# Google the Law?!

Anton Geist

Arbeitsgruppe Rechtsinformatik, Universität Wien
Schottenbastei 10–16/2/5, 1010 Wien
anton.geist@univie.ac.at

**Schlagworte**: Legal Information Retrieval, Ranking, Google, PageRank, Zitationsanalyse

**Abstract:** Google hat durch Linkanalyse („PageRank“) die Websuche revolutioniert und einen erfolgreichen Weg durch den „(Web-)Datendschungel“ präsentiert. Rechtsdatenbanken wenden trotz vergleichbarer Ausgangssituation und guter Eignung noch keine ähnlichen Ranking-Mechanismen an. Ich zeige auf, dass ein Umdenken hier notwendig ist und Techniken der modernen Websuche oft als Vorbild dienen können.

## 1. Einleitung*

### 1.1 Google und die Revolution in der Websuche

„Googeln“ zu beschreiben, ist 2007 nicht mehr nötig: Jede und jeder, der diesen Beitrag liest, hat Google bereits genutzt, um Suchen im Internet durchzuführen. Das Fundament von Google stellt eine so genannte Volltextsuche dar, das heißt die Suchmaschine vergleicht die eingegebenen Suchbegriffe mit allen Webseiten, die von ihr gespeichert („indexiert“) wurden. Jene Webseiten, die die Suchbegriffe enthalten (oder die von anderen gespeicherten Webseiten mit den Suchbegriffen beschrieben werden), werden als Suchergebnisse angezeigt.

Den Erfolg von Google macht allerdings nicht diese Volltextsuche aus, sondern die – nach jeder Volltextsuche durchgeführte – Reihung der Trefferliste („Ranking“): Nur so können alle Webseiten, die die Suchbegriffe enthalten, in jene Reihenfolge gebracht werden, wegen der wir alle Google so ausgiebig benutzen. Heute können wir uns kaum vorstellen, dass es Zeiten gab, in denen Trefferlisten von Websuchmaschinen nicht einmal annä-

* Ich danke Erich Schweighofer für viele wertvolle Hinweise und seine umfangreiche Bereitschaft zur Diskussion.

hernd so treffsicher gereiht waren wie jene von Google: Im November 1997 lieferte nur eine der vier damals größten Websuchmaschinen bei einer Anfrage nach ihrem eigenen Namen ihre eigene Seite in den ersten zehn Treffern [Brin/Page 1998] als Ergebnis. Google hat die damals üblichen Suchtechnologien durch die Einführung von Linkanalyse zur Reihung der Suchergebnisse (genannt PageRank-Algorithmus) auf eine völlig neue Basis gestellt und neue Qualitätsstandards gesetzt [Langville & Meyer 2006].

## 1.2 Googles PageRank

Eine ganz einfache Überlegung begründet (zum überwiegenden Teil) den immensen Erfolg der Suchmaschine Google: Je mehr Links auf eine Seite verweisen, umso wichtiger ist dieser Seite. In weiterer Folge zählen auch Links, die von einer solchen (vermeintlich) wichtigeren Seite auf andere zeigen, bedeutender als andere.

Anders ausgedrückt: Umso mehr und umso wichtigere (weil selbst öfter zitierte) Webseiten auf eine andere Webseite verweisen, umso mehr Punkte sammelt die zitierte Seite. Diese gesammelten Punkte entscheiden dann, wie weit oben in der Trefferliste eine Seite erscheint, die die Suchbegriffe des Users oder der Userin enthält.

Wie es Google Gründer Sergey Brin einmal bei einem Vortrag vor Studierenden der Universität Berkeley im Herbst 2005 plakativ ausgedrückt hat: „Webseiten sind nicht gleich. Menschen schon, aber Webseiten nicht." (abzurufen zum Beispiel unter http://video.google.com/videoplay?docid=7137075178977335350)

Im Detail ist dieser PageRank-Algorithmus freilich geheim, hochkomplex und muss laufend verfeinert werden, um neue Probleme zu bewältigen (zB „Google Bombs"). Das Grundprinzip entspricht allerdings dem eben dargestellten.

## 1.3 Rechtsdatenbanken und die fehlende Innovation

Betrachten wir nun die österreichischen Rechtsdatenbanken, so hat sich die Entwicklung des Internets auf deren Zugänglichkeit selbstverständlich sehr positiv ausgewirkt: Die Rechtsdatenbank (RDB) und LexisNexis Recht Online sind als Volltextdatenbanken nunmehr über jeden Standard-Web-Browser zugänglich, lediglich die Rechts-Index & Datenbank (RIDA) arbeitet noch primär mit CD-ROMs.

In diesem Zusammenhang interessanter ist freilich, in wie fern die oben skizzierten neuen Techniken zur Reihung der Suchergebnisse (Ranking) auch von Rechtsdatenbanken eingesetzt werden.

Kurz gesagt: gar nicht.

Die Trefferlisten in den oben genannten Rechtsdatenbanken werden chronologisch geordnet (RDB [Unilösung]) oder nach Worthäufigkeiten (relative Häufigkeit der Suchbegriffe in den Texten; LexisNexis) gereiht. Alternativ bieten Anbieter die Möglichkeit, nach Autoren oder Gerichten, oder nach der Zahl der Fundstellen (RIDA) zu kategorisieren. Alle diese Reihungsmethoden haben sich im Bereich der Websuche als ineffizient erwiesen, um mit den stetig wachsenden Datenmengen umgehen zu können. Zurecht werden die Rechtsdatenbanken an dieser Stelle anmerken, dass man Websuche und die Suche in Rechtsdatenbanken nicht vergleichen könne. Zum Teil ist das sicherlich richtig, zum Teil ist ein Vergleich allerdings nicht nur möglich, sondern geradezu geboten. Eine genauere Analyse ist daher auf jeden Fall nötig.

## 2. Websuche ≠ Rechtsdatenbankenrecherche

Es wäre tatsächlich vermessen zu behaupten, Rechtsdatenbanken könnten „Google" – gemeint sind allgemein Methoden der modernen Websuche, insbesondere mittels Linkanalyse – einfach 1:1 kopieren. Zwei Unterschiede sprechen sofort dagegen: Die Redundanz der Informationen im Web sowie die viel höhere Einheitlichkeit der Textqualität bei Rechtsdatenbanken.

### 2.1 Redundanz

Im Internet sind viele Informationen (sehr) oft vorhanden (somit redundant) und machen es für Suchmaschinen somit leichter, den User oder die Userin mit relevanten Informationen aus verfügbaren Seiten zu bedienen. Suchen wir beispielsweise im Internet nach Informationen zu einem aktuellen Kinofilm, so gibt es in fast allen Fällen viele verschiedene Webseiten, die unser Informationsbedürfnis erfüllen können. Das führt im Ergebnis dazu, dass Websuchmaschinenbetreiber bei der Optimierung ihrer Systeme weniger „Recall" (Vollständigkeit der Treffer) in Kauf nehmen, wenn dafür eine höhere „Precision" (möglichst passende Treffer) möglich ist [Brin & Paige 1998].

Ganz anders ist die Erwartungshaltung, mit der Juristinnen und Juristen an eine Recherche in Rechtsdatenbanken heran gehen: Ein vollständiges Ergebnis aller Normen und relevanter Entscheidungen ist unbedingt erforderlich. Dafür wird es auch akzeptiert, viele Treffer manuell zu sichten. Schließlich sind doch an eine ungenügende Recherche unter Umständen sogar Haftungsfolgen geknüpft [Thiele 1998].

### 2.2 Qualitätsunterschiede der Texte

Rechtsdatenbanken bieten spezielle und gut redigierte Texte an. In jedem Fall erfolgt irgendeine Art von vorgelagerte Kontrolle, sei es durch Fachverlage (juristische Literatur) oder durch den Prozess der Textproduktion selbst (Gesetze, Entscheidungen). Gleichzeitig wird der Leidensdruck der NutzerInnen durch Handbücher und Kommentare vermindert, sodass sie nicht nur auf die Leistung von Rechtsdatenbanken angewiesen sind. Der Bedarf an ausgefeilter Suchtechnologie im Bereich der Rechtsdatenbanken ist daher sicherlich (noch) nicht so groß wie bei der Websuche.

## 3. Websuche = Rechtsdatenbankenrecherche

Umgekehrt drängt sich eine Verwendung von PageRank-ähnlichen Ranking-Mechanismen im Recht aus zwei ebenso logischen Überlegungen auf: der Entstehungsgeschichte von PageRank, sowie der Prominenz der Verweistechnik im Recht.

### 3.1 Vorläufer von PageRank

Ohne die Genialität von PageRank zu schmälern muss man feststellen, dass die Idee der Verwertung von Zitaten (und nichts anderes sind Links) zwischen Dokumenten nicht von Google stammt. Google hat vorhandene Technologien lediglich für das Webumfeld adaptiert und kombiniert. Sehr schnell landet man nämlich bei der Zitationsanalyse, wo die Überlegung „Wer oft zitiert wird, ist wichtig." schon seit vielen Jahrzehnten untersucht und angewendet wird. Auch wenn die (quantitative) Verweisanalyse im juristischen Bereich noch wenig üblich ist, so wird doch über zweckmäßige Einsatzmöglichkeiten spekuliert [Shapiro 2001]. Eine vermehrte Nutzung

von Passivzitierungen zur Evaluierung von juristischen Fachtexten scheint geboten, schließlich ist eine solche in anderen Fachrichtungen seit Jahrzehnten üblich (als prominentes Beispiel wäre die Zitationsdatenbank CiteSeer für frei zugängliche wissenschaftliche Informationen im Internet zu nennen: http://citeseer.ist.psu.edu/ beziehungsweise http://de.wikipedia.org/wiki/Citeseer).

### 3.2 Juristische Verweise im Textkorpus

Zu einem ganz ähnlichen Ergebnis kommt man, wenn man sich fragt, ob in der täglichen juristischen Arbeit nicht sowieso schon Anhaltspunkte zu erkennen sind, die für die Reihung von Suchergebnissen genützt werden könnten. Verweise sind eine Standardtechnik im Recht, hier muss man nur an den „Hohenecker-Index" denken. Bei diesem handelt es sich um ein in der Praxis laufend verwendetes Verzeichnis der Rechtsmittelentscheidungen sowie der juristischen Fachliteratur. Ob eine rein quantitative Gewichtung von Zitierungen, gerade im Bereich der Judikatur, sinnvoll ist, können freilich erst umfangreiche Tests zeigen. Es erscheint zweckmäßiger, für die Gewichtung Kommentare und Handbücher auszuwerten. Zumindest auf akademischer Ebene gibt es jedoch bereits seit vielen Jahren Bemühungen zur Miteinbeziehung von Verweisen als Hilfe für Suchsysteme im juristischen Bereich [Gelbart & Smith 1991; Rose 1994].

## 4. Schlussfolgerungen

Als Grundvoraussetzung sollten wir das Dilemma der unmöglichen Sichtung aller relevanten Dokumente – auch im juristischen Bereich – akzeptieren und uns nach technischen Lösungsmöglichkeiten umsehen. Technologien wie PageRank können hier über weite Strecken als Vorbild dienen, nachdem die Websuche schon vor mehr als 10 Jahren mit dem eben angesprochenen Problem der Informationsüberflutung konfrontiert war. Rechtsdatenbanken könnten in diesem Zusammenhang einen Google-haften Aufstieg erleben, wenn sie den Schritt von reinen Textprovidern zu Informationsbrokern wagen. Dafür sind freilich noch eingehende Forschungen nötig, um die moderne Websuche an die Bedürfnisse des Rechts anzupassen. Ich hoffe allerdings, mit diesem Beitrag aufgezeigt zu haben, dass neue Ranking-Methoden im Recht notwendig werden und die Websuche

hier ausgezeichnete Vorarbeiten geleistet hat, auf denen juristische Informationsanbieter aufbauen könnten.

## 5. Literatur


[Brin & Page 1998] Sergey Brin, Lawrence Page, *The anatomy of a large-scale hypertextual Web search engine*, Proceedings of the seventh international conference on World Wide Web 7, April 1998, Brisbane, Australia, 107–117.

[Gelbart & Smith 1991] Gelbart, Daphne and Smith, J.C., *Beyond Boolean Search: FLEXICON, a Legal Text-Based Intelligent System*, Proceedings of the Third International Conference on Artificial Intelligence and Law, Oxford, ACM Press, 1991, 225–234.

[Langville & Meyer 2006] Langville, A. N. and Meyer, C. D., *Google's Pagerank and Beyond: the Science of Search Engine Rankings*. Princeton University Press, 2006.

[Rose 1994] Daniel E. Rose, *A symbolic and connectionist approach to legal information retrieval*, Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 1994.

[Shapiro 2001] Shapiro, Fred R. *Collected Papers on Legal Citation Analysis*. Littleton, Colo: F.B. Rothman, 2001.

[Thiele 1998] Thiele, Clemens, *Die Zeitschriftenlektüre des Rechtsanwalts als haftungsrechtliches Problem*, ÖJZ – Österreichische Juristenzeitung 1998, 735.