einer Milliarde Menschen weltweit. Mit der stetig steigenden Anzahl der Internetbenutzer nimmt auch die Zahl der verwendeten Sprachen im World Wide Web (WWW) zu. Global Reach stellt in der aktuellen Untersuchung über die Sprachenverwendung im Internet fest, daß für rund ein Drittel der Internetnutzer Englisch die Muttersprache ist. Weitere bedeutende Sprachen sind Chinesisch, Japanisch, Spanisch und Deutsch (siehe Abbildung 1).³

Die aktuelle Untersuchung unterstreicht die Tatsache, daß das Internet sich längst zu einem Massenmedium mit vielen unterschiedlichen Sprachen entwickelt hat. Eine reine englischsprachige Web Site reicht nicht mehr aus, um eine breite Masse im Internet effektiv anzusprechen.

Die Anforderungen an eine komplexe mehrsprachige Web Site können nicht mehr durch eine manuelle Web-Site-Entwicklung bewältigt werden. Ein systematisches Content-Management ist in der Web-Site-Entwicklung längst eine zwingende Voraussetzung geworden. In der Entwicklung und bei dem Betrieb einer Web Site kommen deshalb immer häufiger Web-Content-Management-Systeme (WCMS) zum Einsatz.

1 Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, 2. überarb. und erw. Auflage, Bonn: Galileo Press GmbH 2002, S. 176 f.

2 Vgl. o. V.: Global Internet Statistics (by Language), Online im Internet: <http://www.global-reach.biz/globstats/index.php3>, 03.09.2004.

3 Vgl. o. V.: Global Internet Statistics (by Language), a. a. O.

Ein modernes WCMS muß die Fähigkeit besitzen, mehrsprachige Inhalte pflegen und verwalten zu können. Für Unternehmen und Institutionen ist die Unterstützung der Mehrsprachigkeit ein wichtiges Auswahlkriterium bei der Entscheidung für ein WCMS.⁴

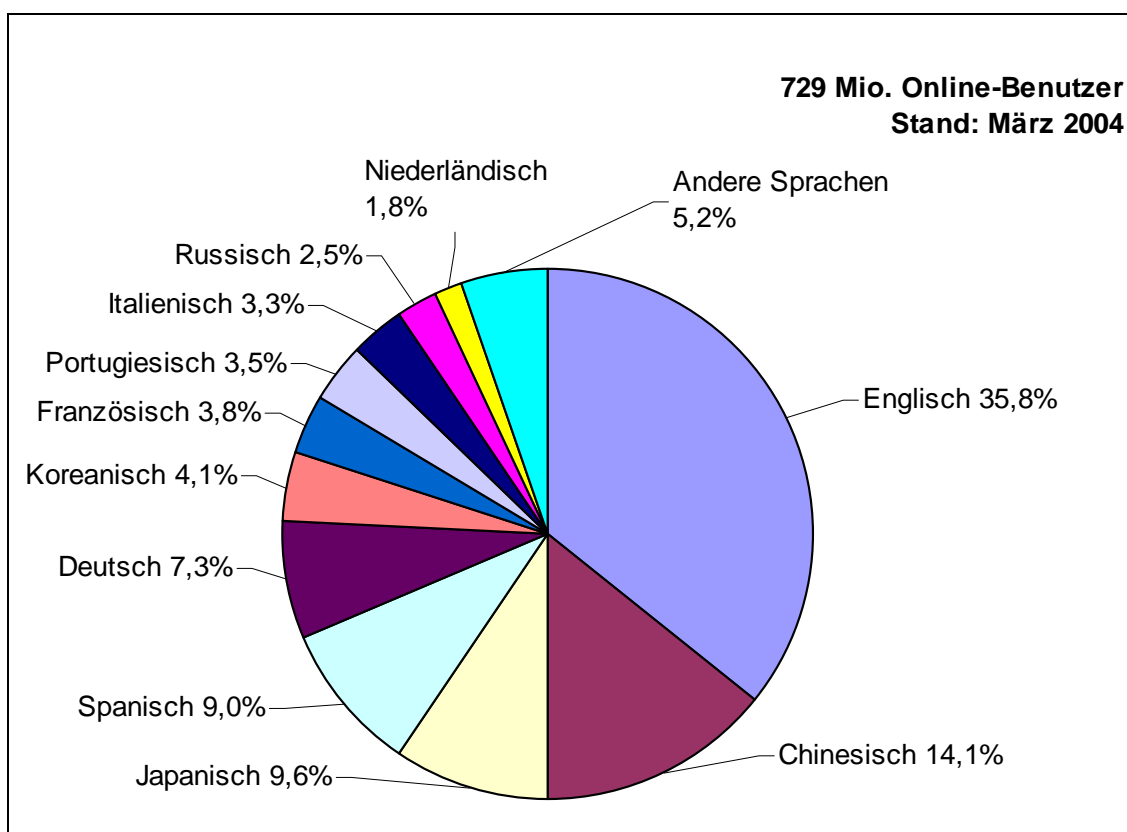


Abb. 1: Anteile der Internetbenutzer nach Muttersprache⁵

Ziel dieser Arbeit ist es, grundlegende Anforderungen der Mehrsprachenfähigkeit einer Web Site und daraus resultierend an ein WCMS abzuleiten. Im zweiten Kapitel werden die Implikationen der Globalisierung auf eine Web Site dargestellt, um daraus Anforderungen und Vorgehensweisen für die Gestaltung einer Web Site (Kapitel 2.1) abzulei-

4 Vgl. Scholz: CMS-Grundlagen: Content-Management – Einführung 1 – Auswahlkriterien für ein WCMS im Detail, Online im Internet: <http://www.content-syndication-broker.de/crm-wcms-artikel-seite31-folge6.html>, 07.09.2004.

5 o. V.: Global Internet Statistics (by Language), Online im Internet: <http://www.global-reach.biz/globstats/index.php3>, 03.09.2004.

ten. In Kapitel 2.2 werden die grundlegende Struktur von WCMS und die Unterstützungsmöglichkeiten bei der Gestaltung einer mehrsprachigen Web Site durch WCMS dargestellt. Im dritten Kapitel werden die grundlegenden Anforderungen an ein mehrsprachiges WCMS erarbeitet. Nach einer kurzen Einführung werden in Kapitel 3.2 die aufgabenspezifischen Anforderungen an eine mehrsprachige Web Site und daraus abgeleitet an ein WCMS beschrieben. In Kapitel 3.3 werden die technikspezifischen Anforderungen näher untersucht. Mit einer Bewertung der Anforderungen sowie einer kritischen Würdigung bildet das vierte Kapitel den Abschluß.

2 Globalisierung im Internet

2.1 Globalisierung und Web Sites

Das Internet hat die traditionellen Geschäftspraktiken verändert und einen direkten weltweiten (globalen) Zugang zu Informationen und Produkten ermöglicht. Globalisierung kann als „das weltweite Zusammenwirken und eine immer stärker werdende Verknüpfung der nationalen Märkte und Gesellschaften aufgrund von Entwicklungen im Bereich des Informations- und Warentransportes“⁶ verstanden werden. Mit dem Begriff Globalisierung wird „der Prozeß beschrieben, der eine Unternehmensorganisation so gestaltet, daß eine internationale Kundschaft effektiv angesprochen und mit Dienstleistungen oder Waren versorgt werden kann.“⁷ Ein global agierendes Unternehmen muß folgende Aspekte berücksichtigen:⁸

- Allgemeine Geschäftsprozesse und Technologien müssen global genutzt und gelebt werden.
- Ansprüche an Qualität und Verfügbarkeit der verbreiteten Waren und Dienstleistungen müssen global gesichert werden.

6 o. V.: Globalisierung, Online im Internet: <http://de.wikipedia.org/wiki/Globalisierung>, 07.09.2004.

7 Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 176.

8 Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 176.

- Eine Lokalisierung von globalen Angeboten muß genutzt werden, um auf regionale Marktunterschiede einzugehen.
- Regionale und internationale Geschäftsbereiche müssen gemeinsam ihre Anforderungen an Marketing und Technologien definieren.
- Regionale Niederlassungen müssen mit den entsprechenden Befugnissen ausgestattet sein, um die globalen Richtlinien lokal umzusetzen.

Viele global agierende Unternehmen folgen der Maxime „Think global, act local“.⁹ Die Unternehmensleitung agiert globalorientiert, während die dezentralen Organisationseinheiten mit entsprechendem Know-how und Marktkenntnissen den lokalen Markt bearbeiten. Daß der Trend zur Dezentralisierung von Unternehmensstrukturen geht, zeigt die wachsende Anzahl an ausgelagerten Vertriebsstrukturen, zeitlich begrenzten virtuellen Zusammenschlüssen, Allianzen und Unternehmensgemeinschaften. Die Web Site des Unternehmens muß diese dezentralen Aspekte in der Konzeption berücksichtigen. In der Literatur hat sich diesbezüglich der Begriff „Globalisierung einer Web Site“ oder „Web Globalisierung“ durchgesetzt. Unter diesem Begriff werden alle Aktivitäten verstanden, die erforderlich sind, um eine Web Site für einen bestimmten Beschaffungs- oder Absatzmarkt vorzubereiten bzw. anzupassen.¹⁰ Bei der Umsetzung dieser Anforderungen werden die beiden zentralen Prozesse Internationalisierung und Lokalisierung nacheinander durchlaufen. Die Bedeutung der Prozesse der Internationalisierung und Lokalisierung sollen deshalb im folgenden näher betrachtet werden.

Internationalisierung (engl.: internationalization) ist ein Schritt im Entwicklungsprozeß einer Web Site. Ziel der Internationalisierung ist es, die Web Site so zu gestalten, daß verschiedene Sprachen und kulturelle Einflüsse ohne zusätzlichen Programmieraufwand eingebaut werden können. Eine zentrale Anforderung der Internationalisierung ist die

9 Vgl. Metzen, Heinz: Das globale Informationszeitalter entsteht auf dem Marktplatz, nicht in Zukunftsministerien oder auf Weltgipfeln, Online im Internet: http://www.sommeruni-uni-osna-brueck.de/04_1.htm, 28.09.2004.

10 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, 2. unveränderte Auflage, New Riders Publishing 2003, S. 475.

Trennung der Sprache und anderer kulturspezifischer Inhalte vom Quellcode der Web Site.¹¹

Folgende Aspekte müssen beim Prozeß der Internationalisierung von Web Sites berücksichtigt werden:¹²

- Isolierung aller sprachlichen und kulturspezifischen Elemente einer Web Site, die angepaßt werden müssen;
- das Design der Oberfläche muß neutral und flexibel gehalten werden;
- Berücksichtigung von regionalen Standards;
- Grafiken müssen anpaßbar sein.

Bevor die Arbeiten an einer Länder- und Sprachenversion, die Lokalisierung, beginnen können, benötigt man eine neutrale, internationalisierte Fassung der Web Site.¹³ Diese Internationalisierung einer Web Site wird einmalig durchgeführt. Je sorgfältiger die Internationalisierung bei der Entwicklung der Web Site durchgeführt wird, desto weniger Aufwand entsteht bei der späteren Lokalisierung.¹⁴ Bei einer bestehenden Web Site ist eine nachträgliche Internationalisierung mit einem erheblichen Zeit- und Kostenaufwand verbunden.¹⁵

Lokalisierung (engl. localization) ist der allgemeine Begriff für die Anpassung von Inhalten (z. B. Bücher, Web Sites), Prozessen, Produkten und insbesondere Computerprogrammen (Software-Lokalisierung) an die in einem bestimmten geografisch und/oder ethnisch umschriebenen Absatz- oder Nutzungsgebiet vorherrschenden "lokalen" sprachlichen und kulturellen Gegebenheiten. Das Absatz- oder Nutzungsgebiet kann ein Land, eine Region oder eine ethnische Gruppe sein.¹⁶ Schwerpunkt der Lokalisierung ist

11 Vgl. o. V.: Globalization, internationalization, localization: an overview, Online im Internet: <http://www.globalization.com/index.cfm?MyCatID=1&MySubcatID=1&pageID=1322>, 20.08.2004.

12 Vgl. o. V.: Globalization, internationalization, localization: an overview, a. a. O.

13 Vgl. Wiegand, Dorothee: Multikulti-Software – Wie Programme fremde Sprachen lernen, in: C't, Heft 7, 2004, S. 186.

14 Vgl. Wiegand, Dorothee: Multikulti-Software – Wie Programme fremde Sprachen lernen, a. a. O., S. 186.

15 Vgl. o. V.: Globalization, internationalization, localization: an overview, a. a. O.

16 Vgl. o. V.: Lokalisation, Online im Internet: <http://de.wikipedia.org/wiki/Globalisierung>, 07.09.2004.

in erster Linie die Übersetzung der Inhalte in andere Sprachen. Daneben zählen zur Anpassung auch Datums-, Zeit-, Währungs- und Temperaturangaben sowie Maßeinheiten und Umrechnungsgrößen. Aber auch Farbpaletten, Schriftarten und Zeichensätze, Audio- und Sprachausgaben sowie Bilder (z. B. Flaggen) müssen angepaßt werden. Gerade bei der Verwendung von Grafiken ist oft eine Anpassung an die sprachlichen und kulturellen Gegebenheiten und die besonderen Gepflogenheiten bzw. Vorlieben der jeweiligen Nutzer wichtig. In jedem Fall sind auch geltende Rechtsvorschriften¹⁷ und ethnische Grundsätze zu beachten.

Die Kosten der Lokalisierung einer Web Site sind nicht zu unterschätzen. Die nachfolgende Abbildung 2 zeigt das Kostenverhältnis in der Entwicklung einer mehrsprachigen Web Site. Jede Lokalisierungsversion einer Web Site verursacht neben den einmaligen Kosten für Entwicklung, Design und Projektmanagement, in der Folge Kosten für die laufende Pflege der Inhalte. Die Übersetzungen können einen beträchtlichen Anteil der laufenden Kosten (35%) verursachen.

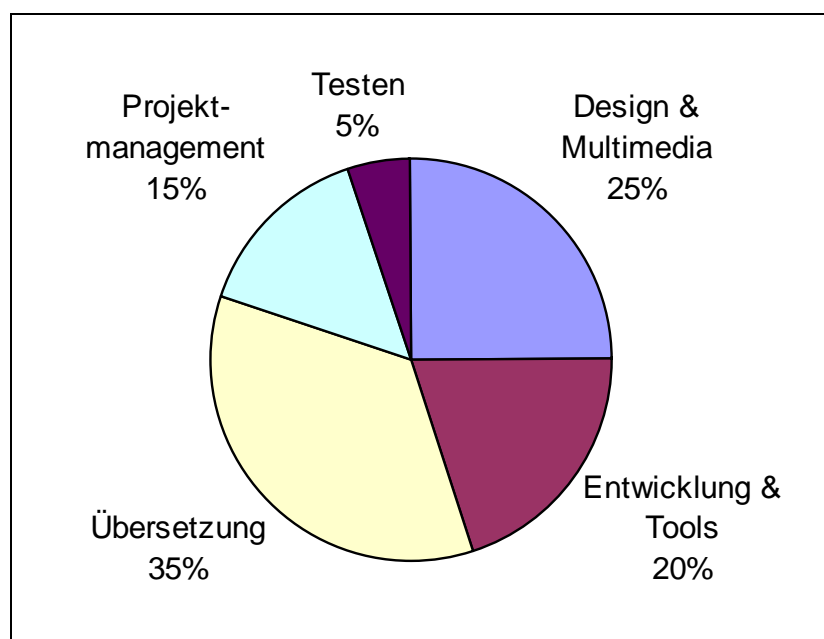


Abb. 2: Kostenzuordnung im Web-Site-Projekt¹⁸

17 Vgl. o. V.: Lokalisation, Online im Internet: <http://de.wikipedia.org/wiki/Globalisierung>, 07.09.2004.

18 Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 141.

Die sprachliche Übersetzung von Inhalten, eine Hauptaktivität in der Lokalisierung einer Web Site, wird überwiegend von menschlichen Arbeitskräften geleistet – nur sie können mit den sprachlichen und kulturellen Feinheiten umgehen. Eine vollautomatische Übersetzung durch Computer führt zur Zeit noch zu kaum brauchbaren Ergebnissen. Die Aufgabe der Computer liegt deshalb nach wie vor darin, die Übersetzer so zu unterstützen, daß Zeit und Geld gespart werden.¹⁹ Dies geschieht durch Computer-Aided Translation (CAT) und beinhaltet z. B. computer-basierte bilinguale und multilinguale Wörterbücher, Managementsoftware für Fachausdrücke oder Übersetzungssoftware, die mit gespeicherten Wörter oder Satzteilen Übersetzungen teilweise automatisieren können.²⁰ Überträgt man die Übersetzungen einem Übersetzungsbüro, z. B. Berlitz, belaufen sich bei westeuropäischen Sprachen die Kosten auf ca. 18 bis 25 US-Cents pro Wort. Bei asiatischen Sprachen und Sprachen des Nahen Ostens können die Kosten auf bis zu 38 US-Cents pro Wort steigen. Hat eine Web Site 100 Seiten mit durchschnittlich 200 Wörtern je Seite, belaufen sich die Übersetzungskosten z. B. vom Englischen ins Deutsche auf immerhin \$ 4600,- (\$ 0,23 je Wort).²¹ Auch die Kosten für Design und Entwicklung sind nicht zu unterschätzen, da man ein gewisses Know-how in der Web-Entwicklung benötigt, über das nur Spezialisten des Grafikdesign und der Programmierung verfügen.²²

Das Zusammenwirken von Internationalisierung und Lokalisierung wird am Beispiel der E-Business-Präsenz von Lands'End in der Abbildung 3 deutlich. Das Unternehmen setzt ein globales Masterdesign ein, das für alle Länder gleich, aber dennoch flexibel genug ist, um eine lokale Anpassung an die Kundenbedürfnisse zu ermöglichen. Im direkten Vergleich der beiden Web Sites von Lands'End Großbritannien und Lands'End Deutschland wird deutlich, daß die Navigationselemente nahezu identisch sind und aus internationalisierten Komponenten bestehen. Lediglich die Navigationsbezeichnungen werden an die lokale Sprache angepaßt. Ebenfalls ist die Werbung für das Abonnieren

19 Vgl. Yunker, John: Building a Global Web Site – The Guide from MultiLingual Computing & Technology, in: Global Web, 55 Supplement, April/May 2003, Hrsg: MultiLingual Computing Inc., Sandpoint, 2003, S. 4.

20 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 220.

21 Vgl. Wiegand, Dorothee: Multikulti-Software – Wie Programme fremde Sprachen lernen, a. a. O., S. 187.

22 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 140.

des Newsletters länderspezifisch bzw. der jeweiligen Mentalität der Briten und der Deutschen angepaßt worden. Während die Funktion des Abonnierens des Newsletters auf der britischen Web Site mit einer Chance auf einen Gewinn in Verbindung gebracht wird, wird auf der deutschen Web Site das kostenlose Abonnement des Newsletters betont. Auf der deutschen Web Site wird zusätzlich ein kostenloser Lands'End-Internetzugang angeboten, während dieser Inhalt auf der britischen Web Site fehlt. Der Vergleich der beiden Web Sites von Lands'End zeigt, daß trotz unterschiedlicher Sprache und Mentalitäten sich die Web Sites in Bezug auf das Design nur in Details unterscheiden. Unternehmen tendieren zu globalen Templates, die dann als Vorlage für nahezu alle Länder- und Sprachversionen innerhalb der E-Business-Präsenz dienen.²³



Abb. 3: Internationalisierung und Lokalisierung bei Lands'End²⁴

23 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 21 f.

24 Screenshots von <http://www.landsend.co.uk> und <http://www.landsend.de>, 25.09.2004.

2.2 Globalisierung und Web-Content-Management-Systeme

Historisch gesehen sind die meisten Web-Auftritte als statische, in einer Sprache realisierte Web Sites entstanden. Vielfach werden diese Web-Auftritte noch mit Web Editoren wie Front Page von Microsoft oder Dreamweaver von Macromedia gepflegt, was zwar eine Erleichterung gegenüber der Pflege in der reinen HTML-Codierung mit einem Text-Editor darstellt, aber keinerlei Möglichkeiten bietet, die Erstellungs- und Publikationsprozesse zu automatisieren. Die konzeptionellen und technischen Schwachstellen bei konventionellen Web Sites führen zu einem hohen Aufwand bei der Erweiterung des Auftritts durch eine neue Sprachvariante. In der Regel entstehen dabei durch Duplizierung, Übersetzung der Texte und Bereitstellung grafikbasierter Navigationselemente für jede Sprachversion parallele Web Sites und redundante Objekte. Dadurch vermehren sich die verwalteten Inhalte mit jeder neuen Sprachversion beträchtlich. Hinzu kommt, daß durch die gestiegene Komplexität aufgrund dieser Spracherweiterung die Pflfegbarkeit leidet und damit zu rechnen ist, daß Inkonsistenzen und die Fehlerrate auf der Web Site steigen. In vielen Fällen liegen die Redaktionen dezentral organisiert bei den nationalen Dependancen, so daß zunehmend Layout und Corporate-Design verloren gehen.²⁵ Eine mehrsprachige Web Site stellt andere Anforderungen an das Layout und Design, an die Funktionalitäten und an die Workflow-Komponenten als eine einsprachige Web Site. Ohne ein geeignetes System, in dem mehrsprachiger Inhalt geplant, erstellt und gepflegt wird, können mehrsprachige E-Business-Präsenzen nicht effizient verwaltet werden. Ein Web-Content-Management-System kann für global agierende Unternehmen bei der sprachlichen Regionalisierung der Inhalte Unterstützung bieten.

Für den Begriff „Web Content Management“ (WCM) existiert bisher keine einheitliche Definition. Selbst die Hersteller professioneller Web-Content-Management-Systeme geben auf die Frage, was sie unter Web Content Management verstehen, sehr unterschiedliche Antworten. Der Begriff „Web Content Management“ soll deshalb zunächst anhand der einzelnen Namensbestandteile erklärt werden.

²⁵ Vgl. o. V.: Internationalisierung – wenn der Webauftritt mehrsprachig werden soll wird es kritisch, a. O., S. 2.

Unter „Web“ werden hier alle drei möglichen Anwendungsgebiete der Web-Technologie, nämlich Internet, Intranet und Extranet verstanden.²⁶ Während die Web-Seiten des Internet für die Öffentlichkeit gedacht und meist uneingeschränkt zugänglich sind, hat das Intranet eine abgegrenzte, unternehmensinterne Zielgruppe. Das Extranet dient der Kommunikation mit definierten Außenstellen, Partnern und Zulieferern und ist genau wie das Intranet gegen unbefugte Zugriffe von außen geschützt.

Als „Content“ wird der gesamte Inhalt einer Web Site bezeichnet. Der Content besteht aus einzelnen digitalen Assets, wie z. B. Bildern, Texten, Grafiken oder auch Video- und Audio-Komponenten, die in einem bestimmten Dateiformat vorliegen. Ein modernes Web-Content-Management-System versteht Content als Summe von Einzelinformationen. Ein WCMS nimmt dabei eine Trennung von Inhalt, Struktur und Darstellung vor.²⁷ Die Struktur definiert die Einzelinformationen eines Dokuments und beschreibt deren Reihenfolge und Verschachtelung. Es erfolgt somit eine Beschreibung der einzelnen Datenelemente, die durch den jeweiligen Inhalt ausgefüllt werden. Die Darstellung ist eine formale Beschreibung, mit deren Hilfe die Ausgabe auf der jeweiligen Web Site realisiert wird.²⁸

Der Namensbestandteil „Management“ weist auf den Prozeß-Charakter des WCM hin. Verschiedene Anwendungsprozesse wie die Sammlung, Erstellung und Verwaltung von Informationen sowie die anschließende Veröffentlichung im Web müssen beim WCM geplant, gesteuert und kontrolliert werden. Dafür werden die Informationen mit Hilfe eines WCMS in ein Workflow-Schema überführt, auf dem der gesamte Publikationsprozess aufbaut.²⁹ Unter Management wird hier die gesamte Verwaltung und Abbildung dieses Workflows in Bezug auf alle inhaltsbezogenen Operationen bezeichnet.³⁰

26 Vgl. Jablonski, Stefan; Meiler, Christian: Web-Content-Managementsysteme, in: Informatik Spektrum 2/2002, S. 102.

27 Vgl. Bullinger Hans-Jörg (Hrsg.): Schuster, Erwin; Wilhelm, Stephan: Content Management Systeme – Auswahlstrategien, Architekturen und Produkte, 5. Aufl., Düsseldorf: Verlagsgruppe Handelsblatt 2001, S. 6.

28 Vgl. Jablonski, Stefan; Meiler, Christian: Web-Content-Managementsysteme, a. a. O., S. 107.

29 Vgl. Jablonski, Stefan; Meiler, Christian: Web-Content-Managementsysteme, a. a. O., S. 102.

30 Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 71.

Der Prozeß des WCM kann als systematische, rechnergestützte Erstellung, Aufbereitung und Aktualisierung, Speicherung, Archivierung und Wiederverwendung sowie Publikation von Web-Inhalten verstanden werden.³¹ Ein Softwaresystem zum WCM wird als *System* bezeichnet, da es sich im wesentlichen mit der Verwaltung von Web Content befaßt³² oder auch als „Softwaresystem für die Administration von Web-Inhalten mit Unterstützung des Erstellungsprozesses, basierend auf der Trennung von Inhalten und Struktur“ verstanden wird.³³ Jablonski/Meiler definieren ein Web-Content-Management-System (WCMS) als ein „Werkzeug zur Beschaffung, Erzeugung, Verwaltung von Inhalten im World Wide Web“.³⁴

Bei der Auswahl oder Entwicklung eines WCMS zur Unterstützung des Globalisierungsprojektes einer Web Site müssen einige konzeptionelle Aspekte betrachtet werden, um im nächsten Schritt die Anforderungen an ein solches WCMS zu spezifizieren. Der gesamte Prozeß eines Globalisierungsprojektes einer Web Site mit WCMS-Unterstützung kann zusammenfassend in die Phasen Internationalisierung, Lokalisierung und Content Management unterteilt werden (Abbildung 4). Der wichtigste Aspekt jedes Globalisierungsprojektes einer Web Site ist die Erstellung und Pflege verschiedener Sprachversionen. Bei der erstmaligen Einrichtung einer Sprachversion werden alle drei Phasen des Gesamtprozesses eines Globalisierungsprojektes der Web Site durchlaufen. Die Aufnahme einer weiteren Sprachversion erfordert nicht mehr den gesamten Prozeßdurchlauf, sondern beschränkt sich auf die Phasen Lokalisierung und Content Management.³⁵

Bei der Unterstützung des Projektes durch ein mehrsprachenfähiges WCMS stehen die Prozesse der Internationalisierung sowie des Content-Management im Vordergrund. Die Prozesse Lokalisierung und Übersetzung stehen hauptsächlich im Zusammenhang mit der Anpassung der Inhalte an die verschiedenen Zielgruppen und sind deshalb zu Be-

31 Vgl. Krüger, Dennis Jörg; Kopp Matthias: Web Content managen – Professioneller Einsatz von Content-Management-Systemen, München: Markt + Technik Verlag 2002, S.21.

32 Vgl. Porst, Axel: Web Content Management und Workgroup Computing, Online im Internet: <http://www.e-gateway.de/eco/contentman.cfm>, 07.07.2004.

33 o. V.: Glossar: Content Management System, Online im Internet: http://www.content-manager.de/ressourcen/glossar_10_content_management_system.html, 09.07.2004.

34 Vgl. Jablonski, Stefan; Meiler, Christian: Web-Content-Managementsysteme, a. a. O., S. 101.

35 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 141.

ginn der Pflege einer neuen Sprachversion von Bedeutung. Im späteren Produktivsystem, also nach der Initiierung der Sprachversionen, werden die Lokalisierung sowie der Hilfsprozeß der Übersetzung in die Workflows des Content-Management überführt. Trotzdem stellen alle vier Prozesse bereits vor der Konzeptionsphase Anforderungen, die das Produktivsystem später erfüllen muß.

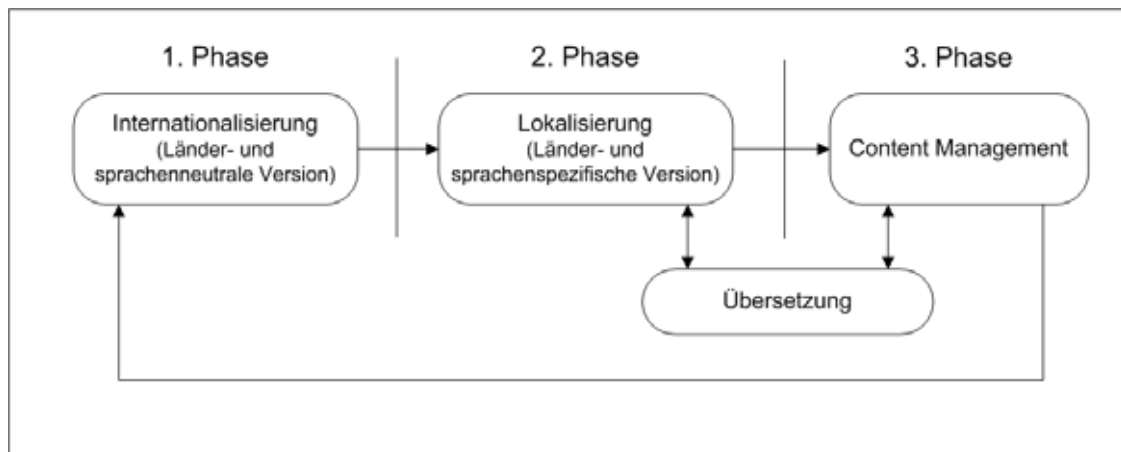


Abb. 4: Prozeß des Globalisierungsprojektes einer Web Site³⁶

Ein ideales mehrsprachenfähiges WCMS unterstützt alle gesprochenen Sprachen und Dialekte im Internet und könnte somit eine höchstmögliche Flexibilität des Systems garantieren. Unter Aufwands- und Kostenbetrachtungen ist die Realisierung eines solchen WCMS utopisch und daher ist es unumgänglich, bereits in der Planungsphase des Globalisierungsprojektes ein bestimmtes Sprachenset zu definieren, welches später vom System unterstützt werden soll. Sollte eine weitere Sprache, die im Sprachenset nicht definiert wurde, im späteren Produktivsystem integriert werden, ändern sich unter Umständen einige Anforderungen oder neue Anforderungen kommen hinzu, was wiederum dazu führen kann, daß Änderungen am System erforderlich werden. Es ist also von großer Bedeutung, vorab das gewünschte/erforderliche Sprachenset sorgfältig zu analysieren.

³⁶ In Anlehnung an Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 141.

Für die Auswahl oder Entwicklung eines geeigneten WCMS zur Unterstützung eines Globalisierungsprojektes einer Web Site existieren eine Reihe von Anforderungen, die im nachfolgenden Kapitel näher betrachtet werden.

3 Anforderungen an ein mehrsprachiges WCMS

3.1 Arten von Anforderungen

Wie aus der Betrachtung des Globalisierungsprozesses einer Web Site ersichtlich wurde, stehen die Prozesse der Internationalisierung und des Web Content Management bei der Auswahl oder Konzeption eines mehrsprachenfähigen WCMS im Vordergrund. Die Prozesse Lokalisierung und Übersetzung stehen hauptsächlich im Zusammenhang mit der Anpassung der Inhalte an die verschiedenen Zielgruppen und sind erst im laufenden System von Bedeutung. Die Hauptaufgabe der Internationalisierung ist die Isolierung aller sprachlichen Bestandteile einer Web Site bzw. eines WCMS. Daraus folgt, daß eine Hauptanforderung der Internationalisierung die Erstellung einer sprachenneutralen Fassung eines WCMS ist. Aus dieser Hauptanforderung wiederum lassen sich zum einen die aufgabenspezifischen Anforderungen an die Gestaltung einer sprachenneutralen Version eines WCMS und zum anderen die technikspezifischen Anforderungen, die Gestaltung der Hardware- und Software-Infrastruktur eines WCMS, bestimmen. Als aufgabenspezifische Anforderungen werden in der Literatur³⁷ hauptsächlich die Isolierung aller sprachlichen Elemente, die Bereitstellung einer sprachneutralen Bedienungsoberfläche, eine strukturierte Ordner-Architektur, die mehrsprachige Dokumentation und rechtliche Rahmenbedingungen genannt. Wichtige genannte technikspezifische Anforderungen sind die Länder- und Sprachencodierung im Quelltext, die Zeichensatzunterstützung im Quellcode und in Datenbanken sowie multilinguale Clients für den Zugriff auf das WCMS. In den nachfolgenden Kapiteln werden die einzelnen Anforderungen näher betrachtet.

37 Vgl. stellvertretend Yunker, John: *Beyond Borders: Web Globalization Strategies*, a. a. O., S. 23 ff. und S. 175 ff. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: *Web Content Management – Websites professionell planen und betreiben*, a. a. O., S. 180 ff. und o. V.: *Globalization, internationalization, localization: an overview*, a. a. O.

3.2 Aufgabenspezifische Anforderungen

3.2.1 Sprachliche Variablen und Konstanten

In einer sprachneutralen Version eines WCMS sind alle sprachlichen und kulturellen Elemente isoliert. Die Isolation der Sprachelemente umfaßt in erster Linie die Benutzeroberflächen und den Quellcode. Die Isolierung kann durch Auslagerung der entsprechenden Inhalte in Datenbanken oder Textdateien durchgeführt werden. Im eigentlichen Quellcode des WCMS sind nur noch Referenzierungen auf die jeweiligen Elemente enthalten. Dabei ist zu beachten, daß es sprachlich konstante und variable Inhalte gibt.

Ein sprachlich konstanter Inhalt ist ein Element einer Web Site, das keine Anpassung an eine Sprachversion benötigt oder so weit sprachlich bereinigt wurde, daß es keiner weiteren Anpassung mehr bedarf. Diese Elemente sind hauptsächlich sogenannten „Corporate Constants“, wie z. B. Markennamen, globale Werbeslogans, Logos, Farben, Navigationsmenüs. Lands’End (Abb. 2) beispielsweise verwendet als „Corporate Constants“ das Logo in der oberen linken Ecke, ein einheitliches Farbenset sowie die Navigationsmenüstruktur.

Sprachlich variable Inhalte sind Elemente einer Web Site, die sich je nach Länder- und Sprachenversion ändern können. Diese Elemente sind zu isolieren und das WCMS muß Strukturen bieten, so daß eine neue Sprachversion einfach integriert werden kann. In der späteren Lokalisierung müssen diese sprachlich variablen Elemente für den Benutzer in einer verständlichen Form aufbereitet werden.³⁸ Der Umfang der Internationalisierung wächst mit der Anzahl der sprachlich variablen Inhalte. Sprachlich variable Elemente sind u. a.:³⁹

- Maßeinheiten und Größen,
- Zahlenformate,
- Preise und Währungen,
- Datums- und Zeitangaben,
- Kontaktinformationen,
- Bilder und Icons.

³⁸ Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 182.

³⁹ Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 175 f.

Einige sprachlich variable Inhalte können in mehreren Sprachversionen benutzt werden, wobei vorab zu klären ist, ob alle Funktionalitäten in allen Sprachversionen angeboten werden sollen, oder ob es Funktionen gibt, die in einer Standardsprache angeboten werden.

1. Maßeinheiten und Größen

In weiten Teilen der Welt gilt das metrische Maßsystem, dennoch gibt es einige Länder, die andere Maßsysteme verwenden. So ist man z. B. in den USA den Umgang mit Meilen und Gallonen gewohnt. In Rußland gibt es die russische Tonne, die ebenfalls vom metrischen System abweicht. In vielen Ländern wird die Temperatur in Grad Celsius angegeben, die meisten englischsprachigen Länder und die meisten Länder des Commonwealth benutzen aber noch die Maßeinheit Fahrenheit. Auch das Papierformat ist ein Ergebnis von unterschiedlichen Standards der Maße in den Ländern. Während das A4-Papierformat in vielen Ländern ein gängiges Format ist und durch die International Standards Organization (ISO) in der ISO-Norm 216 festgelegt ist, benutzen z. B. die Amerikaner den U.S. Standard Letter.⁴⁰ Ein weiteres Beispiel sind die Kleidergrößen, die sogar innerhalb Europas sehr unterschiedlich definiert sind.

2. Zahlenformate

Eine weitere Schwierigkeit ergibt sich aus den unterschiedlichen numerischen Formaten. In einigen Ländern wird das Komma als Dezimalkomma verwendet, dagegen wird im englischsprachigen Raum der Dezimalpunkt verwendet und das Komma kennzeichnet die Tausender-Stelle. Die nachfolgende Tabelle 1 verdeutlicht die unterschiedliche Verwendung des Kommas und des Punktes.⁴¹

Die Notation bei der Angabe von Telefonnummern wird in vielen Ländern sehr unterschiedlich gehandhabt. Selbst in Deutschland ist die Telefonnummer in sehr unterschiedlichen Notationen angegeben. Eine allgemeingültige maximale Länge einer Tele-

40 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 185.

41 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 187 f.

fonnummer ist wegen der Heterogenität kaum feststellbar. Das Eingabefeld der Telefonnummer sollte daher ausreichend Platz bieten und ein Freitextfeld sein. Das entsprechende Feld in der Datenbank sollte nicht nur Zahlen unterstützen, sondern auch Sonderzeichen wie Klammern, Bindestrich und das Pluszeichen. In Tabelle 2 sind einige Beispiele aufgeführt.

Land	Notation
USA	123,456,789.00
Deutschland	123.456.789,00
Frankreich	123 456 789,00

Tab. 1: Zahlenformate im Ländervergleich⁴²

Land	Notation
Deutschland	(1234) 4567-89
Frankreich	12-34-45-67-89
Polen	(12) 345.67.89
Schweden	12 345 67 89
USA	(123) 456 7890
Vereinigtes Königreich	1234 56789

Tab. 2: Schreibweise der Telefonnummern im Ländervergleich⁴³

3. Preise und Währungen

Die Angaben von Preisen und Währungen werden in den einzelnen Ländern ebenfalls unterschiedlich gehandhabt. In einigen Ländern, wie z. B. Deutschland, wird erst der Preis und dann die Währungseinheit genannt, während in anderen Ländern, z. B. den USA, erst das Währungszeichen kommt und dann der Preis.⁴⁴ In einigen Ländern ist es

42 In Anlehnung an Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 187.

43 In Anlehnung an Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 188 f.

44 Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 184.

üblich, die bekannten Währungssymbole (€ \$ und andere) zu verwenden. Diese werden dann oft dem Preis vorangestellt.

Von der International Organization für Standardization (ISO) gibt es für Währungen genormte Abkürzungen aus drei Buchstaben (ISO 4217), die im internationalen Zahlungsverkehr zu eindeutigen Identifizierung benutzt werden sollen. (siehe Tabelle 3). In der Datenbank sollten der Preis und die Währungseinheit getrennt in zwei Feldern gespeichert werden. Die Speicherung der Währungseinheit sollte in der Notation nach ISO 4217 erfolgen. Die Darstellung der Preis- und Währungsangaben kann in der Lokalisierung sprach- und länderbezogen erfolgen.

ISO-Code	Währung	Land
DKK	Krone	Dänemark
GBP	Pfund	Vereinigtes Königreich
JPY	Yen	Japan
PLN	Zloty	Polen
USD	Dollar	USA

Tab. 3: Beispiele für Währungskürzel nach ISO 4217⁴⁵

4. Datums- und Zeitangaben

Die Schreibweise des Datums variiert in den einzelnen Ländern. Sogar in Ländern mit gleicher Sprache, wird das Datumsformat in unterschiedlicher Schreibweise benutzt. Die unterschiedlichen Notationen sorgen dadurch immer wieder für Mißverständnisse. Die Tabelle 4 zeigt unterschiedliche Datumsnotationen.

Mit der Norm ISO 8601:2000 (Data elements and interchange formats – Information interchange – Representation of dates and times) existiert ein einheitliches internationales Format für Datums- und Zeitangaben. Grundsätzlich sind die Elemente vom größten Element zum kleinsten Element (Jahr-Monat-Tag) geordnet.

45 In Anlehnung an o. V.: ISO 4217, Online im Internet: http://de.wikipedia.org/wiki/ISO_4217, 23.10.2004.

Land	Notation
USA	2004/02/28
Frankreich	28/2/04
Deutschland	28.02.04
Niederlande	28-2-04
Schweden	04-2-28

Tab. 4: Datumsformate im Ländervergleich

Das Kalenderdatum nach ISO 8601 wird in der folgenden Notation geschrieben:

YYYY-MM-DD

Um eine Verwechslungsgefahr bei Datumsangaben auszuschließen, ist es eine praktikable Lösung, in der Benutzeroberfläche den Monat teilweise oder ganz in Buchstaben und das Jahr in vier Stellen zu schreiben, z. B. 01. Jan. 2004. Zur Speicherung in der Datenbank sollte das Datums- und Zeitformat nach ISO 8601:2000 verwendet werden. Dadurch kann besser mit Algorithmen sortiert werden. (z. B. die aktuellen News nach dem Datum)⁴⁶

Das Zeitformat wird nach ISO 8601:2000 definiert als:

hh:mm:ss

Genutzt wird die 24-Stunden-Regelung. Die Uhrzeit 12 Uhr, 27 Minuten und 41 Sekunden wird geschrieben als 12:27:41.⁴⁷ Im Gegensatz zum Datumsformat wird das Zeitformat weltweit in einem einheitlichen Format nach ISO 8601 verwendet. Lediglich in den englischsprachigen Ländern ist die 12-stündige Notation mit dem Zusatz a.m./p.m. weit verbreitet. Bei Zeitangaben im Internet sind die Zeitzonen zu beachten (MEZ, EST, etc.). Einige Zeitzonen haben zusätzlich eine Sommer- und Winterzeit, die unter Umständen berücksichtigt werden muß. Ist die Angabe einer einheitlichen Zeitangabe über

⁴⁶ Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 182.

Zeitzone hinweg nötig, greifen die Medien, wie Fernsehen und Radio, auf die Greenwich Mean Time (GMT) oder auf die Coordinated Universal Time (UTC) zurück. Die UTC ist die offizielle Weltzeit und sollte der GMT vorgezogen werden.⁴⁸

5. Kontaktinformationen

Die Adreßangaben sind je nach Region sehr unterschiedlich. Über 100 verschiedene Adressformate sind in der Welt in Gebrauch. Sie unterscheiden sich je nach Land

- in der Formatierung,
- in der Anordnung der Darstellung,
- in der Beschriftung der Felder,
- um Angaben, z. B. des Staates oder der Provinz.

Ein WCMS muß möglichst alle Adreßangabe-Formen unterstützen, wobei es dem Benutzer möglich sein muß, seine Angaben vollständig, z. B. in einem Kontaktformular, tätigen zu können. Hierzu muß ggfs. die Reihenfolge der Kontaktfelder je nach Sprache in der Lokalisierung unterschiedlich angeordnet werden. Die Datenbank muß mit alternativen Feldern umgehen und die Adreßinformationen mit Metadaten versehen können. Metadaten informieren das System über die Anordnung und Verwendung der Felder bei der Anzeige.

6. Bilder und Icons

Neben Texten definiert sich eine Web Site auch über Bilder und Icons. Bei der Verwendung von Bildern und Icons sollte prinzipiell darauf verzichtet werden, sie mit Text zu versehen, da sonst für jede Sprachversion eine Lokalisierung durchlaufen werden muß. Besonders deutlich wird dieses Problem bei den Navigationselementen. Gerade die Navigationselemente sollten in Texte und Grafiken getrennt werden, damit eine Bediener-

47 Vgl. o.V.: ISO 8601 – Numeric representation of dates and time, Online im Internet: <http://www.iso.org/iso/en/prods-services/popstds/datesandtime.html>, 23.10.2004.

48 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 189.

oberfläche einfacher an eine neue Sprachversion angepaßt werden kann und die Pflege der Navigation erleichtert wird.

Bei der Auswahl von Icons ist darauf zu achten, daß diese möglichst sprach- und kulturneutral sind. Die Abbildung 5 zeigt unterschiedliche Möglichkeiten der Verwendung eines Symbols für einen Briefkasten. Der Briefumschlag ist ein universell verwendbares Symbol und wäre für die Symbolisierung „Kontaktaufnahme“ in einer Web-Funktion die beste Wahl.

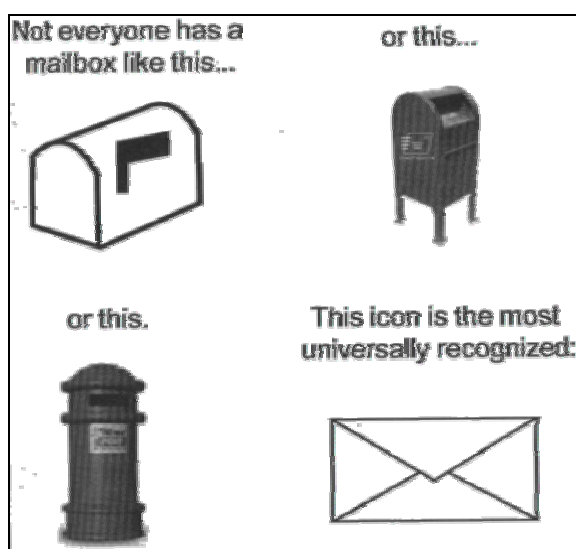


Abb. 5: Verschiedene Symbole für einen Briefkasten⁴⁹

Symbole dienen in einer Software dazu, bestimmte Funktionen zu verbildlichen und für den Nutzer zu vereinfachen. Es gibt zahlreiche Symbole, die eine bestimmte Bedeutung in Software-Produkten haben.⁵⁰ Beispielsweise steht der Papierkorb häufig für die Funktion „Löschen“ und das Fernglas für die Funktion „Suchen“. Die Software-Hersteller orientieren sich bei der Verwendung von Symbolen für bestimmte Funktionen

49 Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 307.

50 Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 182.

vielfach an bekannten und weitverbreiteten Microsoft-Office-Produkten.⁵¹ Der Vorteil von allgemein verständlichen Symbolen liegt gegenüber reinen Textbezeichnungen darin, daß die Symbole nicht an jede Sprachversion angepaßt werden müssen.

3.2.2 Anforderungen an eine sprachneutrale Bedienungsoberfläche

3.2.2.1 Zeichenexpansion

Beim Übersetzen von Texten von einer Sprache in andere Sprachen läßt sich das Phänomen des räumlichen Auseinanderziehens oder Zusammenziehens der Texte bei der Übersetzung beobachten. Das Wort „Katze“ besteht beispielsweise in der deutschen Sprache aus fünf Buchstaben, während es in der spanischen Sprache „Gato“ heißt und vier Buchstaben hat. In der englischen Sprache heißt Katze „Cat“ und besteht nur aus drei Buchstaben. Englische Texte vergrößern sich häufig, wenn sie in andere europäische Sprachen übersetzt werden – von ungefähr 15 % in Spanisch bis zu 35 % in Deutsch. Asiatische Sprachen hingegen benötigen weniger Zeichen als der englische Originaltext.⁵²

Das Auseinanderziehen oder Zusammenziehen des Textes tritt besonders deutlich bei kurzen Texten, wie z. B. Überschriften oder bei Texten in Navigationsleisten auf. Die folgende Abbildung 6 zeigt die Textexpansion bei einem Navigationsbutton.⁵³ Deutlich erkennbar ist, daß sowohl in der französischen, als auch in der deutschen Übersetzung der Text über den Button hinaus läuft.

Gerade Programmierer, die eine englische Sprachversion entwickeln, müssen das Problem der Textexpansion bei der Programmierung von einer Vorlage für mehrere Sprachen mit Platzhaltern berücksichtigen. Das Problem der Textschrumpfung ist ebenfalls nicht zu unterschätzen. Durch zu große Abstände zwischen den einzelnen Platzhaltern kann eine „leere“ Web Site entstehen. Die Tabelle 5 gibt einen Überblick über die Zei-

51 Vgl. Tiedjes, Tobias.: Anforderungen an Multilingualität von Content-Management-Systemen, Online im Internet: http://www.contentmanager.de/magazin/artikel_351_anforderungen_an_multilingualitaet_von_cms.html, 29.04.2004.

52 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 176.

53 Vgl. o. V.: Rubric's International Software Development Guidelines, Online im Internet: http://www.rubric.com/en/pdfs/Rubric_software_dev_guide.pdf, 09.10.2004, S. 8.

chenexpansion bei der Übersetzung vom Englischen in andere europäische Sprachen. Die Benutzeroberfläche bzw. die sprachenneutrale Vorlage muß so angelegt sein, daß durch Zeichenexpansion bzw. -schrumpfung das Web-Site-Design und die Funktonalitäten nicht beeinträchtigt werden.

Anzahl der englischen Zeichen	Zusätzlicher Platzbedarf in anderen Sprachen
1 - 10	200 - 300 %
11 - 20	100%
21 - 30	80%
31 - 50	60%
51 - 70	40%

Tab. 5: Zeichenexpansion bei der Übersetzung⁵⁴

English:	News
Italian:	Notizie
French:	Actualités
German:	Schlagzeilen
Japanese:	ニュース
Chinese:	新聞
Korean:	뉴스

Abb. 6: Zeichenexpansion beim Button „News“⁵⁵

54 Appleby, Stephen; Beires Nuno; Boualem Malek; Boves Losuis; Codogno Maurizio; den Os, Els; Pombo Prol, Marta; Vinesse Jérôme: Multilingual Web sites: Best practice, guidelines and architectures, Deliverable 1, Guidelines for buiding multilingual Web sites, Volume 3 of 5: Annex B, Overview of Language Processing Tools and Techniques, Online im Internet: <http://www.eures-com.de/~pub-deliverables/P900-series/P923/D1Vol3/>, 12.08.2004, S. 9

55 Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 177.

3.2.2.2 Platzhalter

Die wesentliche Grundidee eines WCMS ist die strikte Trennung von Inhalten und Darstellung. Die Verwirklichung der strikten Trennung kann durch die Verwendung von Platzhaltern und Grafiken für die Inhalte in einer Darstellungsvorlage (Template) ermöglicht werden. In einem Template werden die Texte durch Platzhalter repräsentiert. Grafiken hingegen sollten keine Textelemente beinhalten. Das Zusammenspiel von Platzhaltern und Grafiken kann an einem Beispiel erläutert werden.

In der Abbildung 7 setzt sich die Grafik aus einer Tabelle zusammen, deren Zellen zum einen Grafiken beinhalten und somit das (blaue) Kreuz erzeugen und zum anderen aus Platzhaltern, die in der Laufzeit mit den entsprechenden Werten besetzt werden. Damit im Beispiel die Seiten richtig angezeigt werden können, sind die Quellen meist nicht mehr in HTML geschrieben, sondern mit serverseitigen Skripten implementiert. Bei Abruf der Seite werden die Platzhalter ersetzt und die vollständige HTML-Seite wird an den Browser übergeben.



Abb. 7: Konzept für Platzhalter und Grafikelemente⁵⁶

Für jede Sprache wird eine Datei erstellt, die einen Eintrag für die verwendeten Platzhalter besitzen. Diese Dateien (Bezug zu Abb. 7) haben einen sehr einfachen Aufbau:

⁵⁶ Vgl. o. V.: Internationalisierung – wenn der Webauftritt mehrsprachig werden soll wird es kritisch, Online im Internet: <http://www.evodion.de/opencms/export/evodion/Infocenter/resources/internationalisierung.pdf>, 28.07.2004.

Abcdefg = ABCDEFG
Grafik1_1 = Willkommen
Grafik1_2 = auf der multilingualen Website

bzw. für Englisch

Abcdefg = ABCDEFG
Grafik1_1 = Welcome
Grafik1_2 = at the Multilingual Website

Sind alle Textkomponenten auf diese Art und Weise in Textdateien ausgelagert, kann eine weitere Sprache hinzugefügt werden, ohne daß sich der Programmcode der Web Site ändert. Die Abbildungen 11 und 12 im Anhang zeigen eine beispielhafte Web Page wie sie im Internet sichtbar ist und das entsprechende „Template-Gerüst“ dazu.

3.2.2.3 Benutzerspezifische Spracheinstellung

Die Einstellung der benutzerspezifischen Sprache der Bedienungsoberfläche kann prinzipiell auf zwei Arten erfolgen:

- Auswahl der bevorzugten Sprache auf der Bedienungsoberfläche,
- Aufrufen der Web Site mit seinem länderspezifischen Domain-Namen.

Die erste Möglichkeit sollte in einem mehrsprachigen WCMS in jedem Fall verwirklicht werden, d. h. der Benutzer sollte die Auswahlmöglichkeit für seine Sprache haben und diese an einer geeigneten Stelle (z. B. in Verbindung mit der Navigation) auf der Benutzeroberfläche finden. Die Abbildung 8 zeigt Beispiele, wie Sprachauswahl-Funktionen dem Benutzer angeboten werden können. Flaggensymbole für die Auswahl einer Sprache sollten nur eingesetzt werden, wenn der Benutzer explizit sein Heimatland auswählen kann, so wie in der Abbildung 8 bei <http://www.filiago.de>. In der mehrsprachigen Schweiz wird dieses Problem sehr deutlich. Um dieses Flaggenproblem elegant zu umgehen, bietet es sich an, entweder die exakte Bezeichnung der Sprache in die Auswahl aufzunehmen oder mit den Sprachkürzel nach ISO 639 z. B. EN, DE oder FR zu arbeiten (siehe Abschnitt 3.3.1 Sprachen- und Länder-Codierung).

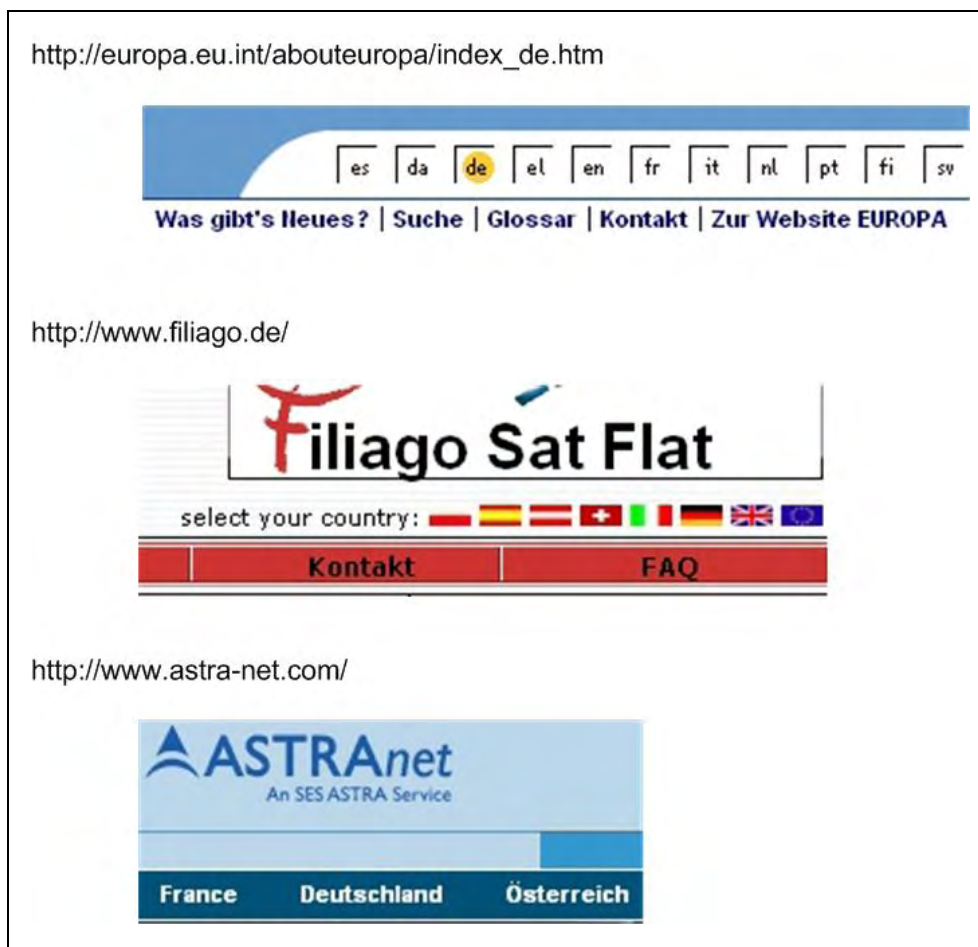


Abb. 8: Beispiele für eine Sprachauswahl⁵⁷

Die zweite Möglichkeit der Einstellung der benutzerspezifischen Sprache einer Web Site kann über die länderspezifischen Domain-Namen realisiert werden. Der Server erkennt entweder die Sprachpräferenz des Benutzers über IP-Adresse oder Browser-Einstellungen, und leitet ihn sofort an die jeweilige Web Site weiter oder der Server kann den User nicht lokalisieren und leitet den Benutzer auf die globale Einstiegsseite, auf der er seine bevorzugte Sprache und/oder Land auswählen kann. Die Abbildung 13 im Anhang verdeutlicht anhand des Beispiels von 3Com diese Funktionsweise mit länderspezifischen Domain-Namen und der Verwendung einer globalen Einstiegsseite.

⁵⁷ Screenshots von http://europa.eu.int/abouteuropa/index_de.htm, <http://www.filiago.de> und <http://www.astra-net.com> am 10.11.2004.

3.2.3 Speicher-Struktur der Web Site

In einem modernen WCMS werden die Inhalte in Datenbanken gespeichert und die Verknüpfung und Platzierung der Inhalte werden dynamisch zur Laufzeit erstellt. Zur besseren Orientierung bei den URL-Pfaden muß bei einem mehrsprachigen WCMS eine Speicher-Struktur festgelegt werden. Um einen Überblick über die verschiedenen Sprachversionen von Inhalten zu behalten und eine funktionierende Verlinkung von Inhalten zu garantieren, ist eine Baumstruktur unumgänglich. Bei der Struktur der Web Site gibt es zwei Möglichkeiten:⁵⁸

- Es existiert nur ein Strukturbaum. Den einzelnen Seiten sind dann jeweils die einzelnen Contentseiten/Artikel in allen verfügbaren Sprachen zugeordnet.
- Die Sprachvarianten der Web Sites werden in getrennten Strukturbäumen (Ordnerstrukturen) innerhalb des WCMS verwaltet.

Welche Speicher-Struktur sich für den Betrieb einer mehrsprachigen Web Site besser eignet, richtet sich nach der inhaltlichen Ausgestaltung der einzelnen Sprachversionen. Soll die Speicher-Struktur der einzelnen Sprachversionen auf Dauer identisch sein, so bietet sich die erste Möglichkeit mit einem gemeinsamen Strukturbaum an. Ein Vorteil ist die Möglichkeit, den Nutzer auf jeder Seite die Sprache individuell bestimmen zu lassen. Die Pflege mit einem gemeinsamen Strukturbaum ist prinzipiell nur möglich, wenn ein Team sämtliche Sprachvarianten pflegt, da alle Sprachvarianten eines Contents im selben Strukturbereich/Unterordner zu finden sind. Wird jedoch jede Sprachversion von einem anderen Team gepflegt oder sollen Abweichungen im Content zwischen den Sprachangeboten möglich sein, bieten sich getrennte Strukturbäume an.

Der linke Teil Abbildung 9 zeigt die Speicher-Struktur bei der Existenz nur eines Strukturbaumes. Aus der Abbildung läßt sich erkennen, daß die spanische und die englische Startseite im Strukturbaum nicht separiert werden. Im rechten Teil der Abbildung 9 sind die Sprachversionen nach getrennten Strukturbäumen geordnet. Für jede Sprachversion existiert ein eigenes Verzeichnis, welches alle lokalisierten Inhalte der jeweiligen Sprache enthält. Der Verzeichnisname im Beispiel orientiert sich dabei am Sprachenkürzel

⁵⁸ Vgl. Tiedjes, Tobias.: Anforderungen an Multilingualität von Content-Management-Systemen, a. a. O.

chern und Hilfstexten verwendet werden. Die Verwendung vieler Grafiken und verdeutlichender Screenshots (Abbildungen der realen Umgebung) ist für das Verständnis wichtig. Die Online-Dokumentation und Online-Hilfe sollte in den Sprachen vorliegen, in denen der Intranet-Zugang zur Pflege der Web Site angeboten wird. Einschränkungen des Sprachangebotes können vorgenommen werden, wenn die Hilfe kontextsensitiv angeboten wird, da die Benutzer hier nur kurze Textpassagen lesen und verstehen müssen. Der Umfang der Übersetzungen hängt von der Anzahl der Nutzer pro Sprache ab. Die Übersetzung von Print-Versionen der Handbücher und Hilfetexte in einzelne Sprachen hängt im wesentlichen ebenfalls von der Anzahl der Nutzer pro Sprache ab. Schulungen sollten ebenfalls in den Sprachen der Benutzer angeboten werden. Durch Schulungen in der Muttersprache der Schulungsteilnehmer wird die Gefahr von Mißverständnissen verringert.⁵⁹

3.2.5 Rechtliche Aspekte

An dieser Stelle soll nur ein kleiner Überblick über rechtliche Aspekte bei einem mehrsprachigen Betrieb einer Web Site gegeben werden. Je nach Art des Betriebes einer mehrsprachigen Web Site können die rechtlichen Aspekte sehr umfangreich sein und größere Anforderungen an ein mehrsprachiges WCMS richten. Die Nichteinhaltung rechtlicher Vorgaben auf einer Sprachversion kann weitreichende Folgen für den Betreiber der Web Site haben. Die Rechtslage kann von Land zu Land sehr unterschiedlich sein und muß in einer Sprach- bzw. Länderversion berücksichtigt werden. Relativ einfach sind der amerikanische und der europäische Markt, da diese Märkte sich durch homogene Richtlinien auszeichnen. Trotzdem müssen Unternehmen mit einer E-Business-Präsenz, die aus einer Niederlassung in einem anderen Land agieren, besonders auf die rechtlichen Richtlinien des betreffenden Landes achten. Zudem ist der Absatz über Ländergrenzen hinweg an internationale Abkommen gebunden, nach denen sich ein im Ausland ansässiges Unternehmen richten muß. Die rechtlichen Aspekte sind im Detail vor der Einführung einer Sprachversion zu prüfen und ggfs. sind zusätzlich rechtliche Anforderungen zu formulieren.

⁵⁹ Vgl. Tiedjes, Tobias.: Anforderungen an Multilingualität von Content-Management-Systemen, a. a. O.

Spezielle rechtliche Aspekte für eine globale Web Site sind:⁶⁰

- Warnhinweise und Dokumentationspflichten,
- Währungsangaben,
- Steuern (z. B. Umsatzsteuer und spezielle Verbrauchsteuern),
- Richtlinien für Werbung (z. B. Verbot der Werbung von Rechtsanwälten in Deutschland),
- Vertriebsrichtlinien und -einschränkungen (z. B. verbotene Substanzen im Handel),
- allgemeine Gesetze, Verordnungen und Richtlinien,
- Disclaimer und Impressum.

Gerade im öffentlichen Sektor ist eine Reihe von Gesetzen, Verordnungen und Richtlinien zu beachten. Das kann dazu führen, daß bestimmte Sprachversionen einer öffentlichen Web Site z. B. einer Stadt mit dem Hinweis versehen werden müssen, daß nur die Amtssprache(n) als rechtsbindend gelten kann.

3.3 Technikspezifische Anforderungen

3.3.1 Sprachen- und Länder-Codierung

Neben den aufgabenspezifischen Anforderungen müssen die technischen Anforderungen vor allem in der Entwicklung oder Weiterentwicklung eines WCMS berücksichtigt werden. Die Anforderungen technischer Natur beziehen sich hauptsächlich auf Programmier- oder Datenspeicherungsverfahren und müssen durch die Entwickler berücksichtigt werden.

Im Bereich des Internets werden spezielle Befehle (Tags) zur Identifikation der Sprache verwendet. Zum einen erkennt der Browser anhand dieser Tags die Sprache, zum anderen können Blindenschrift- und Übersetzungssoftware diese Tags nutzen, damit sie landesspezifische Sonderzeichen nicht falsch oder gar nicht übersetzen. Von der International Organisation for Standardization (ISO) wurde die ISO 639 veröffentlicht, die die

⁶⁰ Vgl. Zschau, Oliver; Traub, Dennis; Zahradka, Rik: Web Content Management – Websites professionell planen und betreiben, a. a. O., S. 186.

Spracherkennungen den Sprachen zuordnet. Gemäß des ISO-Web-Site gibt es ISO 639 in zwei Formen:⁶¹

- ISO 639-1:2002 – Codes for the representation of names of language – Part 1: Alpha-2 code (Kurzbezeichnung aus jeweils zwei Buchstaben)
- ISO 639-2:1998 - Codes for the representation of names of language – Part 2: Alpha-3 code (Kurzbezeichnung aus jeweils drei Buchstaben)

Beispielhafte Sprachkurzzeichen nach ISO 639-1:2002 sind „DE“ für Deutsch und „EN“ für Englisch. Ein beispielhaftes Sprachkurzzeichen nach ISO 639-1998 ist „GER“ für Deutsch.

ISO 3166 ist ein Standard für die Codierung von geographischen Einheiten, und wurde ebenfalls als ISO-Norm veröffentlicht. Der Standard besteht aus drei Teilen:

- der Codierung existierender Staaten und einiger abhängiger Gebiete (ISO 3166-1),
- der Codierung subnationaler Einheiten (ISO 3166-2),
- der Codierung ehemaliger Staaten (ISO 3166-3).

Der erste Teil (ISO 3166-1) listet seit 1974 u. a. zwei- (Alpha-2) und dreibuchstabige (Alpha-3) Länderkurzzeichen auf. Die zweibuchstabigen Kurzzeichen werden unter anderem als Top-Level-Domain-Kennzeichner (TLD) für Staaten verwendet.⁶² Beispielhafte Länderkurzzeichen nach ISO 3166-1 sind „UK“ für Großbritannien, „CA“ für Kanada und „AT“ für Österreich.

Die Norm ISO 3166-2 definiert Kurzformen für die Untergliederungen von Ländern oder abhängigen Gebieten. Im Falle Deutschlands sind es beispielsweise die Bundesländer. Hierzu folgt dem Landescode z. B. DE noch der ISO 3166-2-Variant-Code. Dieser wird durch ein Unter- oder Bindestrich getrennt, dem Landescode angehängt z. B. DE-BY für das Bundesland Bayern.

61 Vgl. o. V.: ISO 639, Online im Internet: http://de.wikipedia.org/wiki/ISO_639, 23.10.2004.

62 Nähere Informationen zur Verwaltung der Top-Level-Domain „DE“ unter <http://www.denic.de>.

ISO 3166-3 bildet seit 1974 alle Länder ab, die nicht mehr existieren bzw. deren Kürzel sich durch eine Umbenennung des Landes geändert haben wie z. B. die Sowjetunion.⁶³

Eine W3C-Richtlinie empfiehlt die richtige Kennzeichnung der Sprache, wenn in einer Web Site mit einer neuen Sprachversion gearbeitet wird. Dies geschieht nach HTML 4.0 mit dem „lang“ Attribut.⁶⁴ Ein Beispiel für den entsprechenden Tag „lang“ kann folgendermaßen aussehen: `How are you`.⁶⁵

Der „lang“-Tag kann ein Sprachenkürzel, ein Länderkürzel oder beides enthalten. Die Kombination (siehe Tabelle 6) ergibt sich aus der geographischen Codierung nach ISO 3166 für Länder mit den Sprachcodes und ISO 639 für Sprachen wird in RFC 1766 näher spezifiziert. Die Sprachkürzel werden durch Kleinbuchstaben symbolisiert, die Länderkürzel durch Großbuchstaben. Neben dem Unterstrich kommt als Trennzeichen auch oft der Bindestrich zum Einsatz. Länder- und Sprachcodes definieren aber nicht nur eine Landessprache bzw. Regionalsprache, sondern werden auch zur Lokalisierung herangezogen.⁶⁶

Sprach- /Länderkürzel	Sprache
en-US	Englisch (USA)
en-UK	Englisch (GB)
de-DE	Deutsch
it-IT	Italienisch
fr-CA	Französisch (Kanada)
fr-FR	Französisch (Frankreich)
es-ES	Spanisch (Spanien)
es-Mx	Spanisch (Mexiko)

Tab. 6: Sprach- und Länderkürzel⁶⁷

63 Vgl. o. V.: ISO 3166, Online im Internet: http://de.wikipedia.org/wiki/ISO_639, 23.10.2004.

64 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 51.

65 Weitere Details in RFC 1766, unter <http://www.ietf.org/rfc/rfc1766.txt>.

66 Vgl. o. V.: ISO 3166, a. a. O.

67 Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 51.

Da die Anzahl der gesprochenen Sprachen im Internet auch in den nächsten Jahren weiter wachsen wird, wird es immer wichtiger, daß der Content mit Sprach- und Länderkürzel und den entsprechenden „lang“-Tag richtig gekennzeichnet wird. Durch die richtige Verwendung des „lang“-Tags kann man folgende Funktionen unterstützen:⁶⁸

- Unterstützung der Suchmaschinen bei der Indizierung der Sites nach Sprachen,
- Unterstützung von Sprachwiedergabesystemen für z. B. Seh-Behinderte,
- Unterstützung der Browser bei der richtigen Textwiedergabe nach Grammatikregeln,
- Unterstützung der Rechtschreib- und Grammatikprüfung.

3.3.2 Zeichensatzunterstützung

Der ASCII-Standard (American Standard Code for Information Interchange) beschreibt einen Zeichensatz, der auf dem lateinischen Alphabet basiert, wie er in der englischen Sprache von Computern und anderen Kommunikationseinrichtungen zur Textdarstellung verwendet wird. In diesem Standard ist die Zuordnung von digital dargestellten Ganzzahlen zu den in der normalen Schriftsprache geschriebenen Zeichen codiert. Mit Hilfe des Codes können digitale Geräte codierte Informationsinhalte senden, empfangen und verarbeiten. Dieser Code verwendet binäre Ganzzahlen, die mit sieben binären Ziffern dargestellt werden (entspricht 0 bis 127 Zeichen), um Informationen darzustellen.⁶⁹

Das Problem bei dem ASCII-Standard ist, daß er keine diakritischen Zeichen (z. B. Umlaute) enthält, die in vielen Sprachen auf der Basis des lateinischen Alphabets verwendet werden. Dazu kommt noch, daß viele Sprachen nicht auf dem lateinischen Alphabet basieren, wie z. B. Chinesisch, Japanisch, Hebräisch oder Arabisch – einige von diesen Sprachen sind in der Schreibung rechts/links, statt links/rechts orientiert und umfassen deutlich mehr Zeichen (Chinesisch ca. 8000 Zeichen).⁷⁰ Deshalb entwickelten verschiedene Hersteller eigene 8-Bit-Codes, um 255 Zeichen darstellen zu können. Der

68 Vgl. Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 51.

69 Vgl. o. V.: ASCII, Online im Internet: <http://de.wikipedia.org/wiki/ASCII>, 05.11.2004.

70 Vgl. o. V.: Rubric's International Software Development Guidelines, a. a. O., S. 5.

Codepage 437 genannte Code kam auf dem IBM PC zur Anwendung.⁷¹ Daraus hervorgegangen ist im Laufe der Zeit die Normenfamilie ISO 8859 (ISO/IEC 8859), die in derzeit 15 verschiedenen Teilnormen unterschiedliche 8-Bit-Zeichensätze für die Informationstechnik definiert. Bekannte Zeichensätze sind:

- ISO 8859-1 (Latin-1) – Westeuropäisch
- ISO 8859-2 (Latin-2) – Osteuropäisch
- ISO 8859-5 (Kyrillisch)
- ISO 8859-6 (Arabisch)
- ISO 8859-7 (Griechisch)
- ISO 8859-15 (Latin-9) – Westeuropäisch mit Eurozeichen

Die ersten 128 Positionen sind bei allen Teilnormen identisch und entsprechen den Zeichen des 7-Bit-Satzes US-ASCII. Es folgen 32 Steuerzeichen, und die letzten 96 Positionen enthalten regionale Sonderzeichen, die den Unterschied zwischen den Teilnormen ausmachen.⁷² Die meisten westeuropäischen Sprachen können zusammengenommen in einem Byte (= 8 Bit) untergebracht werden. Andere Sprachen, insbesondere asiatische, können wegen der Vielzahl unterschiedlicher Zeichen nicht mehr in einem Byte untergebracht werden.⁷³ Systeme müssen demnach die Fähigkeit besitzen, „Multi-Byte-Zeichen“ zu bearbeiten, um eine breite Unterstützung der verschiedenen Zeichen der Sprachen zu gewährleisten.

Aufgrund der Anforderungen an einen Zeichensatz mit Multi-Byte-Unterstützung wurde der internationale Standard „Unicode“ (in Anlehnung an ISO 10646) entwickelt, mit dem Ziel, einen Code für jedes grafische Zeichen oder Element aller bekannten Schriftkulturen und Zeichensysteme (mathematische, technische und musikalische Symbole, etc.) festzulegen.⁷⁴ Unicode und ISO 10646 besitzen die gleichen Codes für die unterschiedlichen Zeichen, aber der Unicode-Standard enthält bedeutend mehr Angaben für die Implementierung von Textverarbeitungslösungen (Darstellung, Sortierreihenfolgen, Schriften, die von rechts nach links laufen etc.). Unicode ist streng genommen eine Un-

71 Vgl. o. V.: ASCII, a. a. O.

72 Vgl. o. V.: ISO 8859, Online im Internet: http://de.wikipedia.org/wiki/ISO_8859, 05.11.2004.

73 Vgl. o. V.: Rubric's International Software Development Guidelines, a. a. O., S. 5.

74 Vgl. o. V.: ISO 10646, Online im Internet: http://de.wikipedia.org/wiki/ISO_10646, 05.11.2004.

termenge von ISO 10646. Während ISO 10646 Zeichencodes mit bis zu 31 Bit zuläßt, sind bei Unicode max. 21 Bit erlaubt. Zur Zeit ist dies aber weniger von Bedeutung, da noch keine Zeichen jenseits des 21-Bit-Bereiches definiert wurden und das auch in absehbarer Zeit nicht der Fall sein wird. Unicode will das Problem der verschiedenen inkompatiblen Codierungen in den unterschiedlichen Ländern beseitigen. Bisher erfordern die bestehenden Standards einen hohen Konvertierungsaufwand beim Datenaustausch. Da die meisten Zeichenkodierungen nur wenige Sprachen unterstützen, war es in der Praxis bisher oft nicht möglich, verschiedene Sprachen im selben Text zu verwenden. Der Unicode-Standard unterstützt 1.114.112 ($=2^{20}+2^{16}$) Zeichen (von U+00000 bis U+10FFFF).

Unicode-Zeichen werden oft in der Form U+XXXX angegeben, wobei das U für die Unicode steht, und XXXX die hexadezimale Zahl ist, mit der das Zeichen codiert ist. So entsprechen beispielsweise die ersten 128 Zeichen, U+0000 bis U+007F, dem ASCII-Zeichensatz. Für die meisten Unicode-Zeichen gibt es frei erhältliche Schriftarten. Es werden auch kommerziell erhältliche Schriftarten angeboten. Dabei gilt für eine unicode-konforme Schriftart, daß die Zeichen an dem vom Unicode vorgeschriebenen Platz zu finden sind, aber nicht, daß ein entsprechendes Zeichen in der Schriftart vorhanden sein muß. Es gibt einige Unicode-Schriften, die nur ca. 600 Zeichen implementiert haben, was aber für viele Zwecke bereits genügt und einen bedeutenden Fortschritt gegenüber den 255 Zeichen der 8-Bit-Schriftarten darstellt. Für die Repräsentation der Unicode-Zeichen zum Zweck der elektronischen Datenverarbeitung gibt es verschiedene Transformationsformate, die sogenannten Unicode Transformation Formats (UTF). In jedem der Formate lassen sich alle 1.114.112 in Unicode-Standard enthaltenen Zeichen darstellen, wobei sich jedes Format ohne Einschränkungen in ein anderes UTF Format konvertieren läßt. Die Formate unterscheiden sich hauptsächlich durch deren Platzbedarf auf Speichermedien sowie Codierungs- und Decodierungsaufwand zur Laufzeit des Codes.

UTF-32 (UCS-4) stellt jedes Unicode-Zeichen mit einem 4 Byte (32 Bit) langen Code dar. Ein Nachteil ist, daß sehr viel mehr Speicherplatz verbraucht wird als bei UTF-8 oder UTF-16, die je nach Zeichen variable Bitlängen verwenden.

UTF-16 (UCS-2) ist die älteste Unicode Kodierung und benötigt 2 Byte (16 Bit) oder 4 Byte (32 Bit). Die Kodierung erlaubt maximal 1.114.112 Codes und schränkt damit aus Kompatibilitätsgründen alle anderen UTF Kodierungen ein.

UTF-8 ist die beliebteste Codierung für Unicode. Im UTF-8 Format wird jedes Unicode-Zeichen in einer speziell codierten Bytekette von variabler Länge abgebildet. Ein erheblicher Vorteil gegenüber allen anderen Formaten ist, daß alle ASCII-Zeichen unverändert gültige Zeichen in UTF-8 sind und somit auch nur 1 Byte (8 Bit) an Daten belegen (siehe Abbildung 10). Anwendungen, die ausschließlich den ASCII-Zeichensatz verwenden, funktionieren unverändert auch mit UTF-8. Für alle auf dem lateinischen Alphabet basierenden Schriften ist UTF-8 die platzsparendste Methode zur Abbildung von Unicode Zeichen.⁷⁵

glyph	character description	code point hex (decimal)	UTF-8
A	LATIN CAPITAL LETTER A	0041 (65)	01000001
ص	ARABIC LETTER SAD	0635 (1.589)	11011000 10110101
ჲ	GEORGIAN LETTER PHAR	10E4 (4.324)	11100001 10000011 10100100
の	HIRAGANA LETTER NO	306E (12.398)	11100011 10000001 10101110

Abb. 10: UTF-8 bei verschiedenen Zeichen⁷⁶

Die Nutzung von Unicode erleichtert die Lokalisierung erheblich. Der Unicode-Standard wurde von vielen führenden IT-Unternehmen wie Apple, Hewlett Packard, IBM, Microsoft, Oracle, SAP etc. übernommen. Unicode wird für moderne Standards wie XML, Java, ECMA-Script (JavaScript) etc. benötigt. Es wird von vielen Softwareprodukten z. B. Linux, Microsoft Windows und den gängigsten Browsern der neueren Ge-

⁷⁵ Vgl. o. V.: Unicode, Online im Internet: <http://de.wikipedia.org/wiki/Unicode>, 05.11.2004.

⁷⁶ Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 445.

neration wie Microsoft Internet Explorer, Mozilla, Opera unterstützt.⁷⁷ Ebenso unterstützt die Datenbank MySQL ab Version 4.1 die Transformationsformate UTF-16 und UTF-8 des Unicoes. MySQL 4.1 besitzt zudem eine sehr viel bessere Unterstützungsfunktion für den Umgang mit wechselnden Zeichensätzen und Collationen (Sortierungen) auf allen Ebenen – von der Anordnung der Spalten bis zur Grundeinstellung des Servers – und jede Ebene kann sogar bestimmen, welchen Zeichensatz sie zur Datenübertragung nutzen will.

Ein Problem in Datenbanken ist die Sortierreihenfolge der Inhalte, da diese sehr oft sprachenabhängig sind. Schwedisch und Deutsch nutzen im allgemeinen den gleichen Zeichensatz ISO-8859-1 (latin 1). Aber es gibt Zeichen, die in beiden Sprachen anders sortiert werden. Im Schwedischen kommt das Z vor dem Ö, während es im Deutschen umgekehrt ist. Unicode stellt mit dem Unicode Collation Algorithm (UCA) einen allgemeinen Spezialalgorithmus für die Sortierreihenfolge der Strings zur Verfügung. Daneben gibt es auch die Default Unicode Collation Element Table (DUCET), die eine Standard-Sortierung für alle Unicode-Zeichen enthält. MySQL 4.1 hat diesen allgemein definierbaren Sortierreihenfolgealgorithmus UCA implementiert.⁷⁸

Unicode ist der ideale Zeichensatz für eine mehrsprachige Umgebung im WWW. Durch Unicode werden alle Zeichen der weltweit bekannten Sprachen abgedeckt, was für eine Sprachenneutralität eines Zeichensatzes sorgt. Dies war bei den vorher verwendeten Zeichensätzen nicht möglich. Durch Unicode werden viele Zeichensatzprobleme für Web Entwickler beseitigt; Unicode hat sich daher schon heute zum inoffiziellen Zeichensatz für das Internet entwickelt.

4 Zusammenfassung

„Think global – act local!“ Ein bekannter Spruch, der im World Wide Web seine Gültigkeit nicht verloren hat. Im Zuge der zunehmenden Globalisierung steigt die Notwen-

⁷⁷ Vgl. Lam, Jason: J2ME And Unicode, Online im Internet: <http://www.webpronews.com/printable.php>, Printable Version, 18.07.2004.

⁷⁸ Vgl. Winstead, Jim: Unicode And Other Funny Characters, Online im Internet: <http://dev.mysql.com/tech-resources/articles/4.1/unicode.html>, 05.11.2004.

digkeit für einen internationalen mehrsprachigen Web-Auftritt, der auf die jeweilige Zielgruppe lokalisiert zugeschnitten wird. Für den Anbieter einer globalen Web Site stellen sich – bedingt durch die unterschiedlichen, angesprochenen Kulturräume – verschiedene Probleme und Aufgaben. Eine globale Web Site zu erstellen heißt unter anderem, kulturelle Unterschiede zu erkennen und entsprechend in der E-Business-Strategie zu berücksichtigen.

Der zentrale Prozeß, der bei einer Globalisierung eines Web Site durchgeführt wird, ist die Internationalisierung. Internationalisierung bedeutet die Erstellung einer sprachenneutralen Version einer Web Site, in die im späteren Betrieb verschiedene Sprachversionen integriert werden können. Je sorgfältiger die Internationalisierung durchgeführt wird, desto weniger Aufwand entsteht bei der Einführung einer weiteren Sprachversionen. Daher resultieren aus der Internationalisierung verschiedene aufgaben- und technikspezifische Anforderungen an ein WCMS, die bei einer Fachkonzeption oder Auswahl eines mehrsprachigen WCMS berücksichtigt werden müssen. Die wichtigsten Anforderungen sind:

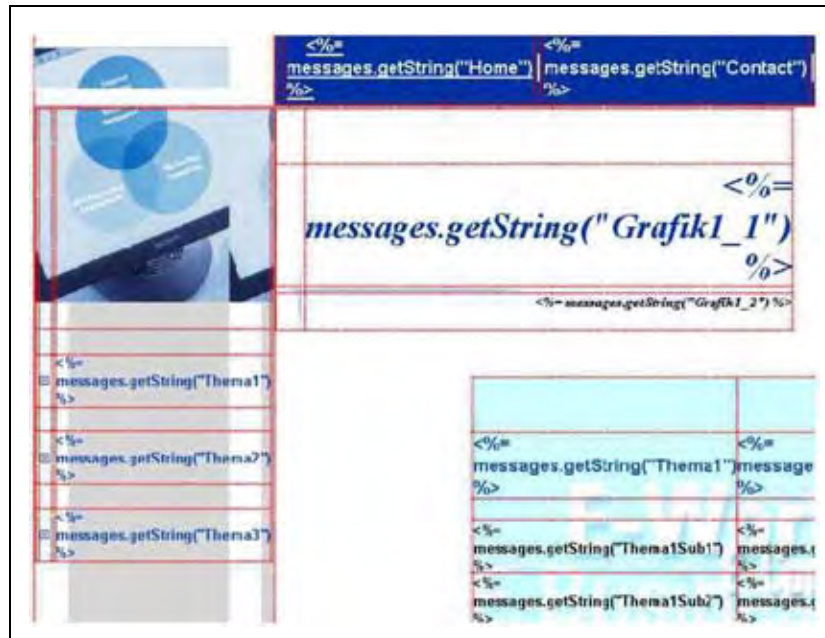
- Eliminierung aller sprachlichen Elemente im Quellcode
- Erstellung einer sprachenneutralen Benutzeroberfläche
- Beschränkung auf ein ausgewähltes Sprachenset
- Zeichensatzunterstützung im Quellcode und in Datenbanken
- Sprachen- und Ländercodierung im Quellcode
- Ordner-Architektur
- Mehrsprachige Hilfe und Dokumentation
- Berücksichtigung von gesetzlichen und kulturellen Rahmenbedingungen

Im Anschluß an die Internationalisierung wird die Lokalisierung für verschiedene Sprachen durchgeführt. Lokalisierung umfaßt die Anpassung einer Web Site an die sprachlichen und kulturellen Gegebenheiten in einem bestimmten geographischen und/oder ethnisch umschriebenen Gebiet. Bei der Lokalisierung einer Web Site müssen neben der Übersetzung der Inhalte auch z. B. Datums-, Zeit-, Währungs- und Temperaturangaben sowie Maßangaben und Umrechnungsgrößen angepaßt werden. Darüber hinaus müssen aber auch Farbgebung oder Grafiken angepaßt werden.

Anhang



Abb. 11: Demo-Site im WWW

Abb. 12: Demo-Site mit JSP-Gerüst⁷⁹

79 o. V.: Internationalisierung – wenn der Webauftritt mehrsprachig werden soll wird es kritisch, a. a. O.

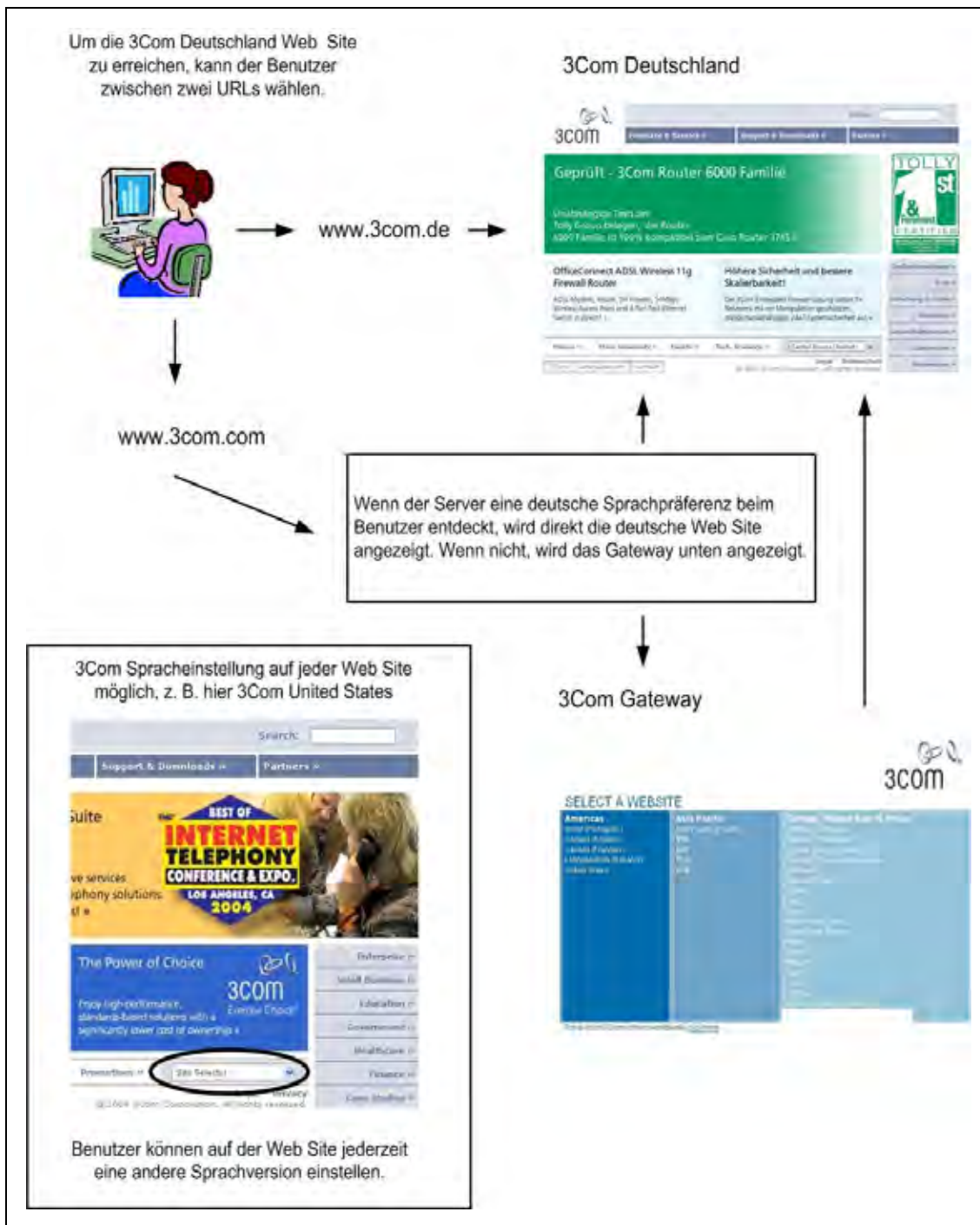


Abb. 13: Erreichbarkeit der deutschen 3Com Web Site⁸⁰

80 Yunker, John: Beyond Borders: Web Globalization Strategies, a. a. O., S. 284.

Literaturverzeichnis

1. **Bashaash, Mahmood:** Ten Major Tips To Develop A Multilingual Web Site To Work, Online im Internet: <http://www.devwebpro.com>, 02.04.2004.
2. **Bauer, Herbert:** Unternehmensportale – Geschäftsmodelle, Design, Technologien, Bonn: Galileo Press GmbH 2001.
3. **Boiko, Bob:** Content Management Bible, New York: Hungry Minds 2002.
4. **Bullinger Hans-Jörg (Hrsg.): Schuster, Erwin; Wilhelm, Stephan:** Content Management Systeme – Auswahlstrategien, Architekturen und Produkte, 5. Aufl., Düsseldorf: Verlagsgruppe Handelsblatt 2001.
5. **Dunlap, Bill:** Using Multilingual Websites, Online im Internet: <http://www.webpronews.com/news/webdevelopmentnews/wpn-42-20020416 UsingMultilingualWebsites.html>, 18.07.2004.
6. **Jablonski, Stefan; Meiler, Christian:** Web-Content-Managementsysteme, in: Informatik Spektrum 2/2002.
7. **Kargl, Herbert:** Management und Controlling von IV-Projekten, München, Wien: Oldenbourg, 2000.
8. **Kraemer, Viktor Hans:** Content-Management - WCMS bringen Ordnung ins Datenchaos - Trends und Technologien im Web-Content-Management-Markt, in: Computerwoche 37/2000, S. 76-77.
9. **Krüger, Dennis Jörg; Kopp Matthias:** Web Content managen – Professioneller Einsatz von Content-Management-Systemen, München: Markt + Technik Verlag, 2002.
10. **Lam, Jason:** J2ME And Unicode, Online im Internet: <http://www.webpronews.com/printable.php>, Printable Version, 18.07.2004.
11. **Manhartsberger, Martina.; Musil, S.:** Web Usability – Das Prinzip des Vertrauens, Bonn: Galileo Press 2002.
12. **Metzen, Heinz:** Das globale Informationszeitalter entsteht auf dem Marktplatz, nicht in Zukunftsministerien oder auf Weltgipfeln, Online im Internet: http://www.sommmeruni-uni-osnabrueck.de/04_1.htm, 28.09.2004.
13. **Niemann, F.:** Anpassungsaufwand ist nicht zu unterschätzen - Content-Management bringt heterogene Daten ins Netz, in: Computerwoche 47/1998, S. 38-40.
14. **o. V.:** ASCII, Online im Internet: <http://de.wikipedia.org/wiki/ASCII>, 05.11.2004.
15. **o. V.:** Content Management Systeme – Software und deren Einsatzfelder für PR, Online im Internet: <http://v.hdm-stuttgart.de/projekte/websitepr/content.php>, 15.10.2004.

16. **o. V.:** Global Internet Statistics (by Language), Online im Internet: <http://www.global-reach.biz/globstats/index.php3>, 03.09.2004.
17. **o. V.:** Globalisierung, Online im Internet: <http://de.wikipedia.org/wiki/Globalisierung>, 07.09.2004.
18. **o. V.:** Globalization, internationalization, localization: an overview, Online im Internet: <http://www.globalization.com>, 20.08.2004.
19. **o. V.:** Glossar: Content Management System, Online im Internet: http://www.contentmanager.de/ressourcen/glossar_10_content_management_system.html, 09.07.2004.
20. **o. V.:** Internationalisierung – wenn der Webauftritt mehrsprachig werden soll wird es kritisch, Whitepaper, Hamburg: evodion Information Technologies GmbH, 2004.
21. **o. V.:** ISO 4217, Online im Internet: http://de.wikipedia.org/wiki/ISO_4217, 23.10.2004.
22. **o. V.:** ISO 639, Online im Internet: http://de.wikipedia.org/wiki/ISO_639, 23.10.2004.
23. **o. V.:** ISO 8601 – Numeric representation of dates and time, Online im Internet: <http://www.iso.org/iso/en/prods-services/popstds/datesandtime.html>, 23.10.2004.
24. **o. V.:** ISO 8859, Online im Internet: http://de.wikipedia.org/wiki/ISO_8859, 05.11.2004.
25. **o. V.:** ISO 10646, Online im Internet: http://de.wikipedia.org/wiki/ISO_10646, 05.11.2004.
26. **o. V.:** Lokalisation, Online im Internet: <http://de.wikipedia.org/wiki/Globalisierung>, 07.09.2004.
27. **o. V.:** Products must first meet the technological requirements of global markets, Online im Internet, <http://www.rubic.com/en/services/internationalization.php>, 10.10.2004.
28. **o. V.:** Rubric's International Software Development Guidelines, Online im Internet: http://www.rubic.com/en/pdfs/Rubic_software_dev_guide.pdf, 09.10.2004.
29. **o. V.:** Subscription offer MultiLingual in: Multilingual Content Management , 45 Supplement, Jan/Feb 2002, Hrsg. MultiLingual Computing, Sandpoint, 2002, S. 24.
30. **o. V.:** The IP-to-Country Handbook, Online im Internet: <http://ip-to-country.webhosting.info/book/print/3>, 19.07.2004.
31. **o. V.:** Unicode, Online im Internet: <http://de.wikipedia.org/wiki/Unicode>, 05.11.2004.
32. **o. V.:** Was ist das Web-Portal-System?, Online im Internet: <http://wi.uni-giessen.de/wps/transmit/editionen/beitrag/Transmit/12>, 19.09.2004.

33. **Porst, Axel:** Web Content Management und Workgroup Computing, Online im Internet: <http://www.e-gateway.de/eco/contentman.cfm>, 07.07.2004.
34. **Scholz:** CMS-Grundlagen: Content-Management – Einführung 1 – Auswahlkriterien für ein WCMS im Detail, Online im Internet: <http://www.content-syndication-broker.de/crm-wcms-artikel-seite31-folge6.html>, 07.09.2004.
35. **Schuster, Erwin; Wilhelm, Stephan:** Content Management, in: Informatik Spektrum, Band 23, Nr. 6, Dezember 2000, Heidelberg: Springer 2000, S. 373-375.
36. **Schwickert, Axel C.;** Ostheimer, Bernhard; Franke, Thomas S.: eUniversity – Web-Site-Generierung und Content Management für Hochschuleinrichtungen, in: Arbeitspapiere WI, Nr. 9/2000, Hrsg.: Lehrstuhl für Allg. BWL und Wirtschaftsinformatik, Johannes Gutenberg-Universität: Mainz 2000.
37. **Schwickert, Axel C.:** Dezentrales Web Content Management, in: Arbeitspapiere WI, Nr. 5/2004, Hrsg.: Professur BWL – Wirtschaftsinformatik, Justus-Liebig-Universität Gießen, Gießen , 2004, S. 3.
38. **Tiedjes, Tobias.:** Anforderungen an Multilingualität von Content-Management-Systemen, Online im Internet: http://www.contentmanager.de/magazin/artikel_351_anforderungen_an_multilingualitaet_von_cms.html, 29.04.2004.
39. **Wiegand, Dorothee:** Multikulti-Software – Wie Programme fremde Sprachen lernen, in: C't, Heft 7, 2004, S. 186-189.
40. **Winstead, Jim:** Unicode And Other Funny Characters, Online im Internet: <http://dev.mysql.com/tech-resources/articles/4.1/unicode.html>, 05.11.2004.
41. **Yunker, John:** Beyond Borders: Web Globalization Strategies, 2. unveränderte Auflage, New Riders Publishing 2003.
42. **Yunker, John:** Building a Global Web Site – The Guide from MultiLingual Computing & Technology, in: Global Web, 55 Supplement, April/May 2003, Hrsg: MultiLingual Computing Inc., Sandpoint, 2003, S. 4 – 9.
43. **Zschau, Oliver.:** Komponenten eines Content Management Systems, Online im Internet: http://www.contentmanager.de/magazin/artikel_16komponenten_eines_content_management_systems.htm, 10/2000.
44. **Zschau, Oliver; Traub, Dennis; Zahradka, Rik:** Web Content Management – Websites professionell planen und betreiben, 2. überarb. und erw. Auflage, Bonn: Galileo Press GmbH 2002.



- Reihe:** **Arbeitspapiere Wirtschaftsinformatik** (ISSN 1613-6667)
- Bezug:** Online-Bestellung unter <http://wi.uni-giessen.de> → Forschung
- Herausgeber:** Univ.-Prof. Dr. Axel C. Schwickert
 Professur BWL – Wirtschaftsinformatik
 Justus-Liebig-Universität Gießen
 Fachbereich Wirtschaftswissenschaften
 Licher Straße 70
 D – 35394 Gießen
 Telefon (0 64 1) 99-22611
 Telefax (0 64 1) 99-22619
 eMail: Axel.Schwickert@wirtschaft.uni-giessen.de
 <http://wi.uni-giessen.de>
- Ziele:** Die Arbeitspapiere dieser Reihe sollen konsistente Überblicke zu den Grundlagen der Wirtschaftsinformatik geben und sich mit speziellen Themenbereichen tiefergehend befassen. Ziel ist die verständliche Vermittlung theoretischer Grundlagen und deren Transfer in praxisorientiertes Wissen.
- Zielgruppen:** Als Zielgruppen sehen wir Forschende, Lehrende und Lernende in der Disziplin Wirtschaftsinformatik sowie das IT-Management und Praktiker in Unternehmen.
- Quellen:** Die Arbeitspapiere entstehen aus Forschungsarbeiten, Diplom-, Studien- und Projektarbeiten sowie Begleitmaterialien zu Lehr- und Vortragsveranstaltungen der Professur BWL – Wirtschaftsinformatik, Univ. Prof. Dr. Axel C. Schwickert, Justus-Liebig-Universität Gießen.
- Hinweise:** Wir nehmen Ihre Anregungen und Kritik zu den Arbeitspapieren aufmerksam zur Kenntnis und werden uns auf Wunsch mit Ihnen in Verbindung setzen.
- Falls Sie selbst ein Arbeitspapier in der Reihe veröffentlichen möchten, nehmen Sie bitte mit dem Herausgeber unter obiger Adresse Kontakt auf.
- Informationen über die bisher erschienenen Arbeitspapiere dieser Reihe und deren Bezug erhalten Sie auf der Web Site der Professur unter der Adresse <http://wi.uni-giessen.de>