

# Erweiterung des Internetangebotes durch automatische Übersetzung

von Michael Mönnich

---

Durch die zunehmend internationale Ausrichtung der deutschen Universitäten ergibt sich auch für die Universitätsbibliotheken die Notwendigkeit, englischsprachige Versionen ihrer Webseiten anzubieten. Dies ist wichtig nicht nur für die Studierenden in Karlsruhe kommen derzeit 20 Prozent der Studierenden aus dem Ausland, sondern auch für Gastwissenschaftler, die für einige Monate an einer deutschen Universität forschen und in dieser kurzen Zeit die Dienste der Bibliothek nutzen wollen, jedoch häufig nur über geringe Deutschkenntnisse verfügen. Hinzu kommt in Karlsruhe, dass in diesem Jahr das sogenannte International Department" den Studienbetrieb aufnahm. An dieser neuen Einrichtung der Universität werden gebührenpflichtige Studiengänge angeboten, die sich vornehmlich an ausländische Studierende richten und ausschließlich in englischer Sprache abgehalten werden. Von daher gibt es also genügend Gründe, ein Webangebot in englischer Sprache aufzubauen.

## 1. Manuell erstellte Webseiten

Es scheint zwar relativ einfach, die vorhandenen Webseiten ins Englische zu übertragen, jedoch zeigen sich bei der konkreten Umsetzung einige Probleme. Zunächst bedarf es Personen, die in der Lage sind, die deutschen Seiten in korrektes Englisch zu übertragen, inklusive der bibliothekarischen und sonstigen Fachtermini. Sofern nicht auf Mitarbeiter der Bibliothek oder anderer Einrichtungen der Universität zurückgegriffen werden kann, wird meist ein Übersetzungsbüro beauftragt.

### Probleme der manuellen Übersetzung

Grundsätzlich bereiten bei der manuellen Übersetzung dynamisch generierte HTML-Seiten wie zum Beispiel Trefferlisten oder Titelaufnahmen aus dem OPAC Schwierigkeiten. Sie können auf diese Weise in der Regel nicht erfasst werden.

In der Praxis als viel größeres Problem zeigt sich jedoch der Aufwand, die einmal erstellten fremdsprachigen Versionen mit den deutschen Originalseiten kongruent zu halten. Die Anzahl statischer Webseiten der Universitätsbibliothek Karlsruhe beträgt zum Beispiel über 600, und allein die inhaltliche Anpassung der deutschen Seiten erweist sich häufig genug als problematisch, da ständig Veränderungen stattfinden, von geänderten Mahngebühren über neue Subito-Tarife, neue Mitarbeiter bis hin zu neuen Dienstleistungen. Es sollte aber vermieden werden, im WWW veraltete oder inkonsistente Informationen zu präsentieren, da dies unprofessionell wirkt.

Eine einfache Lösung wäre natürlich, nur noch englische Webseiten zu produzieren, wie es einige Fakultäten der Universität Karlsruhe tun; angesichts der deutlich inhomogeneren Kundschaft der UB erschien dieser Schritt noch zu radikal.

## 2. Maschinelle Übersetzung

Vor dem Hintergrund dieser Schwierigkeiten hat sich die Universitätsbibliothek Karlsruhe entschlossen, auf die manuelle Übersetzung von Webseiten zu verzichten und statt dessen die maschinelle Übersetzung zu präferieren.

Es gibt im Internet eine ganze Reihe von Übersetzungsdiensten, die das automatische Übersetzen von deutschsprachigen Webseiten ins Englische anbieten. Bekannte Beispiele sind die kostenlosen Dienste von Alta Vista (<http://babelfish.altavista.com>), Heise (<http://systran.heisoft.de>), Free Translation (<http://www.freetranslation.com>) oder des amerikanischen Internetportals Go.com (<http://translator.go.com>). Diese Systeme arbeiten so, dass eine URL eingegeben und die entsprechende übersetzte Version zurückgeliefert wird. Zugrunde liegt meistens eine Software der Firma Systran (<http://www.systransoft.com>).

Bei der Auswahl des Systems wurde von der Universitätsbibliothek dem "Go Translator" der Vorzug gegeben, da dieser explizit die Möglichkeit anbietet, auf eigenen Seiten einen Link zu setzen, der das maschinelle Übersetzen der Seite über den Go-Translator anstößt. Genauere Angaben dazu findet man unter dem Link "Add to your Web site" auf der Go Translator-Homepage. Hinzu kommt auch, dass die Performance von Go akzeptabel ist.

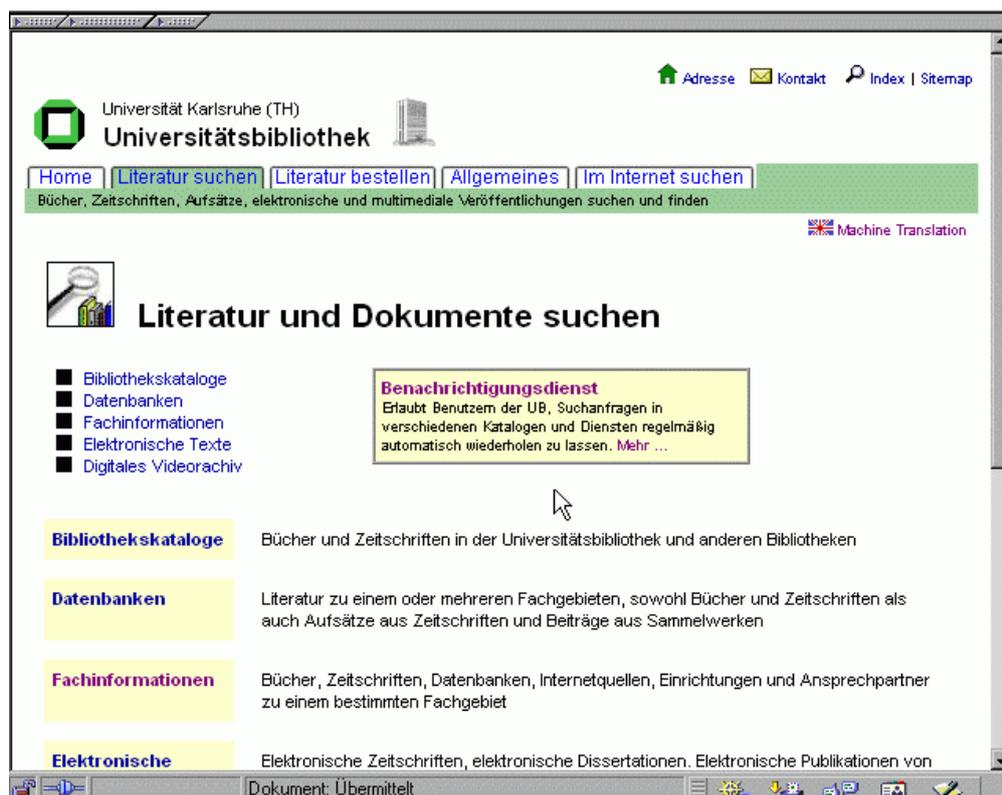


Abbildung 1: Deutsche Originalseite

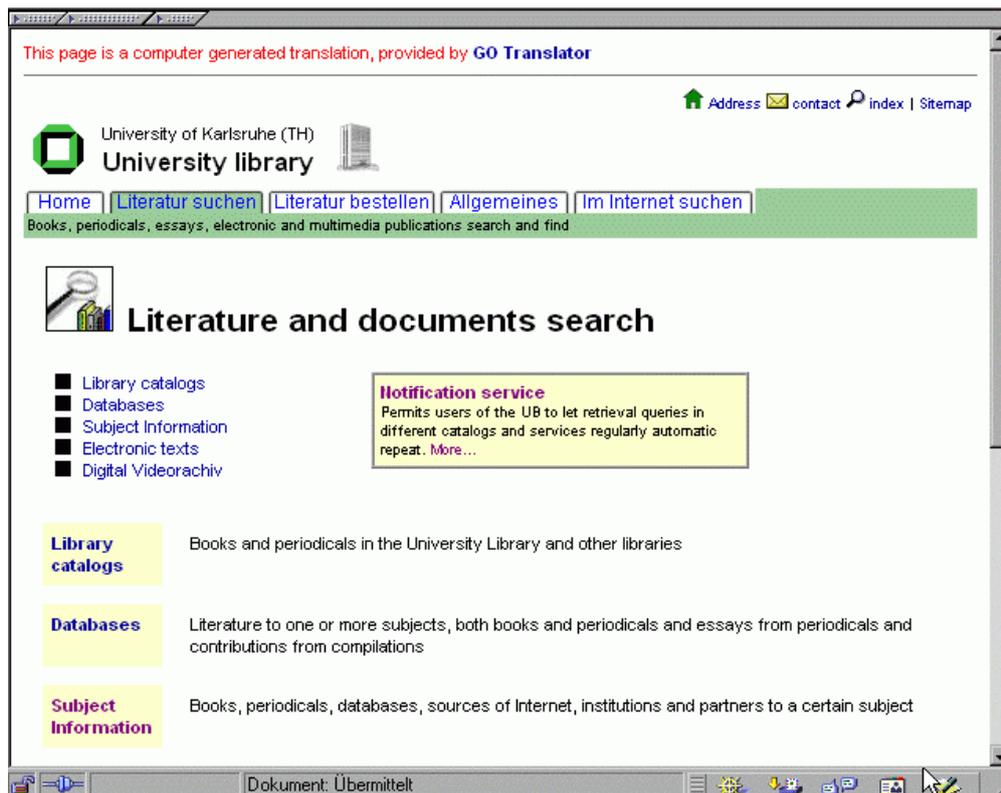


Abbildung 2: Maschinell übersetzte Seite

## Probleme der maschinellen Übersetzung

Übersetzt man die Webseiten der Universitätsbibliothek Karlsruhe in einem ersten Durchgang mit dem Go-Translator, so ergeben sich folgende Ergebnisse: Fehlerhafte Übersetzungen treten wie bei jeder maschinellen Übersetzung natürlich auch beim Go-Translator auf. Das generierte Englisch entspricht häufig nicht dem üblichen Sprachgebrauch und ist mit Sicherheit kein Oxford English". In ca. 90% der Fälle ist die Übersetzung aber so, dass der Sinn verständlich wiedergegeben wird. Wenn man damit leben kann, so können die Ergebnisse als brauchbar beurteilt werden. Der Go Translator arbeitet so, dass Begriffe, die nicht bekannt sind, als deutsche Ausdrücke stehen bleiben.

Wirkliche Probleme treten vor allem in folgenden Fällen auf:

### a. Komplizierte Satzkonstruktionen

Komplexe Sätze wie "mit Ausnahme der vor 1964 erfassten, finden Sie alle in der Universitätsbibliothek vorhandenen Lehrbücher mittels der im 1. Stock befindlichen Katalogterminals" bereiten der automatischen Übersetzung Schwierigkeiten. Dies gilt vor allem für die für die deutsche Sprache typischen geschachtelten Nebensatz- und Partizipialkonstruktionen. Diese Konstrukte sind aber für Webseiten ohnehin nicht geeignet, da Webinformationen möglichst schnell erfassbar sein sollten, was man am besten mit kurzen, einfachen Sätzen erreicht.

Die Übersetzungsleistung kann daher deutlich verbessert werden, wenn man auf komplexe Satzkonstruktionen verzichtet. Dadurch gewinnen häufig auch die deutschen Seiten an Qualität und werden leichter verständlich.

### b. Eigennamen und Fachtermini

Eigennamen bereiten naturgemäß erhebliche Probleme, da zum Beispiel "Zimmermann" mit "Carpenter" oder "Tangen" mit "seaweeds" übersetzt wird. Das selbe gilt für Fachbegriffe wie "Fachinformation" ("drawer information") oder "Fernleihe" ("Remoteborrow"). Diese Ergebnisse sind unbefriedigend und wurden als nicht akzeptabel angesehen.

### **c.) Homonyme**

Auch Homonyme sind problematisch, wie im obigen Beispiel der "1. Stock". In der Regel kann man einen alternativen Begriff benutzen, der eindeutig ist. Die falsche Übersetzung des "1. Stock" als "stick" wird vermieden, wenn man "1. Stockwerk" schreibt.

### **d.) Übersetzung von Folgeseiten**

Standardmäßig arbeitet der Go-Translator so, dass eine Redirektion auf alle in der zu übersetzenden Seite vorhandenen Links angewendet wird. D.h., wenn man einen Link anklickt, wird die Folgeseite ebenfalls übersetzt, die dort verlinkten Seiten ebenfalls und so weiter ad infinitum. Die Übersetzung schließt auch dynamische Seiten ein. Dieses Verfahren mag für viele Fälle die richtige Lösung sein, für die Bibliothek hat sie den gravierenden Nachteil, dass ausgehend von der Suchseite des Online-Kataloges, auch alle Trefferlisten und Titelaufnahmen übersetzt werden. Dies würde zur erheblichen Verwirrung der Benutzer führen und ist daher nicht akzeptabel.

## **3. Kombination von Go-Translator und Skript**

Diese Schwierigkeiten haben dazu geführt, dass die automatische Übersetzung als Stand-Alone-System nicht als ausreichend angesehen wurde. Der Go-Translator wurde daher durch ein Skript ergänzt, das bestimmte Modifikationen an den vom Go-Translator erzeugten Seiten durchführt. Insbesondere wird die Redirektion von Folgeseiten rückgängig gemacht und anhand einer Wortliste werden Wortfolgen durch andere ersetzt, z.B. "drawer information" durch "subject information". Damit können die meisten Probleme mit Fachtermini und Eigennamen abgefangen werden. Die Wortliste der UB Karlsruhe umfasst derzeit knapp 1000 Einträge.

Zudem hat es sich als sinnvoll erwiesen, die übersetzten Seiten auf dem eigenen Webserver zu cachen, um die Performance zu verbessern. Das Skript wird durch Klick auf den Link "Machine translation" gestartet. Es überprüft zunächst, ob bereits eine gecachte Version existiert und greift nur dann auf den Go-Translator zu, wenn diese nicht vorhanden oder älter ist als die deutsche Originalseite.

Zudem wird aus Performance- und ästhetischen Gründen bei den vom Go-Translator erzeugten Seiten die eingblendete Werbung ausgefiltert und statt dessen ein Hinweis "This page is a computer generated translation, provided by GO Translator" ausgegeben. Dieser Hinweis soll auch deutlich machen, dass es sich um eine maschinelle Übersetzung handelt.

Das Skript ist in Python geschrieben, einer leistungsfähigen objektorientierten Skriptsprache, die auf vordefinierte Klassen zur Bearbeitung von Web-Seiten zurückgreift. Die Webseiten der UB werden jetzt nach und nach durch diesen Link ergänzt. Anzumerken ist noch, dass die englische Version der UB-Homepage aus technischen Gründen leider nicht per Skript erzeugt werden kann sondern von Hand aktualisiert werden muss.

## **4. Fazit**

Die automatische Übersetzung in Kombination hat sich als brauchbare Lösung erwiesen. Hauptvorteil ist, dass kein Pflegeaufwand für die Übersetzungen notwendig ist. Lediglich die Wortliste muss erstellt und gepflegt werden. Bei der Erstellung der deutschen Seiten ist zudem eine gewisse Sprachdisziplin einzuhalten. Durch das Caching ist die Performance sehr gut; zudem ist damit auch eine statistische Erfassung der Zugriffe auf diese Seiten möglich. Wenn man bereit ist, das Sprachniveau der maschinell erzeugten Seiten zu akzeptieren, stellt diese Vorgehensweise eine sehr rationelle Methode dar, das Webangebot einer Bibliothek zu internationalisieren.

---

## Zum Autor

**Dr. Michael Mönnich** ist Fachreferent für Informatik der Universitätsbibliothek Karlsruhe und leitet die Abteilung für das Bibliothekssystem

Universitätsbibliothek Karlsruhe  
Postfach 6920  
D-46049 Karlsruhe  
E-Mail: [moennich@ubka.uni-karlsruhe.de](mailto:moennich@ubka.uni-karlsruhe.de)

