

TAXONOMIEN, THESAURI UND ONTOLOGIEN - WEGBEREITER FÜR BESSERE ONLINE- RECHERCHEPRODUKTE(?)

Roman Huditsch

Teamleiter XSLT-Konvertierung DE/AT
LexisNexis Verlag ARD Orac GmbH & CoKG
Marxergasse 25
1030 Wien
roman.huditsch@lexisnexis.at; <http://roman.huditsch.info>

Schlagnworte: *Taxonomie, Ontologie, Thesaurus, Produktentwicklung, Volltextsuche, Benutzerführung, Social Tagging, Folksonomies, Klassifikation*

Abstract: *Klassische Volltextsuchen stellen nach wie vor den primären Informationszugang vieler Online-Rechercheprodukte dar. Mit Hilfe von Methoden und Werkzeugen des Bibliothekswesens wären jedoch auch andere, klassifikationsbasierte Zugangsmöglichkeiten denkbar. Derartige Implementierungen mit konkretem Kundennutzen sind jedoch nach wie vor spärlich gesät. Dieses Dokument soll sowohl interessante Möglichkeiten für den praktischen Einsatz dieser Instrumente als auch mögliche damit einhergehende Problematiken aufzeigen.*

Einführung

Der primäre Zugang zu den zahlreichen Dokumenten der bekannten Online-Rechercheprodukte im Rechtsbereich hat sich im Laufe der vergangenen Jahre nicht oder nur sehr unwesentlich geändert. Nach wie vor bildet die Suche – in den meisten Fällen eine klassische Volltextsuche – das Eintrittsportal zu den juristischen Inhalten.

Diese Tatsache hat sich auch durch die Diskussion rund um Themen wie „Web 2.0“ oder „Semantic Web“ nicht verändert. Es wurden zwar immer wieder Begriffe aus dem klassischen Bibliothekswesen aufgegriffen und auf unterschiedliche Weisen zumindest versuchsweise in den Onlinebereich übertragen, jedoch kamen darauf aufbauende Prototypen bislang nur sehr selten über ein Beta-Stadium hinaus und wurden gerade während der aktuellen Wirtschaftskrise oft wieder eingestellt.

Bleibt die allgegenwärtige Suche also weiterhin alleiniger Zutrittspunkt zum verschriftlichten juristischen Wissen? Welche alternativen Möglichkeiten wären denn mit Taxonomien, Thesauri und artverwandten Werkzeugen denkbar? Wo liegen dabei die Stolpersteine, die es zu überwinden gilt?

Ständig auf der (Volltext-)Suche

Wenn man sich die verschiedenen juristischen Online-Rechercheprodukte anschaut, die derzeit auf dem deutschsprachigen Markt angeboten werden, fällt sofort eines auf: Die meisten – wenn nicht sogar alle – vertrauen darauf, das Informationsbedürfnis ihrer Kunden über eine mehr oder weniger einfach gehaltene Suche befriedigen zu können.

Standardmäßig wird dem Kunden dabei eine klassische Volltextsuche geboten, die durch spezielle Suchfelder für gezielte Metadaten-Suchen ergänzt und abgerundet wird. Dieses Konzept findet sich auch durchgehend bei allen großen Anbietern auf dem Rechtsmarkt: LexisNexis, RDB, juris, beck-online und bei vielen anderen mehr.

Gerade für sehr gezielte Suchen oder für geübte bzw. gut geschulte Kunden stellt diese Art des Informationszugangs auch eine sehr einfache und schnelle Möglichkeit dar, zu den gewünschten Zieldokumenten zu gelangen.

1.1. Ein Weg mit Hindernissen

Leider führt der Weg über eine Suche jedoch nicht immer zum gewünschten Erfolg.

Besonders für Nutzer, die nur sehr sporadisch Gebrauch vom jeweiligen Produkt machen oder ihr Informationsbedürfnis nur sehr vage oder sogar gar nicht in Worte fassen können („da gab’s doch dieses eine Urteil...“), ist diese Art des Zugangs nur bedingt geeignet.

Erschwerend kommt dabei hinzu, dass eine klassische Volltextsuche darauf angewiesen ist, die verwendeten Suchworte auch im jeweiligen Zieldokument vorzufinden. Dabei muss jedoch angemerkt werden, dass sich Suchmaschinen gerade in diesem Punkt zumindest ein wenig durch die Verwendung von Synonymlisten, Stemming und ähnliche Mittel zu helfen wissen.

Mit alternativen Wegen zum Ziel(dokument)

Aus dieser Problemstellung heraus ergibt sich die Frage, ob es daher nicht sinnvoll wäre, neben der klassischen Suche auch alternative Wege zu den gespeicherten Informationen anzubieten. Diese Frage wird wohl in den meisten Fällen von den Produktanbietern bejaht werden, zumal sich daraus Vorteile für die eigene Kunden ergeben, die über kurz oder lang zu einer erhöhten Kundenzufriedenheit führen.

Solche alternative Suchwege sind vor allem für jene Nutzer interessant und wichtig, die ihren Informationsbedarf nicht eindeutig benennen oder – wegen eventuell mangelnder Produktkenntnisse – nicht in eine passende Suchanfrage gießen können.

In den letzten Jahren ist immer wieder vom „Web 2.0“ und dem „Semantic Web“ die Rede gewesen. In engem Zusammenhang damit wurden auch immer wieder Begriffe genannt, die uns aus dem klassischen Bibliothekswesen bekannt sind. Unter diesen Begriffen finden sich zum Beispiel „Taxonomie“, „Thesaurus“, „Ontologie“ und viele andere mehr.

Dass in diesem Zusammenhang oft Ungenauigkeiten und Verwechslungen bei der Begriffsverwendung passiert sind, soll an dieser Stelle nur am Rande erwähnt werden.

Grundsätzlich kann man jedoch festhalten, dass es bei diesen Begriffen, wenn sie in Zusammenhang mit Online-Rechercheprodukten verwendet werden, letztendlich immer um die inhaltliche Klassifizierung von Dokumenten anhand von flexiblen (Tagging, Folksonomies, ...) oder fest vorgegebenen Begriffsstrukturen (kontrollierte Vokabulare) geht.

Je nach Komplexität und Verlinkung der einzelnen Begriffe innerhalb der jeweiligen Struktur lassen sich dann Taxonomien (Ober- und untergeordnete Begriffe), Thesauri („siehe“- und „siehe auch“-Beziehungen) und Ontologien (Begriffsnetze) unterscheiden.

Eine solche inhaltliche Klassifikation eröffnet die Möglichkeit, dem Kunden im Produkt eine einfache – visuell möglichst ansprechende – Einstiegshilfe anzubieten, über die er sich schrittweise immer weiter in den Datenbestand vorarbeiten kann.

1.2. Kein Weg ohne Schlaglöcher...

Auf dem Weg zu solchen Lösungen gilt es jedoch gleich zu Beginn zwei essentielle Fragen zu beantworten:

- Wie lässt sich mit möglichst wenig Aufwand eine Begriffsstruktur erstellen?
- Wie können Dokumente (automatisch?) diesen Begriffen zugeordnet werden?

Gerade die erste Frage wird sehr gerne bei der Produktentwicklung unterschätzt. Hier gilt es beispielsweise zu entscheiden, ob eine Begriffsstruktur manuell von Grund auf neu erstellt werden soll oder ob versucht wird, automatisch – mit entsprechender manueller Nacharbeit - aus den vorhandenen Dokumenten wichtige Begriffe mit Hilfe komplexer Algorithmen und Text-Mining-Technologien zu extrahieren. Weiters gilt es hier auch festzulegen, bis zu welcher Tiefe und in welchem Umfang Strukturen gebildet werden sollen. Gerade der Aufbau einer möglichst vollständigen Begriffsstruktur ist sehr aufwandsintensiv und daher projektkritisch.

Das Hauptaugenmerk bei der zweiten Frage liegt vor allem im Umgang mit Unschärfen bei der Zuordnung von Dokumente zu einer bestimmten Begriffsklasse, Synonymen und sonstigen sprachlichen Herausforderungen.

Social Tagging - der Kunde als „Arbeitskraft“?

In den letzten Jahren machten bekannte Online-Dienste wie del.ici.ous oder Flickr mit sogenannten „Folksonomies“ oder dem „Social Tagging“ von sich reden. Dabei werden online bereit gestellte Inhalte (Links, Fotos, Dokumente, ...) durch eine (möglichst) große Anzahl von Nutzern selbst verstichwortet bzw. „getaggt“. Die zur Beschreibung des Inhalts verwendeten Begriffe werden dabei von den Nutzern individuell vergeben, wobei oft nur sehr rudimentäre Namenskonventionen einzuhalten sind.

Dieser etwas neuere Ansatz stellt sicherlich eine interessante Option zur Beantwortung der in Kapitel 1.2 gestellten Fragen dar, bringt jedoch ebenfalls einige Probleme mit sich, die es zu lösen gilt:

- Sprachenmix (in welcher Sprache soll die Verstichwortung von fremdsprachigen Informationen erfolgen?)
- Singular/Plural (sind nur Singular-Formen als Tags zugelassen, oder werden auch Plural-Nennungen akzeptiert?)
- keine einheitliche Benennungen (Nutzer haben unterschiedliche Vokabulare und weisen Dokumenten daher unterschiedliche Wörter mit gleicher Bedeutung zu)
- keine semantischen Beziehungen zwischen den einzelnen Tags (nach der Eingabe der Tags stehen diese losgelöst neben einander und weisen keinerlei Beziehungen zu anderen Tags auf)

1.3. Also eine Sackgasse?

Trotz dieser Schwierigkeiten muss die Möglichkeit des „Social Taggings“ jedoch nicht gleich als Sackgasse abgetan werden. Sie erfordert allerdings, dass Nutzer bei der Eingabe von Tags produktseitig bestmöglich unterstützt werden. Das kann beispielsweise so aussehen, dass dem User bei der Eingabe der ersten Buchstaben dem System bereits bekannte Begriffe vorgeschlagen werden.

Neben diesen eher technischen Problemstellungen gilt es auch noch organisatorische Fragen zu klären. Soll es beispielsweise nur einer bestimmten Gruppe von Nutzern möglich sein, Dokumenten eigene Tags zuzuordnen, oder soll jeder Nutzer diese Möglichkeit haben?

Welche Anreize können den Nutzern des Produktes überhaupt geboten werden, damit sie diesen – wenn auch nur geringen – Aufwand auf sich nehmen?

Können sich durch dieses Vorgehen negative qualitative bzw. vielleicht sogar rechtliche Folgen für den Produkthanbieter ergeben?

Schlussfolgerungen

Klassifikationsbasierte Informationszugänge fristen derzeit vor allem im Bereich der juristischen Informationsportale noch bestenfalls ein Schattendasein neben der klassischen Volltextsuche. Nichts desto trotz wächst jedoch ständig das Bewusstsein der Produkthanbieter, Informationen auch über zusätzliche, alternative Wege erschließbar machen zu müssen. Damit einhergehend steigt auch wieder die Bereitschaft, in diesem Bereich auch die nötigen Investitionen zu tätigen.

Aus diesen Gründen wird sich die derzeit noch bestehende Lücke zwischen Forschung und den noch in Entwicklung befindlichen Prototypen auf der einen und den tatsächlichen Live-Produkten auf der anderen Seite meiner Einschätzung nach in absehbarer Zeit deutlich verringern.

Ob Nutzer-zentrierte Klassifizierungsmethoden wie das „Social Tagging“ auch für juristische Inhalte bzw. entsprechende Recherche-Produkte eine attraktive und funktionierende Möglichkeit bieten, wird dabei noch getestet werden müssen.

Literatur

Anja Schütte, Anke Buechter, Vom Thesaurus zur Ontologie. http://www.bui.haw-hamburg.de/pers/ulrike.spree/ws0809/buechter_schuette_skos.ppt, aufgerufen 3.2.2010 (2008)

Georg Guentner, Rolf Sint, Rupert Westenthaler, Ein Ansatz zur Unterstützung traditioneller Klassifikation durch Social Tagging. <http://www.ist-live.org/intranet/salzburg-research001-45>, aufgerufen 13.1.2010 (2008)

Jack Park, Sam Hunting, und Douglas C. Engelbart, XML Topic Maps: Creating and Using Topic Maps for the Web, Amsterdam (2002).

Peter Scheir, Ontologie et al.- Begriffsdefinitionen im Kontext Wissensrepräsentation. <http://www.know-center.at/swat>, aufgerufen 13.1.2010 (2006).

Philipp Bachmann, Wissensmanagement mit Wikis und Folksonomies: Einsatzmöglichkeiten in der politischen Kommunikation . http://web20.obm-wahl.de/uploads/Praesentationen/Referat_Wikis_und_Folksonomies.ppt, aufgerufen 3.2.2010 (2006).

Richard Widhalm, Thomas Mück, Topic Maps. Semantische Suche im Internet, Berlin (2002).

Stefanie Panke, Nutzung und Nutzen von Social-Bookmarking-Systemen. www.e-teaching.org/didaktik/tagging-small.ppt, aufgerufen: 13.1.2010 (2007).

Thanh Tran Duc, Ontologie - Konzepte zur Entwicklung und Nutzung . <http://ivs.cs.uni-magdeburg.de/~dumke/OperWeb/Duc/Folien.ppt>, aufgerufen 3.2.2010.