

# KLASSIFIKATION UND RANKING VON DOKUMENTTYPEN IM JURISTISCHEN WISSENSKORPUS

Erich Schweighofer

Ao. Universitätsprofessor, Universität Wien, Arbeitsgruppe Rechtsinformatik (DEICL/AVR)  
Schottenbastei 10-16/2/5, 1010 Wien, AT  
Erich.Schweighofer@univie.ac.at; <http://rechtsinformatik.univie.ac.at>

**Schlagworte:** *Dokumenttypen, Soziale Medien, Wissenskorpus, Verweisung, Ranking*

**Abstract:** *Ein starker Wandel kennzeichnet die juristische Veröffentlichung und Publikationskultur; neben sofortiger Verfügbarkeit ist die Ausrichtung auf einen «virtuellen Diskussionsprozess» im Internet ein weiteres Charakteristikum. Die Dokumenttypen werden mehr und vielfältiger. Es bedarf der Berücksichtigung von bibliografischen Angaben, der Metadaten, der Autoritätsbewertung des Autors, des zeitlichen Kontexts und des Verweisungsnetzwerks, um die Relevanz der Treffer eines Rechtsretrieval-Systems für eine Rechtsfrage nutzerorientiert sortieren zu können. In diesem Beitrag wird ein theoretisches Modell entwickelt, welches in Zukunft unter Einbeziehung vorhandener Instrumente umgesetzt werden soll.*

## 1. Einleitung

Die juristische Veröffentlichung und Publikationskultur war immer vielfältig; ist aber derzeit in einem weiteren starken Wandel begriffen. Die Dokumente der Rechtsetzung und Rechtsprechung sollen in Echtzeit verfügbar sein. Dazu wird der Publikationsprozess der Lehre und Praxis schneller und vielschichtiger. Ein weiteres Merkmal ist die Ausrichtung auf einen «virtuellen Diskussionsprozess». Die zunehmende Digitalisierung unterstützt sowohl die Vielfalt als auch die Schnelligkeit. Homepages und Blogs werden zunehmend eine relevante Publikationsform. Die lange vorherrschende Dominanz von Verlagen bei der Publikation rechtswissenschaftlicher Texte schwächt sich ab. Soziale Medien sind zwar derzeit noch wenig bedeutsam, was sich aber noch ändern wird.

Diese an sich erfreulichen Änderungen bedingen aber, dass wesentlich mehr Sorgfalt bei der Bewertung der Autorität des Dokuments erforderlich ist. Hierzu stehen bibliografische Angaben (insbes. Autor) zeitlicher Kontext des Dokuments, Einordnung in die Vernetzungsstruktur sowie Ranking zur Verfügung. Das Endergebnis – eine hochwertige Sammlung relevanter Quellen zu einem Rechtsproblem – muss gleich wie bisher sein; aber der Weg wird durch die Berücksichtigung von mehr und unterschiedlichen Quellen schwieriger. Der wesentliche Vorteil besteht darin, dass mehr Materialien in kürzerer Zeit zur Verfügung stehen und rascher eine Stabilität in der Diskussion erreicht werden kann (z.B. herrschende Meinung in Rechtsprechung und Literatur).

## 2. Erweiterung des juristischen Wissenskorporus

Das Grundelement des juristischen Wissenskorporus ist und bleibt das Dokument; wenn auch nicht mehr das reine Textdokument, sondern zunehmend auch Multimedia-Dokumente (Bilder, Videos etc.).<sup>1</sup> Sehr häufig werden Dokumente in Dokumentklassen zusammengefasst; insbesondere nach Autoren, Zeiträumen, Veröffentlichungsformen bzw. Inhalten. Eine besondere Einheit bilden oft Autoren und deren Dokumente. Auch hier kommt es zu Differenzierung nach Zeiträumen bzw. Inhalten.

Der Kern des juristischen Wissenskorporus bleibt natürlich unverändert: Gesetzesblätter, Urteilssammlungen und wissenschaftliche Literatur, insbes. Zeitschriften und Kommentare. Die Veröffentlichungsform wechselt zunehmend von Papier auf Online; herausragende Beispiele sind das österreichische Bundesgesetzblatt (online authentisch seit 2004)<sup>2</sup> als auch das EU-Amtsblatt<sup>3</sup>. Bei den Urteilssammlungen setzt sich zunehmend der Trend zur Veröffentlichung aller Urteile durch, was aus Kostengründen nur online machbar ist. Damit sind die gedruckten Urteilssammlungen schleichend überlagert worden und stellen nunmehr vornehmlich eine Auswahl «wichtigerer Urteile» dar. Die Bedeutung hat in der Praxis nicht wirklich abgenommen; ihre Rolle ähnelt zunehmend jener der Gesetzessammlungen (z.B. KODEX Reihe)<sup>4</sup>. Wissenschaftliche Zeitschriften sind schon seit langem (in Österreich seit 1986) auch online verfügbar; für Archivzwecke sind die Online-Versionen schon jetzt dominant. Schleichend verdrängt Online das Print-Produkt; insbes. weil für Lesezwecke jederzeit ein gut verwendbar Ausdruck möglich ist. Kluge Verlage nutzen diesen Trend und fokussieren das Printprodukt auf diesen Zweck.

Das Internet macht jeden zum potentiellen Autor und Selbstverleger. Es bedarf weiterer Kriterien, damit diese schriftlichen und multimedialen «Ergüsse» auch relevant für den juristischen Wissenskorporus sind, insbes. die Autorität des Autors und der Inhalt selbst. Folgende Formen können zur Erweiterung des juristischen Wissenskorporus genannt werden: Websites, Blogs und Soziale Medien.

*Websites:* Verwaltungen, Parlamente und Gerichtshöfe nutzen schon seit langem das Web, um ihre Inhalte bürgernah anzubieten. Herausragendes Beispiel ist die Website EUROPA, die größte juristische Website weltweit, mit den Subsites Amtsblatt der EU bzw. EuGH. Aber auch das österreichische Parlament, die österreichische Gerichtshöfe und die Verwaltung nutzen diese Möglichkeit; wenn auch unter einer Vielzahl von Homepages, aber mit der einheitlichen Subdomain gv.at. Am Beispiel des österreichischen Rechtsinformationssystems RIS sieht man, dass Websites als Rechtsportale zunehmend als Anlaufstelle fungieren. Wer eine österreichische Norm sucht, verwendet unbewusst selbstverständlich das Portal des RIS. Die zusätzliche Indexierung des RIS in der wichtigsten Suchmaschine Google unterstützt diesen Prozess.

*Blogs:*<sup>5</sup> Als neue Publikationsform ist insbesondere der Blog anzuführen. Er zeichnet sich durch Schnelligkeit mit Verzicht auf einen umfassenden Fußnotenapparat aus. Er entspricht damit am ehesten der Glosse. Weiters können sich auch Dritte zu diesem Thema äußern und Kommentare abgeben. Blogs zielen auf den Einstieg in die Diskussion, also wenn noch keinerlei juristische Bewertung zu einem Thema vorhanden ist. Merkmale sind Schnelligkeit, Stellungnahme und Dialog statt umfassender Themenbehandlung. Sowohl die Qualität des

---

<sup>1</sup> Vgl. SCHWEIGHOFER, ERICH, Rechtsinformatik und Wissensrepräsentation. Springer, Wien (1999); SCHWEIGHOFER, ERICH, Rechtsdatalogik – Versuch einer Teiltheorie der Rechtsinformatik. In: Erich Schweighofer, Franz Kummer, Walter Hötzendorfer (Hrsg.), Kooperation: Tagungsband des 18. Internationalen Rechtsinformatik Symposions IRIS 2015. books@ocg.at, Wien 2015, 61–72 (2015).

<sup>2</sup> Website RIS Rechtsinformationssystem des Bundes, <https://www.ris.bka.gv.at>.

<sup>3</sup> Website EUR-Lex, <http://eur-lex.europa.eu/oj/direct-access.html>.

<sup>4</sup> Der KODEX des österreichischen Rechts wird von WERNER DORALT herausgegeben und von den Verlagen LexisNexis und Linde publiziert. Infos: <http://www.lexisnexis.at/unsere-loesungen/rechtsanwaelte-und-notare/rechercheloesungen/nach-buechern-zeitschriften-und-online-medien-suchen/kodex.aspx> (zuletzt abgerufen: 3. Februar 2016).

<sup>5</sup> Wikipedia DE, Blog <https://de.wikipedia.org/wiki/Blog>. Für juristische Blogs hat sich die Bezeichnung «Blawg» etabliert: <http://www.jurawiki.de/Blawg> (zuletzt abgefragt 3. Februar 2015).

Autors als auch die Diskussionsform müssen den Anforderungen entsprechen. Als gutes Beispiel sei auf den Blog des Richters am österreichischen Verwaltungsgerichtshof Peter Lehofer verwiesen.<sup>6</sup> Verlage beginnen nunmehr, diese sehr rasche Kommunikationsform, insbes. für neuere Judikatur, zu nutzen.<sup>7</sup>

*Soziale Medien:* Zunehmend werden auch wichtigere Dialoge über «Facebook & Co.» geführt; soziale Medien sind zur gesellschaftlichen Agora geworden. Regierungen, Gerichtshöfe wie Unternehmen verbessern ihre Präsenz im sozialen Netzwerk, um näher beim Bürger zu sein und ihm eine einfache Möglichkeit zum Dialog zu bilden. Abgesehen vom Aspekt der Privatsphäre muss auch dafür gesorgt werden, dass ausreichende Qualitätskriterien für den Content als auch für den Dialog gegeben sind.<sup>8</sup> Eine neue, noch seltene Form der Publikation, aber auch Kommunikation, ist die Information über Urteile durch Gerichte über Facebook-Seiten.<sup>9</sup>

Entscheidend für die Aufnahme in den juristischen Wissenskorporus ist vornehmlich der Inhalt. Die Präsentation mag sich ändern, aber der Inhalt muss den Qualitätskriterien rechtswissenschaftlicher Methodik entsprechen. Zusammenfassungen oder Wegweiser der Rechtsnormen<sup>10</sup> oder der Urteile mit rechtsdokumentarischer Qualität ergänzen schon seit längerem den Wissenskorporus. Die Rechtsdokumentation wird nicht nur intensiver betrieben, sie ist auch ein wichtiges Produkt zur Dissemination der jeweiligen Inhalte, insbes. über die Medien. Durch das Internet kommt die Argumentationskomponente dazu. Websites, Blogs und Soziale Medien sind auch Ebenen des juristischen Diskurses; auch «best practice» bildet sich langsam heraus. Es bedarf klarer Spielregeln, damit auch die Qualität des Austausches juristischer Argumente wie in Zeitschriften erreicht wird. An sich ist diese Entwicklung sehr begrüßenswert; es dürfen aber niemals die Kriterien eines rationalen Dialogs<sup>11</sup> verletzt werden.

### 3. Metadaten

Es ist unstrittig, dass es ohne Metadaten nicht geht. Hier stehen am Beginn die klassischen Kriterien der Informationswissenschaft zur Beschreibung von Dokumenten: Autor, Titel, Zeit, Veröffentlichungsinstrument. Diese Metadaten umrahmen und ergänzen den eigentlichen Content. An sich ist eine korrekte Autorenbezeichnung unumgänglich. Aber bei der Einbeziehung von Bürgern muss auch darauf geachtet werden, dass typisierte Identitäten bereitgestellt werden, die, falls notwendig, eine ausreichende Anonymität bei Publikation gewährleisten (z.B. Niederösterreicher, 50 Jahre). Bei rechtlichem Bedarf muss aber eine gerichtliche Klärung der Identität möglich sein. Desgleichen muss es akkreditierte Identitäten geben, um an sich personenbezogene Daten in anonymisierter Form verwenden zu können (z.B. 100 RDB-Nutzer haben sich dieses Dokument ausgedruckt).

<sup>6</sup> e-comm, Blog zum österreichischen und europäischen Recht der elektronischen Kommunikationsnetze und -dienste: <http://blog.lehofer.at/> (zuletzt aufgerufen: 24. Januar 2016).

<sup>7</sup> Ein gutes Beispiel ist der Verlag Editions Weblaw in Bern ([http://www.weblaw.ch/competence/editions\\_weblaw.html](http://www.weblaw.ch/competence/editions_weblaw.html)). Blogs werden in das Autorennetzwerk integriert und liefern Content, Weblaw überprüft Quelle und Qualität und nimmt diese dann in den dRSK – Der digitale Rechtsprechungs-Kommentar – auf. Auf Wunsch gibt es einen Push-Service zu neuesten Entscheidungen. Als multimediale Ergänzung sind Videos und Audios (Podcasts) zu Bundesgerichtsurteilen aus dem Haftpflicht-, Sozialversicherungs- und Zivilprozessrecht verfügbar. Das Wissensnetzwerk von WoltersKluwer JURION geht in eine ähnliche Richtung (<https://www.jurion.de/>).

<sup>8</sup> Vgl. dazu SCHILLING, PETER, Nutzen und Zulässigkeit kommerzieller und eigener Social Media durch öffentliche Einrichtungen. In diesem IRIS-Tagungsband (2016).

<sup>9</sup> Als Beispiel sei das Oberste Gericht in Brasilien (Supremo Tribunal Federal) angeführt: <https://www.facebook.com/SupremoTribunalFederal.STF/> (zuletzt abgefragt: 3. Februar 2016). Dieses hat aber auch schon entsprechende Leitlinien ausgearbeitet.

<sup>10</sup> Im Europarecht gibt es Zusammenfassungen der EU-Gesetzgebung seit der Binnemarktinitiative in den 1980er Jahren; ansonsten sind dies eher Erläuterungen oder Wegweiser. Nunmehr sind die Zusammenfassungen der EU-Gesetzgebung in EUR-Lex integriert: <http://eur-lex.europa.eu/browse/summaries.html> (zuletzt abgefragt: 3. Februar 2016).

<sup>11</sup> ALEX, ROBERT, Rights, Legal Reasoning and Rational Discourse. In: Ratio Juris 5, S. 143–152 (1992).

## 4. Autorität des Autors

Wie FRITJOF HAFT richtig bemerkt hat, hängt jede juristische Aussage davon ab, wer diese gemacht hat.<sup>12</sup> Die Autorität des Autors sowie der Kontext sind entscheidend für die Glaubwürdigkeit der Quelle. Bei Parlamenten, Regierungen und Gerichtshöfen und ihren hochentwickelten Publikationsschienen ist dies offensichtlich und relativ einfach festzustellen. Bei Rechtsgelehrten und Praktikern ist dies eine schwierige Bewertungsfrage. Der Ruf des Autors unterliegt dem Wandel der Zeit; er muss aufgebaut werden und geht durch wenige oder keine Publikationen verloren. Nur wenige ausgewählte Autoren werden historisch und bleiben auch über ihren Tod hinaus relevant.

Grundvoraussetzung ist daher jeweils die korrekte Angabe des Autors oder des Autortyps (siehe oben) im zeitlichen Kontext. In vielen Fällen wird die Rechtswissenschaft die evidente Relevanz des Autors problemlos bestimmen können. Hier wird es in erster Linie darum gehen, die jeweils «wichtigeren» Dokumente von den «unwichtigeren Dokumenten», bestimmt für einen beliebigen Zeitpunkt, festzulegen.

Die neuen Medien ermöglichen einen intensiven Nutzerfeedback. Unter Beachtung des Datenschutzes können diese Daten für eine semi-automatische Bewertung der Relevanz des Autors bzw. dessen Dokumente verwendet werden. Die Kriterien hierfür sind: Anzahl der Aufrufe, Downloads, Ausdrücke, Verweildauer auf der Website, Lesedauer des jeweiligen Dokuments etc. Diese Daten können in einen Feedback-Faktor akkumuliert und sodann durch einen Experten bewertet werden.

## 5. Zeitlicher Kontext

Die Kriterien des zeitlichen Kontextes wurden von der Informationswissenschaft bzw. der Rechtsdokumentation bereits ausreichend aufgearbeitet: Datum der Erstellung, Datum der Publikation, Geltungsdauer, Anwendungsdauer, Zeitpunkt des Ereignisses, Daten anderer relevanter Dokumente etc.<sup>13</sup>

Bei den neuen Medien muss der zeitliche Kontext des Dokuments weniger umfangreich sein; aber das Datum der Erstellung der Erstfassung, allfälliger Ergänzungen bzw. auch der Diskussionsbeiträge werden auch hier zu dokumentieren sein.

## 6. Vernetzungsstruktur

Trotz der Unmenge von Dokumenten bemüht sich die Rechtswissenschaft um Konsistenz der jeweiligen Rechtsordnung. Das wesentlichste Element zur Unterstützung dieses Bemühens liegt in der Einordnung des Dokuments in die Vernetzungsstruktur der Rechtsordnung. Es werden explizit die Verweise des Dokuments auf andere Dokumente dokumentiert (Aktivverweise). Gesetze, Judikatur, juristische Fachliteratur etc. greifen auf ein Netzwerk bestehender Dokumente zurück.<sup>14</sup> Sehr wichtig ist Berücksichtigung der Verweise in anderen Dokumenten geworden (Passivverweise); in erster Linie spätere Literatur und Judikatur zu diesem Dokument.

Der juristische Textkorpus ist in seiner Gesamtheit selbst das Recht; zur Autorität benötigt er aber eine weitgehend konsistente Struktur. «Part of the reason for the great importance of citations in legal texts lies in

---

<sup>12</sup> HAFT, FRITJOF, Juristische Schreibschule, Anleitung zum strukturierten Schreiben. Edition Normfall, Frankfurt, S. 225 ff. (2009). Der Text «Das Zitieren von Autoritäten» erschien erstmal 1994.

<sup>13</sup> Vgl. SCHWEIGHOFER (1999), FN 1; SCHARF, JOHANNES, Künstliche Intelligenz und Recht, Von der Wissensrepräsentation zur automatisierten Entscheidungsfindung. OCG Verlag, Wien (2015).

<sup>14</sup> BERGER, ALBRECHT, Die Erschließung von Verweisungen bei der Gesetzesdokumentation. Verlag Dokumentation, München-Pullach (1971); JAHNEL, DIETMAR, Rechtsdatenbanken. In: Dietmar Jähnel, Peter Mader (Hrsg.), Rechtsdatenbanken – Internet. 2. Auflage. Österreichische Verlagsgesellschaft, Wien, S. 7–74 (2000).

the nature of legal texts themselves. There is no physical ‹legal object› that is described in texts, the texts themselves constitute the law.»<sup>15</sup>

Die Erfassung dieser Verweise kann weitgehend mit Listen bzw. sprachlinguistischen Methoden (semi-)automatisiert erfolgen;<sup>16</sup> offen bleibt dabei sowohl die Typisierung der Verweise als auch das Ranking. Es gibt eine sehr vielschichtige Praxis der Zitierformate; neuerdings gewinnt die standardisierte Form des ECLI Bedeutung.<sup>17</sup>

Die derzeit häufigste Typisierung der Verweise sind Aktiv- und Passivzitate. Eine genauere Klassierung ergibt folgende Typen: Vorarbeiten, Änderung/Aufhebung (aktiv oder passiv), Nacharbeiten, zitierende Judikatur bzw. Literatur; Verweise je nach Teil des Dokuments (Rechtsgrundlage, operativer Teil des Urteils, Begründung, etc.). Eine Besonderheit sind Begriffsverweisungen, vor allem innerhalb eines Dokumentes (z.B. Sonderausgaben im Sinne des § 18 Einkommensteuergesetzes). Verweistechiken können aber auch ein Dokument nahezu unlesbar machen.

Beim *citation vectors*-Projekt von COLIN TAPPER<sup>18</sup> wurden Zitierungen als Vektoren dargestellt, wobei diese nach ihrer Bedeutung gewichtet werden (Vektor-Retrieval-System VEXT). Mittels einer Clusteranalyse werden ähnliche Verweisungen festgestellt. Hauptproblem dieses Ansatzes blieb die Wissensakquisition für die Verweisungsvektoren.

Wichtig ist es sich vor Augen zu halten, dass Verweise nur ein Einstieg, eine Erleichterung der juristischen Analyse sein können. «Citation analysis is not a substitute or shortcut for critical thinking; it is, instead, a point of departure for those willing to explore the avenues to thorough evaluation.»<sup>19</sup>

Neben der Darstellung der Stellung des Dokuments im Dokumentennetzwerk der Rechtsordnung – ein wesentlicher Erkenntnisgewinn – sind Verweise ein wichtiges Instrument des Ranking.<sup>20</sup> Ausgangspunkt des Rankings ist eine Formalisierung der Verweise, mit typisierten Dokumentbezeichnungen und, wenn möglich, Berücksichtigung der jeweiligen Dokumentelemente. Dieses Datenmaterial kann durch Nutzerfeedback noch ergänzt werden. Sodann ist vornehmlich eine statistische und analytische Auswertung angesagt: Anzahl der Verweise insgesamt, Anzahl der jeweiligen Typen der Verweise, Zeitschiene der Verweise etc. Die Rechtsinformatik ergänzt diese Analyse durch die Einbeziehung von rechtswissenschaftlichen Relevanzkriterien: Autorität des jeweiligen verweisenden Autors im zeitlichen Kontext, Wichtigkeit der Verweisung im Kontext des jeweiligen Dokuments etc.

Bei den neuen Medien ist die Stellung im Dokumentennetzwerk entscheidend. Ohne Passivzitate eines Blogs (sei es auch nur in der Form des ‹Like›) oder Verweise auf die Homepage ist die Quelle – falls keine besondere Autorität hinzukommt – nahezu unsichtbar und wird kaum Relevanz entfalten können.

<sup>15</sup> GEIST, ANTON, The Open Revolution: Using Citation Analysis to Improve Legal Text Retrieval. In: Meritxell Fernandez-Barrera, Norberto Nuno Gomes de Andrade, Primavera de Filippi, Mario Viola de Azevedo Cunha, Giovanni Sartor, Pompeu Casanovas (Hrsg.), Law and technology: Looking into the future: selected essays. Florence, Italy: European Press Academic Publishing, S. 157–165 (2009). Vgl. auch SCHWEIGHOFER (1999), FN 1.

<sup>16</sup> Vgl. SCHWEIGHOFER (1999), FN 1.

<sup>17</sup> VAN OPIJNEN, MARC, European Case Law Identifier: Indispensable Asset for Legal Information Retrieval. In: Maria Angela Biasiotti and Sebastiano Faro (eds.), From Information to Knowledge – Online access to legal information: methodologies, trends and perspectives IOS Press, December (2011); Wikipedia DE, European Case Law Identifier, [https://de.wikipedia.org/wiki/European\\_Case\\_Law\\_Identifier](https://de.wikipedia.org/wiki/European_Case_Law_Identifier) (zuletzt abgefragt 3. Februar 2016).

<sup>18</sup> TAPPER, COLIN, An Experiment in Use of Citation Vectors in the Area of Legal Data, Complex 9, Universitetsforlaget (1982).

<sup>19</sup> GARFIELD, EUGENE, Uses and Misuses of Citation Frequency. In: Eugene Garfield (Hrsg.), Ghostwriting and other essays. Philadelphia: ISI-Press, S. 403–409 (408) (1986).

<sup>20</sup> GEIST (2009), FN 15.

## 7. Ranking

Bei Rankingverfahren werden die Suchergebnisse absteigend nach ihrer mutmaßlichen Relevanz ausgegeben.<sup>21</sup> Ranking ist ein Kern des Information Retrieval<sup>22</sup> und ist keine Spezialität der Suchmaschinen im Web. Aber erst mit der steigenden Bedeutung von Websuchmaschinen und der wachsenden Größe des Webs wurde ein Ranking der potenziell relevanten Treffer unverzichtbar.<sup>23</sup> Den Durchbruch von textstatistischen Verfahren zum intelligentem Ranking stellt der Erfolg des PageRank-Algorithmus bei Google Ende der 1990er-Jahre dar.<sup>24</sup> Damit konnte dem Wunsch der Nutzer nach einfachen, aber effizienten Suchwerkzeugen entsprochen werden, was ein wesentlicher Wettbewerbsvorteil war und ist.

Für juristische Textkorpora bedeutet dies, dass die Suchsysteme stärker auf die Rigidität der booleschen Logik verzichten und zu einer approximatischen Suche mit Fehlerkalkül wechseln.<sup>25</sup> Die fehlende Qualität der Suchanfrage wird durch effizientes Ranking «überspielt». Dieses spielt somit für die Effizienz dieser Suche eine entscheidende Rolle. In der Praxis ist die «Konkurrenz» zwischen Websuchmaschinen und klassischen Fachinformationssystemen am stärksten bei Textkorpora zu beobachten, die sowohl vom der eigenen Suchmaschine als auch von Google indexiert werden (z.B. EUR-Lex oder RIS).<sup>26</sup>

Leider ist die Relevanz bei juristischen Textkorpora eine schwer fassbare Größe. Die objektive Relevanz – gemessen durch die Relevanzquote (Precision) und Vollständigkeitsquote (Recall) einer Suchanfrage – ist bestimmbar; wird aber praktisch durch die subjektive Relevanz überlagert. Das Ranking soll das Suchergebnis verbessern, weil der Nutzer sein Informationsproblem oft nur vage beschreiben kann, und nur bei entsprechender Geübtheit dies in die rigide boolesche Suchlogik mit Abstandsoperatoren<sup>27</sup> umsetzen kann. Diese fast unlösbare Aufgabe wird dadurch noch erschwert, dass nicht nur eine objektive Relevanz nach Wichtigkeit gegeben werden soll (auch Kommentare sind nur oft nicht sehr klar zur Relevanz), sondern diese auch noch dem Informationsproblem des Nutzers angepasst werden. Dies bedeutet, dass der Fokus auch darauf gelegt werden muss, was der Nutzer noch nicht weiß, aber wissen sollte. Das Ranking wird damit ein wesentlicher Faktor für die Qualität des Rechtsinformationssystems.<sup>28</sup> Das traditionelle inverse chronologische Ranking der Suchergebnisse ist eine gar nicht so schlechte Lösung für diesen Zielkonflikt: Neueste Entscheidungen zu einem Thema werden als erste angezeigt, weil sehr wahrscheinlich ist, dass der Nutzer diese noch nicht kennt.

Der wesentlichste Vorteil des Ranking liegt darin, dass man auch bei großen Treffermengen noch zu brauchbaren Ergebnissen kommt. Der Nutzer muss nur eine überschaubare Menge von Suchergebnissen durchsehen. Relevanz- und Vollständigkeitsquote sind voneinander abhängig; es müssen relevante Dokumente bestmög-

---

<sup>21</sup> Vgl. dazu die mit sehr umfassenden Verweisen versehene Arbeit von GEIST, ANTON, Rechtsdatenbanken und Relevanzsortierung, Dissertation, Universität Wien, in Fertigstellung (2016). Diese war für die Erstellung dieses Kapitels sehr wertvoll. GEIST gilt auch der Dank, für die Zusammenstellung der vielen Zitate, die diese Fragestellung so treffend zusammenfassen. Vgl. auch SCHWEIGHOFER (1999), FN 1, S. 66; LEWANDOWSKI, DIRK, Web Information Retrieval: Technologien zur Informationssuche im Internet. Frankfurt am Main, DGI (2005).

<sup>22</sup> SALTON, GERARD, MCGILL, MICHAEL J., Information Retrieval – Grundlegendes für Informationswissenschaftler. McGraw-Hill, Hamburg (1987).

<sup>23</sup> LEWANDOWSKI (2006), FN 21, S. 73.

<sup>24</sup> LEWANDOWSKI, DIRK, Suchmaschinen verstehen. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 104 (2015).

<sup>25</sup> Dieser Trend wurde von LOSSAU bereits 2004 festgestellt. LOSSAU, NORBERT, Suchmaschinentechnologie und Digitale Bibliotheken – Bibliotheken müssen das wissenschaftliche Internet erschließen. In: ZfBB – Zeitschrift für Bibliothekswesen und Bibliographie 51, Nr. 5–6, S. 284–295 (288) (2004).

<sup>26</sup> Vgl. zum Wettbewerb zwischen Fachinformation und Web KRAFT, MATTHIAS: Metasuche. In: Erich Schweighofer, Franz Kummer (Hrsg.), Europäische Projektkultur als Beitrag zur Rationalisierung des Rechts: Tagungsband des 14. Internationalen Rechtsinformatik-Symposiums IRIS 2011. books@ocg, Wien 2011, 503–508 (505) (2011).

<sup>27</sup> Vgl. SCHWEIGHOFER (1999), FN 1.

<sup>28</sup> BAUER, INGMAR, Moderne Ranking-Verfahren im WWW: Grundlagen, Konzepte und Algorithmen. Saarbrücken, VDM Verlag Dr. Müller (2007).

lichst sortiert werden.<sup>29</sup>

Bei der Bestimmung der Relevanz ist es bedeutsam zu wissen, wie viele Dokumente der Trefferliste der Nutzer intellektuell bewerten kann und will. Juristische Dokumente sind relativ lang und aus den bibliographischen Angaben kann oft nicht die Nicht-Relevanz erkannt werden; d.h. ohne Einlesen geht es oft nicht. Das Anzeigen des Kontexts der Suchwörter ist eine wesentliche Hilfe; bei einigen Suchmaschinen wird großer Wert auf diese Informationskomprimierung gelegt.<sup>30</sup> Bei Praktikern ist es aus Zeitgründen nicht möglich, mehr als 20 Dokumente im Detail zu bewerten, und weniger ist mehr. Daher müssen die relevantesten Dokumente zuerst angeboten werden.

Qualitatives Ranking des Dokuments geht auf diese Ziele ein und entwickelt Verfahren, welche die gefundenen Dokumente absteigend nach ihrer mutmaßlichen Relevanz anzeigen. Damit wird der übliche Weg der chronologischen Trefferliste verlassen, bei welcher die Nutzer diese intellektuell sichten müssen. Die Treffer werden passend zur Relevanz sortiert.

Der Zielkonflikt bedingt, dass es keine alleinige richtige Lösung geben kann. Da die Relevanz selbst sehr unterschiedlich gesehen werden kann, soll eine Vielfalt von Optionen verfügbar sein, damit der Nutzer die passendste wählen kann. Hierbei ist zu beachten, dass der Nutzer selbst oft nicht in der Lage sein wird, explizit die jeweiligen Relevanzkriterien im Detail zu nennen. Aber es gilt wie beim Hypertext: wenn der Nutzer Freude am Arbeiten mit dem Dokumentkorporus hat, wird er die Varianten des Rankings ausprobieren und am Ende ein wesentlich besseres Suchergebnis bekommen. Die Grundfunktionalität der Sortierung nach Datum bzw. nach Relevanz ist der Einstieg.

Jedes Ranking setzt eine gute Kenntnis des juristischen Textkorporus, der Dokumenttypen, der Autoren, des zeitlichen Kontexts sowie der Verweisungsstruktur voraus. Weitere Möglichkeiten sind die datenschutzrechtlich schwierig zu lösende Berücksichtigung des Nutzerverhaltens sowie – insbesondere bei vielen Dokumenten – ein Ranking aufgrund von textstatistische Verfahren.

Ausgangspunkt der Textstatistik im Information Retrieval ist die evidente Tatsache, dass die Häufigkeiten der Wörter in Texten ungleich sind. Dadurch lassen sich Ähnlichkeiten berechnen als auch Clusterverfahren einsetzen. Die Schwierigkeit besteht darin, aus diesen Vektoren und Clustern auch inhaltliche Aussagen zu machen. Die deutsche Sprache mit vielen Regeln und Ausnahmen bedeutet eine Anzahl von grammatikalischen Versionen des gleichen Worts, aber auch durch Wortkomposita sehr unterschiedliche Begriffe mit gleicher Stammform. Bei einer Suche kann auf eine Auflösung dieser Komplexität – vor allem durch Grundformreduktion (Lemmatisierung)<sup>31</sup> – durch den geschickten Einsatz von Trunkierung verzichtet werden; bei der Textstatistik bedarf es zumindest einer Mindestberücksichtigung; das Problem von Negation, Einsatz von Pronomen etc. bleibt dabei ungelöst. Auch im juristischen Bereich beginnt man mit dem Ansatz von *Luhn*, wonach ein bedeutungstragender Begriff (im Recht vor allem ein Rechts- oder Tatsachenbegriff) für ein Dokument umso signifikanter ist, je häufiger er im Dokument auftritt.<sup>32</sup> Statistisch führt dies zur absoluten wie relativen Worthäufigkeit (*term frequency, TF* bzw. *within-document-frequency, WDF*) sowie zur inversen Dokumenthäufigkeit (*inverse document frequency, IDF*). TF spiegelt die Auftretenshäufigkeit eines Terms in einem einzelnen Dokument wieder; WDF relativiert diese zur Gesamtzahl aller Terme des Dokuments. IDF

<sup>29</sup> GAUS, WILHELM, Dokumentations- und Ordnungslehre: Theorie und Praxis des Information Retrieval. 5. Auflage, Berlin, Springer, S. 218 (2005).

<sup>30</sup> Z.B. bei der Lawsearch Enterprise der Weblaw AG: Stichworte, relevante juristische Verweise aus Gesetzgebung, Literatur und Rechtsprechung.

<sup>31</sup> Dies war in einer älteren Version von *juris* der Fall.

<sup>32</sup> LUHN, HANS PETER, The Automatic Creation of Literature Abstracts. In: IBM Journal of Research and Development 2, Nr. 2, S. 159–165 (1958).

berücksichtigt zusätzlich zum Trefferwert eines Begriffs die Häufigkeit seines Auftretens in der Dokumentensammlung.<sup>33</sup>

Forschungen im Rahmen der KONTERM-Projekte haben gezeigt, dass letztlich Juristen – im Gegensatz zu Journalisten – die Bedeutung von Aussagen ungenügend textstatistisch abbilden. Gerichte wiederholen nicht ihre Urteilsprüche; der Gesetzgeber erlässt die Norm nur einmal. Damit können Standardverfahren wie inverse Dokumentenhäufigkeit nicht so erfolgreich wie in anderen Bereichen eingesetzt werden.<sup>34</sup> «Term frequencies do not help as much in law as in other domains. No redundancy exists in legal norms, but a lot of information is irrelevant in case law. Relevant texts parts may consist only of a short paragraph or even only of a single sentence in a very long legal document.»<sup>35</sup>

## 8. Schlussfolgerungen

Ein starker Wandel kennzeichnet die juristische Veröffentlichung und Publikationskultur; neben sofortiger Verfügbarkeit ist die Ausrichtung auf einen «virtuellen Diskussionsprozess» im Internet ein weiteres Charakteristikum.

Eine Weiterentwicklung der Theorie zur Struktur von rechtswissenschaftlichen Wissenskorpora ist unumgänglich, um die neuen Formen relevanter juristischer Dokumente zu berücksichtigen, als auch um sich in der Unmenge diese Dokumente ausreichend zurechtzufinden. Es bedarf der Berücksichtigung von bibliografischen Angaben, der Metadaten, der Autoritätsbewertung des Autors, des zeitlichen Kontexts und des Verweisungsnetzwerks, um die Relevanz der Treffer eines Rechtsretrieval-Systems für eine Rechtsfrage nutzerorientiert sortieren zu können. Erst wenn diesen Formationen vorliegen, kann die inhaltliche Analyse der Dokumente mit Methoden der Computerlinguistik intensiv angegangen werden.

In diesem Beitrag wird ein theoretisches Modell entwickelt, welches in Zukunft unter Einbeziehung vorhandener Instrumente umgesetzt werden soll. Sodann sollen computerlinguistische Methoden zur weiteren Analyse der Dokumente eingesetzt werden.

---

<sup>33</sup> SPARK JONES, KAREN, ALBERS, CHRISTOPH, Eine statistische Interpretation von Begriffsspezifität und ihre Anwendung für das Retrieval. Saarbrücken, Fachrichtung Informationswissenschaft, Universität des Saarlandes, S. 14 (1987).

<sup>34</sup> SCHWEIGHOFER, ERICH, HANNEDER, GOTTFRIED, RAUBER, ANDREAS, DITTENBACH, MICHAEL, Improvement of Vector Representations of Legal Documents with Legal Ontologies. In: 5th International Conference on Business Information Systems, Poznan, April 2002 (Proceedings on CD-ROM) (2002).

<sup>35</sup> SCHWEIGHOFER, ERICH, GEIST, ANTON, Legal Query Expansion using Ontologies and Relevance Feedback. In: Casanovas, Pompeu; Biasiotti, Maria Angela; Francesconi, Enrico; Sagri, Maria-Teresa (Hrsg.), Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT). CEUR-WS.org, S. 149–160 (2007).