

Monika Kuberek

**Volltextserver der Verbände –  
Integrierter Zugriff auf lizenzierte Zeitschriftenartikel  
und freie elektronische Dokumente  
unter Einsatz von Suchmaschinentechnologie**

**Gefördert  
von der Senatsverwaltung für Wissenschaft, Forschung und Kultur des Landes Berlin,  
vom Ministerium für Wissenschaft, Forschung und Kultur des Landes Brandenburg  
und von den Mitgliedsbibliotheken des KOBV**



---

# Volltextserver der Verbände – Integrierter Zugriff auf lizenzierte Zeitschriftenartikel und freie elektronische Dokumente unter Einsatz von Suchmaschinentechnologie

*Monika Kuberek*  
( *kuberek@zib.de* )

*Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)*  
*ZIB Report 07-46*  
*Dezember 2007*

## **Abstract**

Die Intention ist der kooperative Aufbau einer Infrastruktur durch die Bibliotheksverbände, um den Nutzern Volltext-Angebote dauerhaft und komfortabel zur Verfügung zu stellen: Zeitschriftenartikel und elektronische Dokumente werden mittels Suchmaschinentechnologie indiziert und unter Berücksichtigung von Zugriffsrechten zugänglich gemacht. Realisiert ist dies bereits im KOBV-Volltextserver, der seit Ende 2005 im Routinebetrieb läuft.

Vorstellbar ist ein überregionales Netz von Volltextservern der Verbände, die mittels Suchmaschinentechnologie indiziert und nahtlos in das regionale und lokale Literaturangebot integriert werden. Bei den lizenzierten Materialien sind insbesondere auch die Rechte der Verlage zu wahren und entsprechende Rechtemanagement-Verfahren einzusetzen. Es gilt, transparente Verfahren zu konzipieren und umzusetzen, um für die Verlage die notwendige Vertrauensbasis zu schaffen und gleichzeitig den Einrichtungen ihren berechtigten Zugriff auf die Volltexte zu sichern.

Der vorliegende Text ist die schriftliche Fassung eines Vortrages auf dem 3. Leipziger Kongress für Information und Bibliothek "Information und Ethik" im Rahmen der Fortbildungsveranstaltung der Bibliotheksverbände "Netzwerk Bibliotheksverbände", der vom 19.-22. März 2007 im Congress Center Leipzig stattfand.

**Keywords:** KOBV, Kooperativer Bibliotheksverbund Berlin-Brandenburg, Suchmaschinentechnologie, Volltextrecherche, Volltextserver

**CR:** H.0K, 4.0

---

## **Inhalt**

|                                              |          |
|----------------------------------------------|----------|
| <b>Vorbemerkung .....</b>                    | <b>3</b> |
| <b>Recherche-Angebote der Verbände .....</b> | <b>3</b> |
| <b>Volltextsuche .....</b>                   | <b>4</b> |
| <b>KOBV-Volltextserver .....</b>             | <b>5</b> |
| <b>Volltextserver der Verbände.....</b>      | <b>8</b> |

---

## **Vorbemerkung**

Die Intention ist der kooperative Aufbau einer Infrastruktur durch die Bibliotheksverbände, um den Nutzern Volltext-Angebote dauerhaft und komfortabel zur Verfügung zu stellen: Zeitschriftenartikel und elektronische Dokumente werden mittels Suchmaschinentechnologie indiziert und unter Berücksichtigung von Zugriffsrechten zugänglich gemacht.

Realisiert ist dies bereits im KOBV-Volltextserver, der seit Ende 2005 im Routinebetrieb läuft. Den Grundstock bilden Archivdaten dreier Zeitschriftenverlage, deren Verträge ausgelaufen sind sowie die elektronischen Dokumente der regionalen Hochschul-Publikationsserver. Neben der oft diskutierten Vision, die Trennung zwischen Zeitschriftenartikeln und Dokumenten in der Suche aufzuheben, ist im KOBV-Volltextserver ein weiteres Teilziel realisiert: die Gewährleistung des Zugriffes auf die lizenzierten elektronischen Zeitschriften auch nach Ablauf der Lizenzverträge. An den Verträgen waren 32 Einrichtungen aus 9 Bundesländern beteiligt. Die rund 1.400.000 Volltexte sind mittels Suchmaschinentechnologie indiziert, die in Sekundenschnelle Ergebnisse liefert. Während die Metadatenrecherche und die Sicht auf die Abstracts für alle offen sind, ist der Zugriff auf die Artikel-Volltexte auf die an den Verträgen beteiligten Einrichtungen beschränkt.

Vorstellbar ist ein überregionales Netz von Volltextservern der Verbände, die mittels Suchmaschinentechnologie indiziert und nahtlos in das regionale und lokale Literaturangebot integriert werden. Bei den lizenzierten Materialien sind insbesondere auch die Rechte der Verlage zu wahren und entsprechende Rechtemanagement-Verfahren einzusetzen. Dieser organisatorischen Komponente kommt eine ebenso wichtige Rolle zu wie der technischen Umsetzung: Es gilt, transparente Verfahren zu konzipieren und umzusetzen, um für die Verlage die notwendige Vertrauensbasis zu schaffen und gleichzeitig den Einrichtungen ihren berechtigten Zugriff auf die Volltexte zu sichern.

## **Recherche-Angebote der Verbände**

Die Recherche-Angebote der Bibliotheken und Verbände basieren traditionell und bis in die heutige Zeit auf Metadaten, seien es nun die herkömmlichen Bibliotheks-/Verbundkataloge oder neuere Rechercheinstrumente wie die Portale. Um den Informationsgehalt der Kataloge zu erhöhen, haben Bibliotheken und Verbände vor einigen Jahren mit der sogenannten Kataloganreicherung begonnen, d.h. die Metadaten durch gescannte Inhaltsverzeichnisse, Abstracts, Rezensionen u.ä. zu ergänzen. Die ergänzenden Daten werden in eigenen Repositories gespeichert, mit den zugehörigen Metadaten verlinkt und in der Regel indiziert, um sie suchbar zu machen. Mit der Kataloganreicherung bietet man den Benutzern ein größeres Spektrum an Sucheinstiegen und insbesondere die Möglichkeit, die Rechercheergebnisse zu verifizieren: Bevor ein Benutzer ein Medium bestellt, hat er nun die Möglichkeit zu prüfen, ob der Inhalt des Mediums dem entspricht, was er sucht. Auf diese Weise kann er sich manchen Gang zur Bibliothek oder die Ausleihe des Buches, sei es nun in der Bibliothek oder über die Fernleihe, ersparen. Technische Basis für die erweiterte Dienstleistung ist in der Regel die herkömmliche Bibliothekssoftware mit ihren komplexen und diffizilen Suchmechanismen und Ergebnisanzeigen, die seit Jahren in Bibliotheken und Verbänden eingesetzt wird.

Neue Technologien haben in den vergangenen Jahren einen Mehrwert und eine neue Qualität in die Recherche gebracht: So stehen immer mehr elektronische Publikationen als Volltext im Netz und sind oftmals frei zugänglich. Insbesondere hat aber die Suchmaschinentechnologie die Benutzerbedürfnisse nach einfach zu bedienenden und schnellen Suchwerkzeugen grundlegend gewandelt. Die Verbände regieren darauf, indem sie Suchmaschinentechnologie für bibliothekarische Dienste einsetzen. Der Ist-Stand beim Einsatz von Suchmaschinentechnologie besteht allerdings immer noch darin, dass sie für die Recherche in Metadaten eingesetzt wird und Suchmaschinen als Katalogersatz dienen. Dabei kann die neue Technologie mehr: Mit ihrer Hilfe können Volltexte indiziert und mit der Volltextsuche eine neue Qualität in der Recherche erreicht werden.

---

## Volltextsuche

Was ist Volltextsuche? Nach der Definition ermöglicht die Volltextsuche *das Auffinden von Informationen und Texten aus einer großen Menge unstrukturierter Daten. Dazu werden die Suchbereiche mit entsprechenden Index-Werkzeugen indiziert.*<sup>1</sup> Fast automatisch stellt sich für Bibliothekare hier die Frage, ob die Volltextsuche wirklich für bibliothekarische Zwecke benötigt wird, da gerade die bibliographischen Daten in sehr strukturierter Form vorliegen – nämlich als Metadaten, die zudem maschinenlesbar verarbeitet werden können: Jedes Buch ist im Katalog formal und – zumindest die neuere Literatur – häufig auch sachlich nach bestimmten Kriterien erschlossen.

Die Volltextsuche bietet *mehr* als die Metadatensuche, insbesondere im Hinblick auf die sachliche Recherche:

- Sie ermöglicht das Auffinden von Texten bei "ungenauen" Metadaten: Auch bei groben Tippfehlern (wo selbst fuzzy- und Meinten-Sie-Funktionalität nicht mehr greifen) kann der richtige Treffer gefunden werden, da die Begriffe auch im Volltext vorkommen.
- Sie ermöglicht das Auffinden von Medien ohne irgendeine sachliche Erschließung: Terme, die nicht beschlagwortet sind, werden im Volltext gefunden.
- Sie ermöglicht das Auffinden "versteckter" Informationen: Textbereiche, die nicht verschlagwortet sind, weil der Gesamttext einen anderen Fokus hat, werden gefunden.
- Sie ermöglicht das Auffinden von Medien mit interdisziplinären Inhalten: Medien, die von einem Institut erworben und mit der dortigen Klassifikation erschlossen sind, werden von Nutzern anderer Disziplinen gefunden.
- Und last but not least braucht der Nutzer keine Vorkenntnisse, um qualifizierte Suchergebnisse zu erhalten: Er muss weder normierte Schlagwörter kennen noch eine Fachklassifikation, sondern kann mit "natürlich-sprachigen" Suchbegriffen suchen.

An dieser Stelle ein kurzes Wort zu Metadaten: Metadaten sind eigentlich aus einem Mangel heraus entstanden. Die ausgefeilten Kriterien zur Erschließung von Büchern sind in den Bibliotheken entwickelt worden, *weil* man nicht die Werkzeuge und Mittel hatte, den vorliegenden Volltext in seiner Komplexität vollständig für die Suche bereitzustellen, trotzdem aber bestrebt war, den Benutzern die Suche in einer großen Zahl von Büchern zu ermöglichen. Die Einbeziehung des gesamten Textes in die Recherche ist erst im "digitalen Zeitalter" möglich. Betrachtet man die Frage vor diesem Hintergrund, so müsste die oben aufgeworfene Frage eigentlich umgekehrt gestellt werden – nicht: Wozu benötigt man die Volltextsuche?, sondern vielmehr: Wieso gibt es die Metadatensuche?

Dies soll nun nicht heißen, dass Metadaten oder die Suche mit Metadaten heute überflüssig sind. So liegt vom Gesamtbestand der Literatur erst ein sehr geringer Teil in elektronischer Form vor. Zwar entsteht die heutige Literatur genuin als elektronische Publikation und durch Digitalisierungsprojekte ist der Bestand an elektronischen Volltexten in den vergangenen Jahren immens angewachsen, doch eine Vollständigkeit ist bei weitem noch nicht erreicht. Solange dies nicht der Fall ist, werden Metadaten für die Suche benötigt. Metadaten unterstützen darüber hinaus eine gezielte Suche und sind ein hilfreiches Instrumentarium, solange es keine adäquaten Ranking-Algorithmen für die Ergebnisanzeige der Volltextrecherchen gibt.

Allerdings gibt es durch die Volltexte heute mehr und bessere Sucheinstiege als ausschließlich Metadaten und diese Möglichkeiten sollte man nutzen, gerade auch im Hinblick auf die sachliche Erschließung, die sehr teuer, bei weitem nicht umfassend und daher für die Suche nur bedingt einsetzbar ist. Bei der Implementierung der Volltextsuche können die oben angerissenen Vorteile der Volltextsuche genutzt und gleichzeitig die vorhandenen Metadaten mit den Volltexten kombiniert und eingesetzt werden, um die Suchergebnisse zu strukturieren.

---

<sup>1</sup> s. den Artikel in Wikipedia, Stichwort "volltextsuche"; <http://de.wikipedia.org/wiki/Volltextsuche>

## KOBV-Volltextserver

Die Methode, Volltexte und Metadaten in der Suche zu kombinieren, verwendet auch der KOBV-Volltextserver, der hier als Beispiel für einen Volltextserver mit integriertem Zugriff auf lizenzierte Zeitschriftenartikel und freie elektronische Dokumente vorgestellt wird. Der KOBV-Volltextserver wurde auf der Basis von Lucene aufgebaut, einer Open-Source-Programmbibliothek, und läuft seit 2005 im Routinebetrieb.<sup>2</sup>

Wie in anderen Suchmaschinen auch gibt es im KOBV-Volltextserver eine Einfeld-Suche (dahinter steht die Volltextsuche), die "Meinten Sie"- und fuzzy-Funktionalität, um nur einige zu nennen. Selbstverständlich erfolgen Recherche und Anzeige der Ergebnisse auch im KOBV-Volltextserver in Sekundenbruchteilen.

Die Metadaten-Suche ist in die Erweiterte Suche integriert. Mittels der Metadaten kann der Benutzer – abgestimmt auf den Inhalt des Volltextservers – spezifisch nach diversen Informationen suchen, wie "Volltext", "Autor", "Titelstichwörtern", "Abstract", "Schlagwörtern", "Jahr", "Zeitschriftentitel", "ISSN" und "Verlag".



Abb. 1: KOBV-Volltextserver, Metadaten in der Erweiterten Suche

Im KOBV-Volltextserver ist der integrierte Zugang zu lizenzierten Zeitschriftenartikeln und freien elektronischen Dokumenten realisiert. Der Server wurde von der KOBV-Zentrale in Zusammenarbeit mit dem Friedrich-Althoff-Konsortium (FAK) und den regionalen Bibliotheken aufgebaut und enthält

- die Archivdaten der Verlage Elsevier, Kluwer Academic Publishers und Springer, deren Verträge ausgelaufen sind und für die der langfristige Zugriff auf diese Weise gesichert werden soll. An den Verträgen des FAK waren Hochschulen und wissenschaftliche Institute beteiligt, insgesamt 32 Einrichtungen aus 9 Bundesländern. Recherchierbar sind derzeit 1.600 Zeitschriften mit rund 1.400.000 Artikeln.
- die elektronischen Dokumente aus den regionalen Hochschul-Publikationsservern, derzeit rund 5.000 Publikationen.

Während die Metadatenrecherche und die Sicht auf die Abstracts für alle Benutzer frei zugänglich sind, ist der Zugriff auf die Volltexte der lizenzierten Zeitschriften selbstverständlich auf die ehemaligen Lizenznehmer beschränkt. Für das RechteManagement wurde dazu in Absprache mit den Verlagen analog zum Verfahren der Verlage ein IP-Checking implementiert. Der Freischaltmodus ist durch eine Ampel visualisiert, die – abhängig von der IP-Adresse – rot oder grün geschaltet wird; bei den freien e-Dokumenten ist sie grün. Bei roter Ampel wird dem Benutzer – ebenfalls kontextsensitiv – der Button "KOBV-Dienste" (SFX) angezeigt, über den ihm weitere Suchmöglichkeiten angeboten werden.

<sup>2</sup> KOBV-Volltextserver: <http://volltexte.kobv.de>

Anhand einer Beispielrecherche soll der integrierte Zugang zu Zeitschriftenartikeln und e-Dokumenten verdeutlicht werden. Die Suche mit dem Suchbegriff "non-anthropocentric" bringt eine Ergebnisliste mit 172 Treffern, von denen Treffer 1 und Treffer 17 hier vorgestellt werden:

- *Treffer 1: Zeitschriftenartikel*

Die Ergebnisliste zeigt als 1. Treffer einen Zeitschriftenartikel. Der Suchbegriff wurde im Abstract gefunden. Durch Klick auf den Titel des Artikels gelangt der Benutzer zum Volltext, wobei die grüne Ampel signalisiert, dass der Benutzer zugriffsberechtigt ist.

Über den verlinkten Zeitschriftentitel könnte der Benutzer weiter gehen und sich den gesamten Zeitschriftenband mit weiteren Volltexten anzeigen lassen.

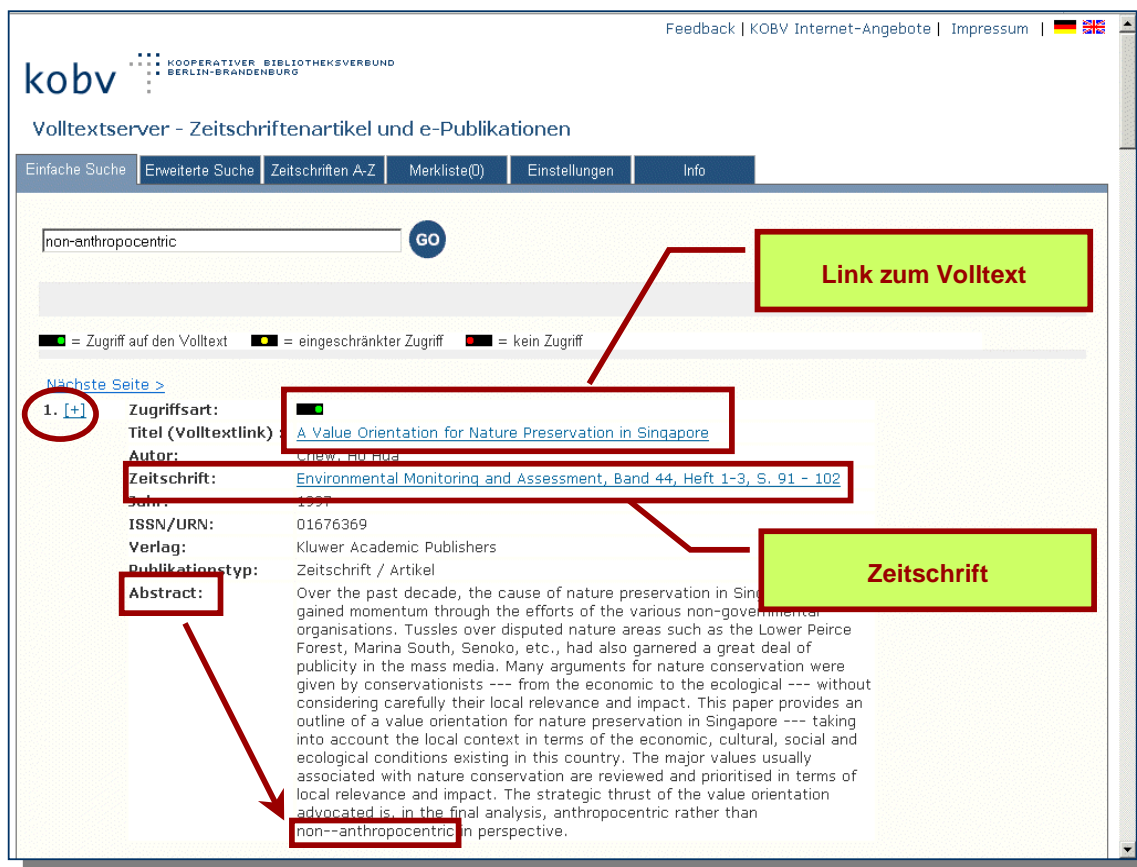


Abb. 2: Treffer 1 – Zeitschriftenartikel



- Treffer 17: e-Dokument

Die Ergebnisliste zeigt als 17. Treffer ein elektronisches Dokument. Es handelt sich um eine Publikation der Technischen Universität Berlin. Der Suchbegriff wurde ebenfalls im Abstract gefunden. Durch Klick auf den Titel des Dokumentes gelangt der Benutzer auch hier zum Volltext, wobei auch in diesem Fall die grüne Ampel die Zugriffsberechtigung signalisiert.

The screenshot shows a search result for an e-document. The following fields are visible:

- Zugriffsart:** [Green Ampel]
- Titel (Volltextlink):** [A framework of values and criteria for interdisciplinary evaluations of nature and landscapes](#)
- Autor:** Ziserlis, Marcus
- Jahr:** 2006
- ISSN/URN:** <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:kobv:83-opus-12327>
- Verlag:** Technische Universität Berlin
- Publikationstyp:** Dissertation
- Abstract:** Nature and landscapes cover all fields of life of the human being... of fauna and flora in an interacting relation. Therefore, it is a survey, to analyse, and to evaluate these particular parts of comprehensively. The present research is therefore devoted to comprehensive values and criteria for interdisciplinary evaluations of nature and landscapes. Thereby, it distinguishes between a level of values of the different anthropocentric demands of individuals and of the society on nature and landscapes, as well as a non-anthropocentric value on their own. Already developed evaluation criteria are condensed to a common framework of the same evaluation criteria of the different scientific disciplines for comparable evaluations of these comprehensive values of nature and landscapes on application level. Nevertheless, each scientific discipline works according to its own appropriate survey and analysis methods. Thus, scientists of the different disciplines carry out an evaluation of the particular case in spatial and time relation according to the proposed common evaluation criteria under participation of the public. The resulting competing values must be weighed up against each other for concrete implementation recommendations and decisions. There are frequently implementation and acceptance problems of nature conservation, which are based on communication problems due to the one-sided orientated evaluation methods of natural sciences and their inhomogeneous evaluation criteria, and thereby often rather less comprehensible evaluation methods. The present thesis fills this scientific gap by providing the approach of a comprehensive interdisciplinary evaluation framework, and a discussion of common survey and analysis methods of the different scientific disciplines. Moreover, the scientific research discusses backgrounds of the individual and the society of the particular values within an interactive dynamic development process towards nature and landscapes, and it draws conclusions for the application of an interdisciplinary evaluation method. The application of the interdisciplinary evaluation framework is discussed on the basis of a theoretical and a practical evaluation example of Neobiota, as well as, an empirical study of 51 Environmental Impact Studies from Germany dating from the 1990s, which reveals the high demand of it. Finally, there is an overview provided of some research deficits, especially the need to develop appropriate indicators for the generally limited available time and possible extent of the survey, analysis, and evaluation parts.

Abb. 3: Treffer 17 – e-Dokument

Wie das Beispiel zeigt, ist die integrierte Volltextsuche von Zeitschriftenartikeln und elektronischen Dokumenten möglich. Im KOBV ist man bei den konzeptionellen Überlegungen zum Aufbau des Volltextservers davon ausgegangen, dass der Benutzer den Text haben möchte. Die Frage, wo ein Text veröffentlicht wurde, ob in einer Zeitschrift oder als Monographie ist für ihn gleichgültig; vielfach weiss er den Unterschied überhaupt nicht.

Die Unterscheidung zwischen den Veröffentlichungsarten ist – insbesondere, wenn es um Volltexte geht – eher eine künstliche und in den Bibliotheken durch die Erschließungskonventionen bedingt. Für die Volltextsuche ist sie irrelevant. Auch vor diesem Hintergrund bringt die Volltextsuche eine neue Qualität in der Benutzerrecherche.

---

## **Volltextserver der Verbände**

Das am Beispiel KOBV-Volltextserver vorgestellte Konzept der integrierten Volltextsuche ist auf einen überregionalen Dienst der Verbände übertragbar. Dabei geht es nicht darum, eine Konkurrenz zu Google aufzubauen, sondern den in Bibliotheken und Verbänden vorhandenen Content, auf den Google keinen Zugriff hat, aufzubereiten und anzubieten.

Ein überregionaler Volltextserver existiert bislang nicht. Die Intention ist der kooperative Aufbau einer Infrastruktur durch die Bibliotheksverbände, um den Nutzern Volltexte, die in Bibliotheken und Verbänden inzwischen in einer Vielzahl vorhanden sind, dauerhaft und komfortabel zur Verfügung zu stellen. Zunächst geht es darum, ein Konzept über die Inhalte zu erstellen, diese zu benennen und zu strukturieren, beispielsweise: Artikel von Zeitschriften, deren Lizenzierung ausgelaufen ist (in Zusammenarbeit mit den Konsortien), e-Publikationen (die im Zusammenhang mit Open Access entstehen), Nationallizenzen (in Zusammenarbeit mit den lizenzverhandelnden Einrichtungen), ... Weiterhin ist ein Konzept für die technische und organisatorische Realisierung zu entwickeln – z.B. Erstellung und Verfügbarmachung von Volltextindices; Aufbau eines Dienstes, der in lokale und regionale Angebote integriert werden kann; gemeinsamer verteilter Betrieb in den Verbänden; ... Der Zugriff auf lizenzierte Materialien erfordert ein differenziertes Rechtemanagement, wobei das in der Universität Freiburg entwickelte System zur verteilten Authentifizierung, Autorisierung und Rechteverwaltung (AAR) zum Einsatz kommen kann.

Dabei kann es im Zusammenhang mit dem hier vorgestellten Kontext nur darum gehen, ein *überregionales* Angebot aufzubauen, keine regional begrenzten Dienste. Bei elektronisch über das Internet verfügbaren Volltexten, die mittels moderner Technologien und vernetzter Systeme von überall zugänglich und technisch keinerlei Standortbeschränkungen unterworfen sind, macht für den Benutzer nur ein möglichst umfassendes Angebot Sinn.