

Einsatz der Transliterierung in der multilingualen Recherche

Zusammenfassung: Das Internet wird als ein internationaler Informationsraum, an dem jeder partizipieren und von dem jeder profitieren kann, angesehen. Es wurde untersucht, inwieweit im Internet verfügbare Suchmaschinen, Online-Wörterbücher und OPAC Systeme ihre Dienste einem internationalen Nutzerkreis zugänglich machen. Der hier untersuchte Aspekt betrifft die multilinguale Recherche, bei der eine Transliterierung von sprachabhängigen Sonderzeichen erforderlich ist. Die Frage, die es zu beantworten galt, war, wie wird Transliterierung eingesetzt und welche Qualität die Ergebnisse haben, wenn Nutzer nur einen eingeschränkten Zeichensatz verwenden und Sonderzeichen transliterieren (müssen). Die Untersuchungen haben gezeigt, dass Transliterierung, sogar innerhalb einer Anwendung, uneinheitlich eingesetzt wird und damit auch die Qualität der erzielten Rechercheergebnisse sehr unterschiedlich ist.

1 Einführung

Im Rahmen der Internationalisierung besteht für Nutzer die Notwendigkeit, gleichlautende oder ähnliche Information aus verschiedenen Sprachräumen zu erhalten. Um diese Funktionalität zu gewährleisten ist es notwendig, den für die Recherche verwendeten Suchraum auf unterschiedliche Sprachräume zu erweitern und den Suchbegriff diesen Sprachräumen anzupassen. Rein informationstechnisch gesehen, stellt es kein Problem dar, mehrere Sprachräume mit einzubeziehen. Der zweite Teil der Aufgabe, den Suchbegriff den einzelnen Sprachräumen anzupassen, um so eine bessere Qualität der Rechercheergebnisse zu erzielen, erscheint umso komplexer.

Diese Aufgabenstellung kann unter verschiedenen Gesichtspunkten betrachtet werden. Zunächst stellt sich die Frage, welche Recherchemöglichkeiten bieten sich Nutzern, die außerhalb ihres angestammten Sprachraums Informationen suchen, dafür aber weder die technische Unterstützung zur Eingabe von sprachabhängigen Sonderzeichen noch die notwendigen speziellen Kenntnisse zur alternativen Eingabe von Sonderzeichen, z.B. über Codes, haben. Der vorliegende Beitrag wird vornehmlich diese Aufgabenstellung untersuchen. Man kann die Fragestellung aber auch dahingehend erweitern, dass Rechercheunterstützungsmöglichkeiten, für den Fall, dass der Begriff nur in der Benutzersprache vorliegt, untersucht werden.

Nach dem vorliegenden einführenden Abschnitt geht der Beitrag zunächst auf den Begriff der Multilingualität, danach auf Aspekte der Transliterierung und des Textretrievals ein. Dabei wird eine Übersicht der besonderen Aspekte, die sich im Zusam-

menspiel mit dem Internet ergeben, herausgearbeitet. Im letzten Teil des Beitrages werden die durchgeführten Untersuchungen und die Ergebnisse vorgestellt.

2 Multilingualität im Internet

Das gängige Verständnis von Multilingualität im Internet bedeutet, dass eine Webseite in mehreren Sprachversionen zur Verfügung gestellt wird. Von einer (meist in Englisch gehaltenen) Startseite aus, gibt es die Möglichkeit sich auf die einzelnen Sprachversionen durchzuklicken (vergleiche Melnick). Dies kommt Internetnutzern, die sich in einem einzigen Sprachraum bewegen, entgegen. Internationale Suchmaschinen haben sich darauf eingestellt und bieten verstärkt die Möglichkeit, die Suche in vordefinierten Sprachräumen durchzuführen. Die Anzahl der von Suchmaschinen unterstützten Sprachen ist unterschiedlich und wächst kontinuierlich.

In einer typischen Anwendungssituation gehen Nutzer üblicherweise von wenigen Begriffen aus, z.B., in einer e-commerce Anwendung oder bei einer Literaturrecherche sind Produzenten-, Autoren- oder Produktnamen bekannt. Eine Recherche, die einer erweiterten Bedeutung des Begriffs Multilingualität genügt, sollte Nutzern die Möglichkeit bieten, die Suche *sprachübergreifend* durchzuführen und *in einem* Recherche-schritt, Seiten mit dem gesuchten Begriff zu finden, unabhängig von deren Sprache.

Alltäglich, aber in diesem Fall erschwerend, ist die Tatsache, dass Computer eine dem jeweiligen Sprachraum angepasste Tastatur haben. Somit benötigt ein deutscher Nutzer Spezialkenntnisse, um ein französisches „ç“ einzugeben. Ohne Spezialkenntnisse wird der Nutzer einen Ausweg suchen, das nicht vorhandene Zeichen wird transliteriert und es wird, zum Beispiel, statt eines ç ein c eingegeben. Die Frage ist, inwieweit Anwendungen diese typische Vorgehensweise berücksichtigen und Nutzer dabei unterstützen. In der Praxis gibt es dafür verschiedene Lösungsansätze:

- Alternative Eingaben. Die fehlenden Sonderzeichen werden zur Verfügung gestellt, zum Beispiel als Schaltflächen auf der Nutzeroberfläche, so dass, sie durch einfaches Klicken eingefügt werden können (vergleiche Zeus).
- Selbstdefinierte Ersatzdarstellungen. Für Sonderzeichen wird eine ASCII Ersatzdarstellung definiert. Nutzer können bei Bedarf diese Ersatzdarstellungen verwenden (vergleiche TITUS).
- Transliteration. Statt Sonderzeichen werden durch internationale Normen festgelegte, sprachabhängige Ersatzdarstellungen verwendet.

Im vorliegenden Beitrag werden wir uns der Frage widmen, wie das Problem der fehlenden Sonderzeichen mit auf Transliteration basierenden Lösungen angegangen wird.

3 Transliterierung

Die Transliterierung wurde ursprünglich entwickelt, um Zeichen von einem Zeichensatz in einen anderen umzuwandeln, z.B. von Kyrillisch nach Latein. Die Transliterierung hat sich mit verstärktem Einsatz der Multilingualität in text- und sprachverarbeitenden Anwendungen zu einem flexiblen Werkzeug, welches viele Aufgaben wahrnimmt, entwickelt. Bibliotheken zur Entwicklung von Anwendungen für den internationalen Markt wie die ICU Bibliothek von IBM (vergleiche Transliteration) enthalten vielfältig einsetzbare, multifunktionale Transliterierungsfunktionen, die neben den durch ISO-Normen festgelegten Transformationen zwischen Zeichensätzen auch eine Reihe von sprachabhängigen Zeichenvereinfachungen implementieren.

Die Komplexität einer Recherche, aber auch die erzielte Funktionalität und die zu erwartende Güte der Ergebnisse, wachsen mit der Berücksichtigung von zusätzlichen Parametern. Beim Einsatz von Transliterierung im Rahmen der multilingualen Recherche können verschiedene Aspekte Berücksichtigung finden, die Nutzersprache dient in diesen Anwendungen als Quellsprache, Zielsprachen sind die Sprachen der zu berücksichtigenden Suchräume, d.h., die Sprachen der zu recherchierenden Dokumente.

- Keine Transliterierung. Hierbei wird die Recherche mit dem eingegebenen Suchstring, ohne Berücksichtigung von Transliterierung, durchgeführt. Zum Beispiel, solch eine Recherche würde im deutschen Sprachraum bei Eingabe der Suchbegriffe „Münchhausen“ und „Muenchhausen“ nur Dokumente berücksichtigen, die das Wort in der jeweiligen Schreibweise enthalten.
- Quellsprachenabhängige Transliterierung. Bei der Recherche werden alle Schreibvarianten, die durch Transliterierung der Eingabe in der Quellsprache erhalten werden, berücksichtigt. Obiges Beispiel: die Recherche berücksichtigt sowohl Dokumente mit „Münchhausen“ als auch solche mit „Muenchhausen“. Um den speziellen Bedürfnissen der multilingualen Internetrecherche Rechnung zu tragen, muss dabei auch die Möglichkeit in Betracht gezogen werden, dass der Nutzer die Transliterierung bereits durchgeführt und „Muenchhausen“ eingegeben hat. Das Ergebnis muss identisch mit dem der Recherche ohne Transliterierung sein.
- Zielsprachenabhängige Transliterierung. Wir gehen davon aus, dass sowohl die Quellsprache als auch die Sprache der zu recherchierenden Dokumente bekannt sind. In diesem Falle erfolgt die Bearbeitung der Eingabe unter Berücksichtigung der Transliterierungsregeln aus der Quellsprache in die Zielsprache(n). Zum Beispiel, bei Eingabe von „Хрущев“ (Russisch) ist eine oder mehrere der folgenden Transliterierungsvarianten zu wählen: „Chruschtschow“ (Deutsch), „Khrushchev“ (Englisch) oder „Khrouchtchev“ (Französisch), wenn

Deutsch, Englisch und Französisch als Zielsprachen von vornherein feststehen. Die Recherche erfolgt simultan mit allen obigen Transliterationsvarianten.

- Quell- und zielsprachenabhängige Transliteration. Die umfangreichste Recherche kann durchgeführt werden, wenn Informationen über Quell- und Zielsprache vorliegen. Zum Beispiel, bei Eingabe „Chruschtschow“ (Deutsch) erkennt das System, dass die dazugehörigen zielsprachabhängigen Varianten „Хрущев“ (Russisch), „Khrushchev“ (Englisch) oder „Khrouchtchev“ (Französisch) lauten und führt mit diesen Suchwörtern die Recherche simultan in dem durch die Zielsprachen aufgespannten Suchraum durch.

4 Textretrieval

Die Transliteration kann nicht für sich alleine untersucht werden. Erst durch den sich daran anschließenden Einsatz vom Textretrieval werden die Rechercheergebnisse den Nutzern zugänglich gemacht.

Textretrieval, die Wiedergewinnung von maschinell gespeichertem Text, beeinflusst aber auch die erzielten Ergebnisse. Die zugrunde liegenden Algorithmen zur Berechnung der Übereinstimmung zwischen Suchbegriff und zu recherchierende Dokumente sind sehr unterschiedlich und erlauben die Wiedergewinnung der Dokumente auf unterschiedlichen Niveaus. In diesem Beitrag werden die im Einzelnen verwendeten Algorithmen, ihre Komplexität oder Laufzeit nicht näher betrachtet, die Unterscheidung wird ausschließlich anhand der erzielten Ergebnisse gemacht. Beim Textretrieval werden, im Allgemeinen, folgende Möglichkeiten unterschieden:

- Exakte Suche (exact match). Diese Retrievaltechnik erfordert eine exakte Übereinstimmung des Suchbegriffs mit dem Wort im Dokument. Die Übereinstimmung bezieht sich, beispielsweise, auch auf Groß-/Kleinschreibung.
- Ungenaue Suche (partial match). Dabei werden teilweise Übereinstimmungen zwischen Suchbegriff und Dokument bewertet. In diesem Bereich sind, z.B., die Normalisierung von Groß-/Kleinschreibung und insbesondere auch der Einsatz der Transliteration anzusetzen.
- Diffuse Suche (fuzzy matching), die Suchtechnik, mit der nicht nur das genaue Äquivalent des gesuchten Begriffs gefunden werden kann, sondern auch alle Elemente, die eine gewisse Ähnlichkeit aufweisen. Dabei können verschiedenartige Suchfunktionen eingesetzt werden, z.B., Grundformenrecherche, mit der Wörter unabhängig von ihrer Flexionsform gefunden werden (Suche mit „Apfel“, Fundstelle mit „Äpfel“) oder phonetische Recherche, mit der phonetisch gleichlautende oder ähnlich klingende Wörter gefunden werden.

5 Untersuchungsprämissen und Ergebnisse

Bei den durchgeführten Untersuchungen haben wir von Anfang an angenommen, dass die betrachteten Systeme eine ungenaue Suche (partial match) beim Textretrieval verwenden. Die Benutzerinformation, die zum Teil von den Systemen zur Verfügung gestellt wird, sowie die alltäglichen Erfahrungen zeigen, dass Normalisierung von Groß- und Kleinschreibung zum Standard der überwiegenden Mehrheit der Retrievalsysteme gehören. Mehr noch, es kommt sogar vor, dass die Berücksichtigung von Groß-/Kleinschreibung als eine zusätzliche Funktionalität von Systemen zur Verfügung gestellt wird (die von uns untersuchten Systeme gehören nicht dazu).

Fuzzy matching, welches auf der Skala der Retrievalmöglichkeiten am anderen Komplexitätssende angesiedelt ist, erfordert sprachabhängige Retrievalalgorithmen. Die Komplexität der dabei verwendeten Algorithmen ließ von Anfang an vermuten, dass fuzzy matching nicht oder nur selten eingesetzt wird, Tests haben dies bestätigt.

Die Untersuchungen haben sich auf die Frage konzentriert, wie Transliterierung, ein durch Normen spezifiziertes und algorithmisch gut fassbares Verfahren, für Textretrieval eingesetzt wird. Transliterierung kann als Recherchewerkzeug unterschiedlich parametrisiert werden. Bei den Untersuchungen wurde davon ausgegangen, dass die vorkommenden Transliterierungen einer Vereinfachung der Eingabe dienen, d.h., Weglassen von Akzenten und Transliterierung von Sonderzeichen. Die zentrale Frage war, ob und wie die Information über Quell- und Zielsprache dabei eingesetzt wird.

- Zunächst wurde mit Deutsch, Französisch und Portugiesisch eine Gruppe von Sprachen, die das lateinische Alphabet verwenden, untersucht. Die Besonderheiten der Sprachen wurden anhand der enthaltenen diakritischen Zeichen und Sonderzeichen dargestellt. Der Schwerpunkt lag auf der Untersuchung der Transliterierung von Zeichen, die in anderen Sprachen nicht vorkommen. In den zu untersuchenden Systemen wurde zunächst ein Testkorpus eingegeben, der das jeweilige Sonderzeichen, z.B., den Umlaut „ü“, enthält, dann derselbe Testkorpus mit dem transliterierten Zeichen, in unserem Beispiel „ue“ für die deutsche oder „u“ für die französische Sprache.
- Außerdem wurde mit Russisch eine Sprache, die kein lateinisches Alphabet verwendet, untersucht. Russisch enthält diakritische Zeichen und Sonderzeichen, ё, й, ъ, ѓ, etc. Für Russisch wurde, wie bei den anderen Sprachen, anhand eines Testkorpus untersucht, wie die Ergebnisse der untersuchten Systeme mit und ohne Verwendung der Sonderzeichen aussehen. Darüber hinaus ist beim Russischen die Verwendung der Transliterierung ins lateinische Alphabet vom besonderen Interesse.

Die Untersuchungen umfassten Online-Wörterbücher, Web-Suchmaschinen und O-PAC-Systeme. Die Ergebnisse werden nachfolgend exemplarisch dargestellt.

5.1 Online-Wörterbücher

Wörterbuchanwendungen haben per Definition eine Quellsprache; dadurch ist es wenig sinnvoll, einen Textkorpus mit Sonderzeichen aus anderen Sprachen zu betrachten. Transliterierungen sind nur quellsprachenabhängig und dienen der Vereinfachung der Eingabe, bei Sprachen mit nicht lateinischem Alphabeten wird zusätzlich die Transliterierung in das lateinische Alphabet (einfache ASCII Transliterierung) betrachtet. Nachfolgend die Zusammenfassung der Ergebnisse, die recht unterschiedlich ausfallen. Details können den nachfolgenden Tabellen entnommen werden.

- Transliterierungen werden in den angebotenen Sprachversionen korrekt bearbeitet (Leo, Linguadict, Pauker).
- Transliterierungen werden zum Teil korrekt bearbeitet, eine Systematik konnte nicht erkannt werden. Es ist zu vermuten, dass die Streuung der Ergebnisse durch Fehler in der Datenerfassung verursacht wird (Eurodicautom).
- Granddictionnaire, Multitran, Rambler, Travlang und Zeus wenden keine Transliterierung an. Granddictionnaire und Zeus stellen alternative Eingabemöglichkeiten zur Verfügung, Multitran bietet eine online Nutzerunterstützung an, d.h., bei fehlenden diakritischen Zeichen wird die Schreibweise mit diakritischen Zeichen (erkannt mittels Transliterierung?) als Eingabealternative angeboten.

Testkorpus enthält	Ergebnisse enthalten
Eurodicautom (Deutsches Wörterbuch)	
Deutsche diakritische Zeichen (ä, ö, ü)	Suchbegriff mit diakritischen Zeichen und mit Transliterierung nach deutschen Regeln
Transliterierte deutsche diakritische Zeichen (ae, oe, ue)	Suchbegriff mit und ohne Translit.; Ergebnis stimmt nur zum Teil mit dem obigen überein.
Deutsche Sonderzeichen (ß)	Suchbegriff mit diakritischen Zeichen und mit nach deutschen Regeln transliterierten diakritischen Zeichen; Suchergebnisse identisch.
Transliterierte deutsche Sonderzeichen (ß -> ss)	
Leo (Deutsches Wörterbuch)	
Deutsche diakritische Zeichen mit / ohne Transliterierung	Suchbegriff immer mit diakritischen Zeichen; Suchergebnisse in beiden Fällen identisch
Deutsche Sonderzeichen mit / ohne Transliterierung	Suchbegriff mit Sonderzeichen; Suchergebnisse in beiden Fällen identisch
Linguadict (Deutsches Wörterbuch)	
Deutsche diakritische Zeichen mit / ohne Transliterierung	Suchbegriff mit diakritischen Zeichen; Suchergebnisse in beiden Fällen identisch
Deutsche Sonderzeichen mit / ohne Transliterierung	Suchbegriff mit Sonderzeichen; Suchergebnisse in beiden Fällen identisch

Eurodicautom (Französisches Wörterbuch)

Franz. Sonderzeichen mit / ohne Translit. (ç -> c, œ-> oe)	Suchbegriff mit Sonderzeichen; Suchergebnisse in beiden Fällen identisch
Französische diakritische Zeichen (à, é, ê, è, ï, ô, etc.) mit / ohne Transliteration	Suchbegriff mit diakritischen Zeichen, zusätzlich Ergebnisse zu Begriffen mit identischer Translit. Ergebnisse in allen Fällen identisch.

Granddictionnaire (Französisches Wörterbuch)

Franz. diakritische Zeichen	Suchbegriff mit diakritischen Zeichen
Translit. franz. diakr. Zeichen	Keine Treffer
Französische Sonderzeichen	Keine Treffer
Translit. Sonderzeich. (ç -> c)	Keine Treffer
Transliteriertes französisches Sonderzeichen (œ -> oe)	Suchbegriff mit translit. Sonderzeichen; System kennt kein œ, es wird transliteriert.

Linguadict (Französisches Wörterbuch)

Französische Sonderzeichen	Suchbegriff mit Sonderzeichen
Translit. franz. Sonderzeichen	Suchbegriff
Franz. diakritische Zeichen	Suchbegriff mit diakritischen Zeichen

Transliterierte französische diakritische Zeichen	Suchbegriff, in seltenen Fällen auch mit Sonderzeichen, keine Systematik erkennbar.
---	---

Zeus (Französisches Wörterbuch)

Französische Sonderzeichen	Suchbegriff mit Sonderzeichen
Translit. franz. Sonderzeichen	Keine Treffer
Franz. diakritische Zeichen	Suchbegriff mit diakritischen Zeichen
Translit. franz. diakr. Zeichen	Keine Treffer

Eurodicautom (Portugiesisches Wörterbuch)

Portugiesische diakritische Zeichen und Sonderzeichen mit / ohne Transliterierung	Suchbegriff mit diakritischen Zeichen / Sonderzeichen; Ergebnisse mit / ohne Transliterierung identisch.
---	---

Pauker (Portugiesisches Wörterbuch)

Portug. diakrit. Zeichen, Sonderzeichen mit / ohne Translit.	Suchbegriff mit diakrit. Zeichen, Sonderzeichen; Ergebnisse mit / ohne Transliterierung identisch.
--	---

 Travlang (Portugiesisches Wörterbuch)

Portugiesische diakritische Zeichen oder Sonderzeichen	Suchbegriff mit diakritischen Zeichen oder Sonderzeichen.
Translit. portug. diakritische Zeichen oder Sonderzeichen	keine Treffer

 Multitran (Russisches Wörterbuch)

Kyrillische diakrit. Zeichen	Suchbegriff;
Kyrillische Zeichen, diakritische Zeichen transliteriert	System schlägt alternativ den Suchbegriff mit kyrillischen diakritischen Zeichen vor;
Kyrillisch nach ASCII translit.	keine Treffer

 Rambler (Russisches Wörterbuch)

Kyrillische diakrit. Zeichen	Suchbegriff;
Kyrillisch, diakrit. Z. translit.	keine Treffer
Kyrillisch nach ASCII translit.	keine Treffer

5.2 Web-Suchmaschinen

Bei Web-Suchmaschinen wurden in der Regel mehrere Sprachversionen untersucht. Für den russischen Sprachraum wurde eine russische Suchmaschine (Aport) zusätzlich hinzugezogen. Die Analyse der Ergebnisse zeigt, dass Suchmaschinen sich auf quellsprachenabhängige Transliterierung beschränken. Die Qualität der Ergebnisse ist unterschiedlich, sogar innerhalb der einzelnen Suchmaschinen, sie wird wesentlich davon beeinflusst, welche Quellsprachen berücksichtigt werden.

- Innerhalb der Suchmaschinen zeigte Google, aus Nutzersicht, das beste Verhalten. Diakritische Zeichen und Sonderzeichen aus lateinischen Alphabeten wurden transliteriert und die Rechercheergebnisse sind für alle Sprachversionen identisch. Eingaben aus dem kyrillischen Alphabet wurden, auch in der russischen Version, nicht transliteriert, es wurde jeweils nur mit dem Suchbegriff in der eingegebenen Form recherchiert.
- Abacho ist eine Suchmaschine, die quellsprachenabhängige Transliterierung konsequent einsetzt. Die Quellsprache wird immer mit der Sprache der verwendeten Version gleichgesetzt und nur Sonderzeichen, die in dieser Sprache vorkommen, werden auch vollständig transliteriert (eine Ausnahme bildet die französische œ Ligatur). Abacho hat keine russische Version und verarbeitet auch keine kyrillischen Zeichen.
- Altavista verwendet keine Transliterierung, obwohl es, z.B., bei Verwendung oder Transliterierung von diakritischen Zeichen, Ergebnisse liefert, die beide Zeichenarten enthalten. Beim näheren Betrachten stellt sich aber heraus, dass die Ergebnisse sehr unterschiedlich sind, eine Systematik ist nicht erkennbar, es ist unklar, wie die Ergebnisse zustande kommen.
- Aport, die untersuchte Suchmaschine aus dem russischen Sprachraum, führt ebenfalls keine Transliterierung und keine Vereinfachung durch.

	Testkorpus	Rechercheergebnis enthält
Abacho		
.de	Deutsche Sonderzeichen und diakritische Zeichen / Transliteration	Suchbegriff mit und ohne Transliteration; Identische Treffermengen.
	Franz. / portug. diakrit. Zeichen	Suchbegriff;

	Transliterierte französische / portugiesische diakritische Zeichen	Begriff mit und ohne diakrit. Zeichen; Ergeb. unterscheiden sich von obigen.
	Französische Sonderzeichen	Suchbegriff; Sehr wenige Treffer
	Transliterierte französische Sonderzeichen	Suchbegriff mit und ohne Translit.; Ergebnis unterscheidet sich vom obigen.
	Kyrillische diakritische Zeichen	Suchbegriff;
	Kyrillische Zeichen, diakritische Zeichen transliteriert	Suchbegriff; Treffermenge unterscheidet sich von der obigen Suche.
	Kyrillisch nach ASCII transliteriert	Keine Treffer.
.fr	Deutsche diakritische Zeichen	Suchbegriff;
	Nach deutschen Regeln transliterierte deutsche diakritische Zeichen	Suchbegriff; Treffermenge unterscheidet sich von der obigen Suche.
	Deutsche diakritische Zeichen, französisch transliteriert	Suchbegriff; Treffermenge unterscheidet sich von beiden obigen.
	Deutsche Sonderzeichen	Suchbegriff mit und ohne Translit.
	Nach deutschen Regeln transliterierte deutsche Sonderzeichen	Suchbegriff mit und ohne Translit.; identisch zur obigen Recherche
	Französische / portugiesische diakritische Zeichen oder Translit.	Suchbegriff mit und ohne Translit.; Identische Ergebnisse.

.fr	Transliterierte französische Sonderzeichen (ç -> c, œ -> oe)	Suchbegriffe mit „c“ führen zu Ergebnissen mit und ohne Translit.; Suchbegriffe mit „oe“ führen nur zu Ergebnissen mit Transliterierung;
	Kyrillische diakritische Zeichen	keine Treffer
	Kyrillisch, diakrit. Zeichen translit.	keine Treffer
	Kyrillisch nach ASCII transliteriert	keine Treffer

Altavista

.de	Deutsche diakritische Zeichen	Suchbegriff;
.fr		
.pt	Transliterierte deutsche diakritische Zeichen	Suchbegriff; Treffermengen unterscheiden sich von den obigen.
	Deutsche Sonderzeichen	Suchbegriff;
.de	Translit. deutsche Sonderzeichen	Suchbegriff mit und ohne Translit.; Ergebnis unterscheidet sich vom obigen.
.fr		
.pt	Translit. deutsche Sonderzeichen	Suchbegriff; Treffermenge unterscheidet sich von der obigen.
.de	Französische / portugiesische diakritische Zeichen und Sonderzeichen	Suchbegriff;
.fr		
.pt	Translit. franz. / portug. diakritische Zeichen und Sonderzeichen	Suchbegriff mit und ohne Translit.

	Kyrillische diakritische Zeichen	Suchbegriff;
	Kyrillisch, diakrit. Zeichen translit.	Keine Treffer;
	Kyrillisch nach ASCII transliteriert	Wenige bzw. keine Treffer;

Treffermengen von einer Sprachversion zur anderen unterschiedlich.

Google

.de .fr .pt	Deutsche / franz. / portug. diakrit. Zeichen oder Sonderzeichen mit / ohne Transliterierung nach Regeln der Sprache des Zeichens	Suchbegriff mit diakrit. Zeichen / Sonderzeichen und mit Translit. nach Regeln der Sprache des Zeichens; Ergebnisse mit / ohne Translit. identisch.
	Kyrillische diakritische Zeichen	Suchbegriff;
	Kyrillische Zeichen, diakritische Zeichen transliteriert	Suchbegriff; Ergebnis unterscheidet sich vom vorherigen
	Kyrillisch nach ASCII transliteriert	Suchbegriff; Ergebnis unterscheidet sich von den obigen;
Obige Ergebnisse sind in allen Sprachversionen der Suchmaschine identisch.		
.fr .pt	Deutsche diakritische Zeichen, Sonderzeichen nach Regeln der Sprache der Suchmaschine translit.	Nachfrage zur Orthographie (Vorschlag Schreibweise mit Sonderzeichen), danach Suche wie oben.

.ru	Deutsche / französische / portugiesische diakritische Zeichen oder Sonderzeichen	Suchbegriff; Ergebnisse unterscheiden sich wesentlich von den Ergebnissen der .de, .fr und .pt Versionen.
.ru	Deutsche / franz. / portug. diakrit. Zeichen oder Sonderzeichen, translit. nach Sprache des Zeichens	Suchbegriff; Ergebnisse mit / ohne Transliterierung unterschiedlich.
	Deutsche diakritische Zeichen, Sonderzeichen translit. nach Regeln der Sprache der Suchmaschine	Suchbegriff transliteriert ; Ergebnis unterscheidet sich von den Ergebnissen der beiden obigen Recherchen
Rechercheergebnisse der .ru Version unterscheiden sich wesentlich von den Ergebnissen der .de, .fr und .pt Versionen.		
	Kyrillische diakrit. Zeichen	Suchbegriff
	Kyrillische Zeichen, diakritische Zeichen transliteriert	Suchbegriff; Ergebnisse mit / ohne Transliterierung unterschiedlich.
	Kyrillisch nach ASCII transliteriert	Suchbegriff; Ergebnis unterscheidet sich von den obigen Ergebnissen.

Aport

.ru	Kyrillische diakritische Zeichen	Suchbegriff mit diakritischen Zeichen;
	Kyrillisch, diakrit. Zeichen translit.	Rechercheergebnisse identisch.
	Kyrillische Zeichen nach ASCII transliteriert	Suchbegriff; Ergebnis unterscheidet sich von den obigen Ergebnissen.

5.3 OPAC-Systeme

Die untersuchten OPAC-Systeme sind an öffentlichen Bibliotheken angesiedelt, sie haben, genau wie Online-Wörterbücher, per Definition eine Quellsprache. Die Ergebnisse der Untersuchungen lassen sich folgendermaßen resümieren:

- Alle untersuchten Systeme haben Sonderzeichen und ASCII-Transliterationen der Quellsprache korrekt bearbeitet.
- Transliterationen von Sonderzeichen aus anderen Sprachen werden zum Teil akzeptiert, eine Systematik ist schwer erkennbar (Beispiel Behandlung des Zeichens ð im OPAC-Bibliotheksverbund Bayern).
- Kyrillische Zeichen werden nur von russischen OPAC-Systemen verarbeitet, die anderen Systeme akzeptieren keine kyrillischen Eingaben.
- Untersuchungen mit transliterierten kyrillischen Zeichen zeigen, dass die von Systemen verwendete Transliteration nicht immer konform mit internationalen Normen ist.

Testkorpus enthält	Ergebnis enthält
Saarländische Universitäts- und Landesbibliothek (UniSb)	
Deutsche diakritische Zeichen mit / ohne Transliteration.	Suchbegriff; Identische Treffermengen, unterschiedliche Auflistung der Ergebnisse.
De. Sonderz. mit/ohne Translit.	Suchbegriff; identische Treffermengen.
Französische / portugiesische diakritische Zeichen mit / ohne Translit.	Suchbegriff mit diakritischen Zeichen
Kyrillische diakritische Zeichen	Anfrage kann nicht bearbeitet werden.
Kyrillisch, diakrit. Zeichen translit.	Anfrage kann nicht bearbeitet werden.
Kyrillisch nach ASCII transliteriert	Suchbegriff transliteriert.

Bibliotheksverbund Bayern (BVB)

Deutsche diakritische Zeichen mit / ohne Transliterierung	Suchbegriff mit und ohne Transliterierung; Identische Treffermengen.
Deutsche Sonderzeichen mit / ohne Transliterierung	Suchbegriff mit und ohne Transliterierung; Identische Treffermengen.
Franz. / portug. diakrit. Zeichen	Suchbegriff
Transliterierte französische diakritische Zeichen	Suchbegriff mit diakritischem Zeichen; Treffermenge identisch zur obigen Suche.
Portugiesische diakritische Zeichen	Keine Treffer
Translit. port. diakrit. Zeich. (õ->o)	Suchbegriff mit diakritischen Zeichen
Kyrillische diakritische Zeichen	Anfrage kann nicht bearbeitet werden.
Kyrillisch, diakrit. Zeichen translit.	Anfrage kann nicht bearbeitet werden.
Kyrillisch nach ASCII transliteriert	Suchbegriff transliteriert.

Hessisches Bibliotheksinformationssystem, Lokalsystem Frankfurt/M (UniF)

Das System ermöglicht die Recherche mit Sonder- und diakritischen Zeichen, aber die Datensätze enthalten nur transliterierte Zeichen.

Deutsche diakritische Zeichen mit / ohne Transliteration	Suchbegriff mit transliterierten Zeichen. Identische Treffermengen
Deutsche Sonderzeichen mit / ohne Transliteration	Suchbegriff mit transliterierten Sonderzeichen; Identische Treffermengen.
Französische / portugiesische diakritische Zeichen mit / ohne Translit.	Suchbegriff mit transliterierten Sonderzeichen; Identische Treffermengen.
Kyrillische diakritische Zeichen	Keine Treffer
Kyrillisch, diakrit. Zeichen translit.	Keine Treffer
Kyrillisch nach ASCII transliteriert	Suchbegriff transliteriert.
Universitätsbibliothek Université de la Réunion (Absys Opac)	
Deutsche diakritische Zeichen	Suchbegriff;
Deutsche diakritische Zeichen französisch transliteriert (ö -> o)	Suchbegriff mit diakritischem Zeichen; Identische Treffermenge
Deutsche Sonderzeichen	Suchbegriff mit Sonderzeichen
Deutsche Sonderzeichen französisch transliteriert (ß-> s)	Suchbegriff mit Sonderzeichen; Rechercheergebnis identisch zur vorherigen Suche.
Deutsche diakritische Zeichen oder Sonderzeichen deutsch transliteriert	Keine Treffer

Franz. / portug. diakritische Zeichen mit / ohne Translit.	Suchbegriff mit und ohne diakritischen Zeichen; Identische Treffermengen.
Kyrillische diakritische Zeichen	Anfrage kann nicht bearbeitet werden.
Kyrillisch, diakrit. Zeichen translit.	Anfrage kann nicht bearbeitet werden.
Kyrillisch nach ASCII transliteriert	Suchbegriff transliteriert.

Portugiesische Nationalbibliothek (Porbase)

Deutsche diakritische Zeichen	Suchbegriff portugiesisch transliteriert;
Deutsche diakritische Zeichen portugiesisch transliteriert	Suchbegriff portugiesisch transliteriert; Identische Treffermenge.
Deutsche Sonderzeichen (ß)	Sonderzeichen wird nicht erkannt, z.B. „Reißner“ ≙ „Reiner“
Französische / portugiesische diakritische Zeichen mit / ohne Translit.	Suchbegriff mit / ohne Sonderzeichen, zusätzlich Begriffe mit identischer Translit.
Kyrillische diakritische Zeichen	Anfrage kann nicht bearbeitet werden.
Kyrillisch, diakrit. Zeichen translit.	Anfrage kann nicht bearbeitet werden.
Kyrillisch nach ASCII transliteriert	Suchbegriff transliteriert.

Russische Staatsbibliothek (RuSB)	
Deutsche / franz. / portug. Sonderzeichen / diakritische Zeichen	Suchbegriff
Sonderzeichen oder diakriti- sche Zeichen russisch trans- lit. (ö-> o)	Suchbegriff mit diakritischen Zei- chen;
Nach deutschen Regeln transliterierte deutsche Son- derzeichen	System schlägt alternative Schreib- weisen vor, darunter auch die russi- sche Translit.
Kyrillische diakritische Zei- chen	Suchbegriff;
Kyrillische Zeichen, diakriti- sche Zeichen transliteriert	System schlägt Alternativen vor, darunter auch die Form mit diakrit. Zeichen.
Kyrillische Zeichen nach AS- CII transliteriert	Suchbegriff kyrillisch und translit.; Ergebnisse unterscheiden sich von den obigen.

6 Zusammenfassung und weiterführende Arbeiten

Die Untersuchungen haben gezeigt, dass Transliterierung als Werkzeug in der multi-lingualen Recherche unterschiedlich umgesetzt wird. Die untersuchten Anwendungen setzen nur quellsprachenabhängige Transliterierung ein, bei der Umsetzung spielen Suchmaschinen, bedingt durch den herrschenden Wettbewerb, eine Vorreiterrolle.

Um umfassende Rechercheergebnisse zu erzielen, ist zukünftig der Einsatz der ziel-sprachenabhängigen Transliterierung unerlässlich. Solche Anwendungen werden ge-genwärtig in Systemen bei internationalen Institutionen verstärkt eingesetzt (siehe

ABEL, Inhaltsverzeichnis des EU-Amtsblattes oder CATEL, Katalog des Amtes für EU-Veröffentlichungen, vergleiche Eur-OP). Zielsprachenabhängige Transliterierung setzt umfangreiche Vorarbeiten bei der Indexierung voraus, insbesondere die Erstellung entsprechender Wörterbücher oder den Einsatz mehrsprachiger Thesauri. Die Untersuchung solche Systeme soll Gegenstand weiterführender Arbeiten sein.

Literatur

- Eur-OP: Amt für amtliche Veröffentlichungen der EU, <http://eur-op.eu.int/>
- Lavoie, Brian F. / O'Neill, Edward T. (1999): How "World Wide" Is the Web? Trends in the Internationalization of Web Sites. *Annual Review of OCLC Research 1999*, Dublin, OCLC Online Computer Library Center, <http://www.oclc.org/oclc/research/publications/review99/>
- Melnick, Jim (2000): Multilingual forms in Unicode. Developing Unicode-based global Software Applications. *developerWorks, IBM Open Source*, September 2000, <http://www-106.ibm.com/developerworks/unicode/library/os-mult.html?dwzone=unicode>
- Titus: Titus, Thesaurus Indogermanischer Text- und Sprachmaterialien, <http://titus.fkidg1.uni-frankfurt.de/indexd.htm>
- Transliteration (2002): *ICU – International Components for Unicode*, Version 2.2., Userguide, IBM, San José, August 2002, <http://oss.software.ibm.com/icu/userguide/>

Online-Wörterbücher, Web-Suchmaschinen, OPAC-Systeme

- | | |
|---------------------|---|
| Abacho : | http://www.abacho.com (.de, .fr) |
| Absys : | http://www.univ-reunion.fr/ |
| Altavista : | http://www.altavista.com (.de, .fr, .pt, .ru) |
| Aport : | http://sm.afort.ru |
| BVB : | http://www-opac.bib-bvb.de/ |
| Google : | http://www.google.com (.de, .fr, .pt, .ru) |
| Eurodicautom : | http://europa.eu.int/eurodicautom/ |
| Granddictionnaire : | http://www.granddictionnaire.com |
| Leo : | http://dict.leo.org |
| Linguadict : | http://www.linguatec.net/online/dict/ |
| Multitran : | http://www.multitran.ru |
| Pauker : | http://pauker.at |
| Porbase : | http://ipac.bn.pt/ipac-cgi/ipac.exe |
| Rambler: | http://www.rambler.ru |
| RuSB | http://www.rsl.ru/ |
| Travlang: | http://dictionaries.travlang.com |
| UniF : | http://webopac.server.uni-frankfurt.de/ |
| UniSb : | http://opac.sulb.uni-saarland.de/ |
| Zeus : | http://zeus.inalf.fr/tlf.htm |