

Unterausschuss Neue Medien (22)
Wortprotokoll *
30. Sitzung

Berlin, den 25.02.2013, 13:00 Uhr
Sitzungsort: Paul-Löbe-Haus
Konrad-Adenauer-Str. 1
10557 Berlin
Sitzungssaal: E.300

Vorsitz: Sebastian Blumenthal, MdB

TAGESORDNUNG:

Öffentliche Anhörung zu technischen Fragen eines Leistungsschutzrechts für Presseverlage

Experten:

Dr. Arnd Haller, Google Germany GmbH
Dr. Thomas Höppner, Rechtsanwalt, Vertreter des Bundesverbands Deutscher Zeitungsverleger
Prof. Dr. Dirk Lewandowski, Hochschule für Angewandte Wissenschaften Hamburg
Michael Steidl, International Press Telecommunications Council, London

Gesetzentwurf der Bundesregierung

Entwurf eines Siebenten Gesetzes zur Änderung des Urheberrechtsgesetzes

BT-Drucksache 17/11470

* Redaktionell überarbeitete Abschrift der Tonaufzeichnung

Anwesenheitsliste*

Mitglieder des Ausschusses

Ordentliche Mitglieder des Ausschusses

Stellvertretende Mitglieder des Ausschusses

CDU/CSU

Brandl, Dr. Reinhard
Jarzombek, Thomas
Schipanski, Tankred
Wanderwitz, Marco

Heveling, Ansgar

SPD

Dörmann, Martin
Klingbeil, Lars
Zypries, Brigitte

FDP

Blumenthal, Sebastian
Schulz, Jimmy

DIE LINKE.

Behrens, Herbert
Sitte, Petra, Dr.

BÜNDNIS 90/DIE GRÜNEN

Notz, Konstantin von, Dr.

*) Der Urschrift des Protokolls ist die Liste der Unterschriften beigelegt.

Bundesregierung

Linden	BMWi
Ernst	BMJ
Pakuscher	BMJ
Finkenberger	BMJ
Ernstschneider	BKM

Bundesrat

Denove	LV Bayern
--------	-----------

Fraktionen und Gruppen

Dunker	CDU/CSU
Scheele	DIE LINKE.
Braun	DIE LINKE.
Pohl	B90/GRÜNE
Morschhäuser	B90/GRÜNE
Volkmann	FDP
Kannapin	DIE LINKE.
Piallat	B90/GRÜNE
Romanowsky	CDU/CSU
Kollbeck	SPD
Schreiber	FDP
Plettenberg	CDU/CSU

Der Vorsitzende: Ich begrüße Sie zur 30. Sitzung des Unterausschusses Neue Medien zu technischen Fragen eines Leistungsschutzrechts für Presseverlage. So lautet der Titel unserer heutigen Anhörung, die öffentlich stattfindet und, wie bei uns im Unterausschuss Neue Medien allgemein üblich, per Livestream im Internet verfügbar ist. Alles, was hier von den Sachverständigen dargestellt wird, kann später über die Archivseite des Bundestages abgerufen werden und natürlich wird dann auch eine Videoaufzeichnung für diejenigen, die die Sitzung jetzt nicht per Livestream verfolgen können, bereitgestellt.

Die Obleute haben sich auf folgende Vorgehensweise verständigt: Die Beratungszeit soll 90 Minuten betragen und es soll unmittelbar mit Fragen der Fraktionen begonnen werden, also keine Eingangstatements geben. Mit dem Thema sind die hier Anwesenden allesamt schon seit Längerem intensiv beschäftigt, so dass wir ohne einführende Worte auskommen. Wir haben heute die Vertreter von vier Sachverständigenparteien eingeladen, denn Schwerpunkt ist, wie gesagt, der technische Bereich.

In alphabetischer Reihenfolge darf ich begrüßen: Herrn Dr. Wieland Holfelder, Technikdirektor von Google, Herrn Dr. Thomas Höppner, Rechtsanwalt, als Vertreter des Bundesverbands Deutscher Zeitungsverleger, Herrn Prof. Dr. Dirk Lewandowski, Hochschule für angewandte Wissenschaften, Hamburg, und Herrn Michael Steidl, International Press Telecommunications Council, London. Herzlich Willkommen. Wir haben ferner von Seiten der Bundesregierung aus dem Bundesministerium der Justiz, den zuständigen Unterabteilungsleiter, Herrn Dr. Christoph Ernst, bei uns. Ich begrüße ergänzend auch Herrn Dr. Arnd Haller als Vertreter von Google Germany. Da in der heutigen Anhörung der Fokus auf der Technik liegt, hatten wir uns darauf verständigt, dass Herr Dr. Holfelder in der ersten Reihe die Vertretung von Google übernimmt. Für den Fall, dass er sich rückversichern will oder Fragen nicht beantworten kann, gibt es die Verabredung der Obleute, dass Herr Dr. Haller auch zu Wort kommen kann.

Soweit zu den Regularien und zum Ablauf. Ich denke, ich kann mir weitere erläuternde Worte als Vorsitzender sparen. Wie Sie sicher wissen, hat die Anhörung im federführenden Rechtsausschuss eher die juristische Perspektive beleuchtet, weshalb sich eine Reihe von technischen Fragen ergeben haben, die nicht abschließend beantwortet werden konnten. Dafür ist heute hier im Unterausschuss Neue Medien Raum und Zeit. Das ist unsere Zielvorstellung.

Wir kommen dann auch direkt zu den Fragerunden. Ich bitte um Beachtung, dass eine Frage an zwei Sachverständige oder zwei Fragen an einen Sachverständigen gestellt werden können. Bitte halten Sie nach Möglichkeit keine Koreferate, sondern kommen Sie möglichst konkret auf den Punkt. Das gilt auch für die Antworten. Ich bitte Sie, präzise und prägnant zu antworten, damit wir die Zeit möglichst effektiv nutzen können. Wir beginnen mit der Fraktion der CDU/CSU. Herr Heveling, bitte.

Abg. Ansgar Heveling (CDU/CSU): Ich habe eine Frage an Herrn Dr. Höppner und zwar, ob es aus technischer Sicht zutreffend ist, dass die Nachrichtenaggregatoren den Webseiten der Presseverleger lediglich Leser zuführen, so wie Taxifahrer Restaurantbesitzern Gäste vor die Tür bringen. Eine zweite Frage habe ich an Herrn Steidl: Ist es technisch nicht einfach möglich, sich über robots.txt oder

Robot-Meta-Tags abzumelden und sich so der Verwertung durch Aggregatoren zu entziehen? Vielen Dank.

Der Vorsitzende: Für die SPD-Fraktion Herr Klingbeil, bitte.

Abg. Lars Klingbeil (SPD): Vielen Dank, Herr Vorsitzender. Ich habe eine Frage an Herrn Dr. Höppner. Es mag banal klingen, aber gleichwohl ist mir die Beantwortung der Frage wichtig. Klaut Google aus Ihrer Sicht Daten und hat damit ein Geschäftsmodell aufgebaut? Meine zweite Frage geht an Herrn Dr. Holfelder. Gibt es Ihrer Meinung nach heute bereits technische Möglichkeiten, dass Verlage Einfluss darauf nehmen, ob und dass etwas im Internet erscheint und in welcher Tiefe es einsehbar ist?

Der Vorsitzende: Für die Fraktion der FDP Jimmy Schulz, bitte.

Abg. Jimmy Schulz (FDP): Ich habe eine Frage an zwei Sachverständige und zwar an Herrn Dr. Holfelder und an Herrn Dr. Höppner. Es gibt ja immer wieder die Diskussion um die technischen Standards, das wurde ja gerade auch schon angeschnitten. Verlage haben die Möglichkeit, den Zugriff von automatisierten Suchmaschinen oder Aggregatoren zu beschränken bzw. ganz auszuschließen, zum Beispiel mit robots.txt. Es gibt aber auch erweiterte Möglichkeiten, zum Beispiel bietet das „nosnippet“-tag die Möglichkeit, Snippets ganz auszuschließen. Es gibt aber auch die Möglichkeit, in weiteren HTML-Tags bestimmte Vorkehrungen zu treffen, um festzulegen, welche Inhalte in welcher Art und Weise zu verwenden sind. Bieten diese Möglichkeiten Ihrer Einschätzung nach die Möglichkeit, genau das, was die Verlage ja auch wünschen, fein granuliert zu steuern und festzulegen, welche Inhalte in welcher Art und Weise von den Suchmaschinen oder Aggregatoren verwendet werden?

Würde zu diesem Zweck nicht schon eine Regelung genügen, dass Sie das anhand der gegebenen technischen Möglichkeiten vornehmen? Gebe es darüber hinaus Standards, mittels derer eine weitere Einschränkung bzw. feinere Einstellung der Wünsche der Verlage möglich wäre, die diesen helfen könnte?

Der Vorsitzende: Für die Fraktion DIE LINKE., Frau Dr. Sitte, bitte.

Abg. Dr. Petra Sitte (DIE LINKE.): Ich glaube, die Fragen haben eine Schnittmenge und das ist, finde ich, auch nicht schlimm. Die Verleger behaupten, robots.txt reiche nicht aus, benötigt würde eine maschinenlesbare Rechtesprache, wie sie etwa mit dem Automated Content Access Protocol (ACAP) zwar bereits entwickelt worden sei, aber eben noch nicht zur Anwendung komme. Damit könnten, das ist ja vorhin schon in der Frage von Herrn Klingbeil angeklungen, Copyright-Informationen und Nutzungsbedingungen maschinenlesbar übermittelt werden. Es könnte der Preis, es könnte die Dauer, ein Verbot für eine Nutzung usw. maschinenlesbar mitgeteilt werden. Ich habe den Eindruck, damit würden Suchmaschinen zu bloßen Schaufenstern für die Medienindustrie verwandelt. Ich möchte vor diesem Hintergrund fragen, welchen Aufwand die Umsetzung beispielsweise des ACAP-Standards für Google bedeuten würde und warum Google diesen Standard nicht adaptiert. Diese Frage richtet sich an Herrn Dr.

Holfelder. Meine zweite Frage geht an Herrn Prof. Dr. Lewandowski. Wie gesagt, meine Fragen haben eine Schnittmenge zu den Fragen der Kollegen, aber das liegt ja auf der Hand. Crawler müssen sich nicht an die Vorgaben von robots.txt halten, sie können sich auch darüber hinwegsetzen. Halten sich denn derzeit Suchmaschinenanbieter daran? Wie sind Ihre Erfahrungen und gibt es dazu eigentlich schon wissenschaftliche Untersuchungen?

Der Vorsitzende: Die erste Fragerunde wird komplettiert durch Abg. Dr. von Notz.

Abg. Dr. Konstantin von Notz (BÜNDNIS 90/DIE GRÜNEN): Vielen Dank, Herr Vorsitzender. Meine Frage richtet sich an die beiden Sachverständigen Herrn Prof. Dr. Lewandowski und an Herrn Dr. Holfelder. Wenn das Leistungsschutzrecht für Presseverlage nun tatsächlich kommt, ist es dann nicht unabdingbar notwendig, festzulegen, durch welche technischen Maßnahmen die Suchmaschinenanbieter und Newsaggregatoren Presseerzeugnisse automatisch erkennen?

Der Vorsitzende: Damit ist die erste Fragerunde komplett. Ich schlage vor, dass wir bei der Beantwortung alphabetisch vorgehen. Insofern bitte ich darum, dass Sie, Herr Dr. Holfelder, mit der Beantwortung der an Sie gerichteten Fragen beginnen und wir dann die Reihe fortsetzen.

Dr. Wieland Holfelder (Technikdirektor, Google Germany GmbH): Vielen Dank, Herr Vorsitzender. Ich bin Entwickler und mein Hintergrund ist die Softwareentwicklung. Vor 14 Jahren habe ich, als ich in Silicon Valley für ein Start-up-Unternehmen gearbeitet habe, einen Crawler implementiert und kenne mich deswegen auch mit robots.txt und dem Robots Exclusion Protocol (REP) einigermaßen aus. Es freut mich sehr, dass wir heute ein bisschen über die Technik reden können, die nämlich bereits schon seit vielen, vielen Jahren da ist und eigentlich auch sehr erfolgreich funktioniert.

Zu der Frage von Herrn Klingbeil, ob die vorhandenen technischen Möglichkeiten nicht bereits ausreichend sind und was wir da vielleicht noch erweitern können, kann ich sagen, dass robots.txt de facto ein Internetstandard ist, der seit vielen Jahren, nämlich seit fast zwei Jahrzehnten, existiert und es Webseitenbetreibern – auch Verlagswebseitenbetreibern – erlaubt, relativ feingranular zu bestimmen, welche Inhalte Crawler, die Suchmaschinen einsetzen, abgreifen dürfen. Die Granularität ist an dieser Stelle tatsächlich auf Webseitenebene individuell festlegbar. Ich kann zum einen sagen, ob eine Suchmaschine einen Inhalt crawlen, also erfassen darf, und Links auf dieser Webseite weiterverfolgen darf, um neue Inhalte zu finden. Auf der anderen Seite kann ich festlegen, ob ein Snippet für diese Webseite erzeugt werden soll. Ich kann weiter über sogenannte Meta-Tags festlegen, welche Informationen zum Beispiel eine Suchmaschine über eine Webseite anzeigen soll. Das funktioniert mit dem sogenannten Meta-Tag Description. Man kann in eine Webseite schreiben: Für diesen Text möchte ich gern, dass du, Suchmaschine, bestimmte Informationen anzeigst. Das kann die Suchmaschine übernehmen, sie muss es aber nicht, und das hat gute Gründe, denn manchmal macht es keinen Sinn, einen Text eins zu eins zu übernehmen. Zum Beispiel im Falle einer Suche, wenn ich nach einem bestimmten Schlüsselwort suche und die Webseite zwar das Schlüsselwort, aber die Zusammenfassung

das Schlüsselwort nicht enthält. In einem solchen Fall würde nämlich die Zusammenfassung angezeigt und der Nutzer den Kontext verlieren. Das zugrunde liegende Prinzip „Keyword in context“ ist ein sehr altes Prinzip, das von einem deutschen Wissenschaftler erfunden wurde und für Bibliotheksindizierungen und andere Dinge erfolgreich eingesetzt wird. Eine andere Möglichkeit wäre, Bilder auszuschließen, wenn ich möchte, dass auf einer Webseite Bilder nicht indiziert werden sollen. Ich kann auch sagen, dass ich verschiedene Dienste, also nicht nur verschiedene Suchmaschinen, sondern auch innerhalb der Suchmaschine verschiedene Dienste, unterscheiden möchte. Zum Beispiel kann ein Verlag festlegen, zwar von Google News ausgenommen zu werden, aber in der Websuche aufzutauchen. Das ist aufgrund der Unterscheidung durch den User-agent, der einen Crawler eindeutig identifiziert, möglich. Google setzt beispielsweise verschiedene Crawler für die verschiedenen Dienste ein. Ich habe als Handout (vgl. Anlage) ein Beispiel vorbereitet, mit dem wir Ihnen gern zeigen möchten, wie das aussieht. Es beinhaltet auch das Beispiel einer News-Suche, in der wir genau das übernehmen, was der Verlag uns vorgibt, das wir übernehmen sollen.

Ich habe eine Suche nach „Bild“ vorgenommen, indem ich „Bild“ eingegeben habe. Es kommt dann die Übersicht der Webseite www.bild.de. Wenn ich mir dann den HTML-Quelltext, der darunter liegt, ansehe, dann sehe ich einen Meta-Tag, der Description heißt. Da steht dann das Snippet drin, das wir hier anzeigen. Das heißt, wir übernehmen die Empfehlung, die uns www.bild.de an der Stelle gegeben hat, indem wir schreiben: Bild.de, die Seite 1 für aktuelle Nachrichten und Themen, Bilder und Videos aus den Bereichen News, Wirtschaft, Politik, Show, Sport und Promis. Das Beispiel zwei in den Unterlagen zeigt dasselbe für „Spiegel“. Man kann sehen, dass diese Technologie weit verbreitet ist. Interessant ist vielleicht noch das dritte Beispiel. Dort habe ich eine Google-Websuche mit den Schlüsselworten „Nordkorea warnt USA“ vorgenommen. Es erscheint dann ein Bericht von Spiegel-Online mit einem kleinen Snippet, das sagt: „Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer Befehlshaber hat einen US-General telefonisch vor einer Eskalation der Lage ...“.

Bezeichnend ist, dass das die Google-Websuche ist und hier keine Werbeflächen genutzt oder angezeigt werden. Dies ist der Fall, weil es keine relevante Werbung für dieses Thema gibt. Eines unserer Kriterien ist, Werbung nur dann zu zeigen, wenn sie für den Benutzer relevant ist. Bei Nachrichten trifft das in den seltensten Fällen zu. Wenn Sie dieselbe Suche auf Google News machen, sehen Sie das gleiche Ergebnis: Keine Werbung – auf Google News wird grundsätzlich keine Werbung angezeigt – und auch wieder dasselbe Snippet. Klickt man dann auf den Link, um ein bisschen mehr darüber zu erfahren, so gelangt man dann zum tatsächlichen Artikel. Auf Seite 7 des Handouts sehen sie den Artikel. Hier nun haben wir vier verschiedene Werbeflächen, die von dem Webseitenbetreiber zur Refinanzierung genutzt werden. In dem blauen Kästchen kann man die Zusammenfassung sehen, die deutlich länger ausfällt als unser Snippet. Als letztes sehe ich mir schließlich den Quelltext der Webseite, den HTML-Text, an. Dort gibt es wieder einen Meta-Tag namens Description, der festlegt, dass in dem Fall, dass jemand auf diesen Artikel geht, der Suchmaschine empfohlen würde, die Zusammenfassung zu verwenden. Der Text, den Sie dort finden, ist der Text, der sowohl in der News-Suche als auch in der Web-Suche angezeigt wird. Das heißt, wir reden ja schon miteinander und es funktioniert eigentlich wunderbar.

Genauso könnte jetzt hier geschrieben stehen: Ich möchte kein Snippet anzeigen. Oder zum Beispiel: Ich möchte die Seite gar nicht im Index haben, so dass sie überhaupt nicht in der Suche auftaucht. Das waren jetzt die kleinen Beispiele für die News-Suche. Darüber hinaus sind noch zwei Seiten angefügt, die zeigen, wie einfach robots.txt oder auch Meta-Tags tatsächlich funktionieren. Man kann also relativ granular festlegen, was ein User-agent, also Crawler, eines Suchdienstes auf welcher Seite vornehmen darf. Das kann auf Verzeichnisebene stattfinden, auf individueller Seitenebene geschehen, mit sogenannten Wildcards bestimmte Bereiche betreffen und so weiter. Von daher sage ich: Die Möglichkeiten sind da und sie werden erfolgreich genutzt. Wir haben an dem Beispiel gesehen, dass wir uns auch daran halten. Von daher sehe ich überhaupt keinen Bedarf, diesbezüglich in irgendeiner Form gesetzlich einzugreifen.

Dann möchte ich noch die Frage, die Herr Schulz gestellt hat, aufgreifen. Über die erweiterten Möglichkeiten haben wir schon ein bisschen gesprochen. Der zweite Teil ihrer Frage betraf Standards, die darüber hinaus vielleicht noch notwendig wären, um weitere Dinge zu tun. Diesbezüglich sind wir immer offen, denn so funktioniert nun einmal das Internet und die Standards entwickeln sich mit der Zeit. Es gibt zum Beispiel ein Meta-Tag, das „unavailableafter“ heißt. Es legt fest, bis wann ein Inhalt sichtbar sein soll, den eine Suchmaschine finden soll. Dieses Tag gab es am Anfang von robots.txt nicht. Er wurde im Laufe der Zeit nach Diskussionen auch auf internationaler Ebene mit Webseitenbetreibern und Verlegern eingeführt. Auf diese Art und Weise funktionieren internationale Standards, denn sie sind permanent in Bewegung und werden angepasst. Es macht wenig Sinn, einen Status quo einzufrieren. Von daher lautet meine Antwort: Ja, wir können uns vorstellen, dass es da durchaus noch Möglichkeiten gibt, das eine oder andere zu erweitern und es gibt viele Beispiele, dass es in der Vergangenheit so bereits auch funktioniert hat.

Dann zur Frage von Frau Dr. Sitte, die die Rechtesprache ACAP angesprochen hat. Wie gesagt, es gibt Aspekte in ACAP, die wir inzwischen auch in robots.txt sehen. „Unavailable after“ ist eines der Beispiele. Es gibt andere Dinge in ACAP, die aus Sicht der Verleger, da sie den positiven Nutzen im Sinne haben, vielleicht etwas außer Acht gelassen werden, aber man muss dabei eben auch bedenken, dass im Internet auch Spammer, Schurken und Leute, die das Internet für ihre böartigen Zwecke nutzen wollen, unterwegs sind. Insofern gibt es eben einige Formulierungen in ACAP, die an der Stelle für uns durchaus schwierig umzusetzen sind, ohne den Nutzen, den unsere Nutzer im Internet sehen, stark einzuschränken. Das ist zum einen die Vorgabe, dass ein Snippet bindend vorgegeben wird. Das heißt, wir können dann eben nicht entscheiden, ob wir es übernehmen, sondern müssen das tun. Stellen Sie sich die Situation vor: Wir haben eine Websuche abgesetzt für eine Schauspielerin und in den Ergebnissen kommt dann eine Reihe von Links auf Webseiten, in denen die Schauspielerin erwähnt wird. Einer der Webseitenbetreiber hat Böses im Sinne und zielt darauf ab, in seiner Beschreibung zu suggerieren, es seien alle relevanten Informationen zu der Schauspielerin zu finden und es würden die besten Informationen dazu geliefert. Klickt man nun darauf, landet man auf einer Porno-Webseite, einer Malware-Webseite oder einer Spam-Webseite und der Suchmaschinenbetreiber darf es nicht einschätzen, obwohl er es vielleicht könnte, weil ihm der Snippet an dieser Stelle bindend vorgegeben wird. Das heißt, neben den wirtschaftlichen Aspekten, wo Sie sagen, dass sei „das Schaufenster der Verlage“, kommen solche Aspekte noch hinzu und man sollte sie einfach nicht unterschätzen. Wir müssen mit solchen Vorgängen

tagtäglich in sehr, sehr hohen Fallzahlen umgehen. Von daher sind wir zurückhaltend, die Entscheidung, was in ACAP steht, nur einer Interessengruppe zu überlassen, denn das Internet hat einfach sehr viele verschiedene Interessengruppen. Insofern ist es sicherlich besser, sich anzuschauen, was es bereits gibt und was auch funktioniert, nämlich robots.txt und REP und dann sehen wir, ob es vielleicht noch Dinge gibt, die wir zusätzlich hinzunehmen müssen, um das abzurunden. Aber, wie gesagt, das machen wir ja sowieso und haben auch in der Vergangenheit gezeigt, dass es funktioniert.

Herr Dr. von Notz hat danach gefragt, wenn das Leistungsschutzrecht komme, ob es dann nicht unabdingbar notwendig sei, über technische Möglichkeiten zu verfügen, um zwischen Verlagswebseiten und anderen Webseiten zu unterscheiden. Die Situation stellt sich momentan so dar, dass, wenn ich mit meinen algorithmischen Methoden bzw. mit Heuristiken die Suche angehe und vermute, eine Seite könnte eine Nachrichtenwebseite sein und sie entsprechend klassifiziere, sich dann aber konkret, weil jemand danach sucht, herausstellt, dass es keine Nachrichtenseite war, weil mein Algorithmus falsch war, dann ist das natürlich kein schönes Produktergebnis, aber es ist auch keine Rechtsverletzung. Das Leistungsschutzrecht hingegen ist als Verbotrecht aufgesetzt. Bei einem Verbotrecht reichen Heuristiken nicht mehr aus, sondern man benötigt eine bindende Liste von Verlagen, die unter dieses Gesetz fallen. Das heißt, entweder muss es eine Liste geben, die uns in irgendeiner Form zukommt, wobei es auch der Klärung bedarf, wer sich „Verlag“ nennen darf und wer darüber entscheidet, ob das so ist. Auf technischer Ebene fällt es uns sehr schwer, auseinanderzuhalten, ob sich um ein Dokument handelt, welches auf der Privatseite eines Journalisten steht, der das Recht an diesem Dokument hat und es veröffentlicht, aber kein Verleger ist, oder um eine Verlagswebseite, die genau das gleiche Dokument veröffentlicht, aber im Kontext eben eines Verlages und dann plötzlich unter das Leistungsschutzrecht fällt. Rein technisch gesehen gibt es von meiner Seite aus einfach keine Möglichkeit, diese beiden zu unterscheiden. Von daher wäre das von unserer Seite auch eine der Schwierigkeiten. Sollte die Regelung kommen, müssen wir im Zweifel sehr konservativ vorgehen, damit wir keine Rechtsverletzung verursachen.

Der Vorsitzende: Vielen Dank für die Ausführungen. Wir fahren dann mit Herrn Dr. Höppner fort.

Dr. Thomas Höppner (Rechtsanwalt, Vertreter des Bundesverbands Deutscher Zeitungsverleger):

Sehr geehrte Damen und Herren Abgeordnete, im Namen des Bundesverbands Deutscher Zeitungsverleger, für den ich heute hier sprechen darf, danke ich Ihnen für die Einladung und die Möglichkeit, zu technischen Aspekten Stellung zu nehmen.

Die Frage von Herrn Heveling zielte vermutlich auf die aktuelle Google-Kampagne des Vergleichs mit einem Taxi ab. Dahinter verbirgt sich aber eine ganz zentrale und vielleicht die entscheidende Frage, nämlich die nach der Rechtfertigung eines Leistungsschutzrechts. Um es vorwegzunehmen, ich bin kein Fan des Vergleichs, das haben Sie wahrscheinlich vermutet, denn ich kenne einfach kein Taxi, das sich zunächst bei allen Restaurants bedient, um dann einem Kunden, der ein Restaurant sucht, direkt im Taxi das Gericht zu servieren. Newsaggregatoren würden sich selbst auch nie als Taxi im Internet ansehen

wollen. So würde der Business-Case nicht funktionieren und so würden sie niemals das Venture-Capital bekommen, das gerade in den Markt fließt. Vielmehr präsentieren sich Newsaggregatoren ja gerade, um im Vergleich zu bleiben, als die zentralen Restaurantmeilen, wo Sie, wenn Sie so wollen, die Köstlichkeiten der ganzen Welt serviert bekommen und diese auch gleich konsumieren können. Insofern blendet der Vergleich das Wesentliche aus. Er unterstellt, dass Aggregatoren 100 Prozent der User, die nach Presseerzeugnissen suchen, auch an Verlagsseiten weitergeben würden, und dass Verleger für die Weiterleitung, quasi für einen Dienst, nunmehr Geld verlangen wollen. Beides trifft aber nicht zu. Das Leistungsschutzrecht soll ja vor solchen Diensten schützen, die den Verlegern, wenn man sie zusammengenommen betrachtet, nicht zusätzlichen Traffic bescheren, sondern diesen geradezu wegnehmen. Sie nehmen Traffic weg, indem sie mit den Leistungen, die sie von den Verlagsseiten abgreifen, Verbraucher auf ihre eigenen Homepages ziehen und ihnen dort das Informationsbedürfnis, das sie haben, unmittelbar befriedigen.

Insofern muss man sich auch klar für diesen Vergleich von der Illusion bzw. der Vorstellung lösen, ein Taxi mit der normalen Google-Suche vergleichen zu können, wo man einen Suchbegriff eingibt und einem auf diese Suchanfrage dann verschiedene Webadressen vorgeschlagen werden. So werden ja Nachrichten im Internet nicht konsumiert und so funktionieren auch Newsaggregatoren nicht. Bei typischen Newsaggregatoren finden Sie die Nachrichtentexte automatisch auf der Homepage voreingestellt, mehrere untereinander. Sie werden nach Rubriken oder Kategorien, die der User vorher vorgibt, systematisiert, mit anderen Inhalten kombiniert und um Vorschaubilder ergänzt. Dann werden aus ganz anderen Kontexten weitere Texte hinzugefügt. Alles nur, um dem User eine eigenständige User-Experience zu bieten, die mit einer Suche eigentlich gar nichts mehr zu tun hat. Deshalb ist der Vergleich mit einer Suchanfrage, den man jetzt vielleicht ziehen wollte, gar nicht mehr gegeben. Heute zielt doch das Geschäftsmodell vieler Newsaggregatoren geradezu darauf ab, mit Hilfe der Verlagsleistungen eigene Kundenbeziehungen zu Lesern aufzubauen und diese vom Anklicken der Verlagsseiten abzuhalten, indem sie bereits auf ihrer eigenen Website die Informationen geben, die Leser eigentlich haben wollen. Deshalb haben wir es, anders als beim Taxi und beim Restaurant, nicht mit einer Symbiose oder einem Komplementärverhältnis zu tun, wie das manchmal dargestellt wird, sondern wir haben hier ein direktes Konkurrenzverhältnis. Aggregatoren- und Verlagsseiten bedienen dieselbe Nachfrage derselben Interessenten und leider derzeit mit denselben Inhalten, die der eine teuer produziert und der andere billig kopiert.

Für die Antwort auf Ihre Frage ist insofern auch relevant, wie Nachrichten heute eigentlich gelesen werden. Studien zeigen, dass Verbraucher in erster Linie einen schnellen Überblick gewinnen und nichts verpassen wollen. Für dieses Interesse reicht es aber oft schon aus, wenn man allein die Überschriften und die Zusammenfassungen liest. Warum genügt es uns denn oftmals, dass wir Zeitungen durchblättern und überfliegen und trotzdem einen Informationsgewinn haben? Warum verbringen User im Durchschnitt nur 30 Sekunden auf einer Verlagsseite um sich zu informieren? Weil dieser kurze Blick genügt, um das aktuelle Bedürfnis, zu schauen, was gerade auf der Welt läuft, zu befriedigen. Nun kann man fragen, ist das vielleicht ein Beweis für die mangelnde Bedeutung der Berichterstattung im Internet? Nein, im

Gegenteil, diese Komprimierung auf das Wesentliche hat einen eigenständigen ökonomischen Wert. Nur wird derzeit dieser Wert von den Aggregatoren abgefasst, wenn sie nämlich systematisch und zeitnah selbst über die aktuellen Ereignisse berichten, die sie gerade von Verlagsseiten übernommen haben. Insofern muss man sich auch klar vor Augen halten, dass die weitere Botschaft, die auch in der Werbeanzeige enthalten ist, nämlich: Wir, Google, schicken doch jeden Monat vier Milliarden Klicks an die Verlage, das Wesentliche ausblendet. Verschwiegen wird nämlich, dass monatlich auf Google 100 Milliarden Suchanfragen gestartet werden. Das heißt, die vier Milliarden würden gerade einmal vier Prozent entsprechen. Gleichzeitig hat Google an anderer Stelle gesagt, dass Verlagsseiten acht Prozent der Suchergebnisse ausmachen. Man kann sich auch fragen, warum bisher nicht mitgeteilt wurde, wie viele User eigentlich auf Google News gehen, wie lange sie dort verweilen, wie viele von ihnen dann tatsächlich noch durchklicken, wann sie zurückkommen und wie lange sie dann wieder bei Google News verweilen. Genau diese Informationen kann aber nur der Aggregator liefern. Die Verlage können das nämlich nicht messen, sie messen lediglich die Klicks, die zu ihnen durchdringen. Auf Ihre Frage würde ich also ganz eindeutig antworten: Nein, viele Aggregatoren erhöhen nicht den Traffic zu den Verlagsseiten, sondern nehmen ihn geradezu weg. Sie befriedigen selber einen Teil des Informationsbedürfnisses, nämlich den nach aktuellen Informationen, und leiten lediglich den Teil der User, der sich noch für einzelne Details interessiert, an die Verlagsseiten weiter. Einen großen Teil der Aufmerksamkeit nehmen die Aggregatoren folglich für sich selbst ein.

Das technische Problem daran ist, dass Verlage derzeit keine hinreichenden Instrumente haben, um sich dagegen zu wehren. Ihnen bleibt vielmehr faktisch nur übrig, das Spiel, das ihnen aufgedrängt wird, diesen Wettbewerb, den Aggregatoren erzeugen, indem sie vorgeben, ausschließlich nach Qualität und Renommee zu ranken, mitzuspielen. Nur deshalb optimieren sie ihre Seiten für Aggregatoren und schreiben Meta-Descriptions hinein, damit sie, wenn sie schon darauf angewiesen sind, wenigstens von diesem kleinen Topf, der ihnen noch verbleibt, also den Klicks, die durchgeleitet werden, die besten Ergebnisse bekommen und so am meisten vom Rest des Kuchens bekommen. Das ist der Grund, warum Webseiten optimiert werden.

Der Kampf um den Rest darf nun wirklich nicht verwechselt werden mit einer generellen Einwilligung, dass Aggregatoren Verlagsleistungen nutzen sollten. Insofern ist da auch kein Widerspruch vorhanden. Aggregatoren, die ungefragt Verlagsseiten abgreifen und ihren Nutzern vermitteln: Wir positionieren diese Verlagsseiten allein nach Qualität, Renommee und Relevanz, zwingen Verlage doch geradezu, den ihnen aufgedrängten Wettbewerb ernstzunehmen. Verleger hätten ansonsten doch einen erheblichen Schaden in Bezug auf ihr Renommee in Kauf zu nehmen, wenn sie nämlich plötzlich überhaupt nicht mehr bei den Aggregatoren erscheinen, diese aber der ganzen Welt erklären, dass sie die Suchergebnisse nach Renommee und Qualität ordnen. Dann würden sie einen Schaden erleiden, der zudem weit über die Onlinewelt hinausginge. Und das kann doch von Verlagen nicht erwartet werden. Aus diesem Grund denke ich, gerade weil die technischen Hilfsmittel keine Lösung bieten, muss der Rechtsrahmen sicherstellen, dass hinreichende Anreize geschaffen werden, damit Verlage weiterhin Inhalte für User

kostenlos ins Netz stellen. Das ist eine typische Aufgabe fürs Urheberrecht. Und deshalb unterstützen die deutschen Verlage auch den Gesetzentwurf.

Zur zweiten Frage, ob Google Daten klaut. Ja selbstverständlich hat Google damit ein Geschäftsmodell aufgebaut. Warum gibt es denn Google News? Eben wurde gesagt, Google News schalte keine Werbung. Was dabei ausgeblendet wird, ist, dass selbstverständlich mit Google News Geld verdient wird. 2007 hat Marissa Mayer im Interview gesagt, Google News sei jährlich 100 Millionen Dollar wert. Das war kurz nachdem Google News überhaupt an den Start gegangen war. Sie erklärte, dass man Google News als einen Zulieferer ansehen müsse, denn User gingen dort hin, wo die Informationen zu finden sind.

Wenn ein Aggregator ein tolles Nachrichtenangebot hat, lockt das natürlich die User an. Sind die User schon einmal auf der Plattform, suchen sie nicht nur nach Nachrichten, sondern auch nach ein paar Produkten, gehen wieder zurück zur Websuche und klicken hin und her. Nicht umsonst verkauft Google ja Google News als integralen Bestandteil der Suche. Es lässt sich schwer behaupten, dieser integrale Bestandteil sei nun etwas ganz anderes, da dort keine Werbung geschaltet werde, denn es ist ja doch alles ein System. Und in diesem System wird selbstverständlich Werbung geschaltet, womit Google auch sehr viel verdient. Deshalb denke ich, dass das mit dem Geschäftsmodell leicht zu beantworten ist. Und schauen Sie sich doch einmal Google Knowledge Graph, einen weiteren Dienst, an, der mit der eigentlichen Suche an sich nicht mehr viel zu tun hat. Sie geben einen Begriff ein, zum Beispiel eine Persönlichkeit, und schon kumuliert Google verschiedenste Informationen aus verschiedensten Webseiten und macht ein völlig neues Produkt daraus, das dann mit den Quellen, aus denen es sich bedient hat, nicht mehr viel zu tun hat. Sie können dann auch gar nicht mehr weiterklicken, weil Google Ihnen die Antwort schon direkt gegeben hat. Sie finden bestenfalls vielleicht unten irgendwo einen Link, aber zusammengesetzt ist das ganze Produkt aus mindestens zehn verschiedenen Quellen.

Die dritte Frage von Herrn Schulz zielte direkt auf die technischen Möglichkeiten und er fragte, ob es nicht ausreichen würde, aus diesen einen gesetzlichen Standard zu machen. Das ist natürlich verknüpft mit der Frage, was denn die Technik überhaupt leisten kann. Auch wenn hier alles gerade so rosig dargestellt wurde, so rosig ist es in Wirklichkeit gar nicht. Man muss sich nur vergegenwärtigen, wie robots.txt und Meta-Tags in Wirklichkeit funktionieren. Erster wichtiger Punkt ist, wir haben es hier mit einer freiwilligen Kooperation zu tun. Natürlich, die großen Suchmaschinen halten sich daran, das ist ja auch gut so. Aber, wie viele Aggregatoren gibt es, gerade im kleinen Bereich, denen es einfach egal ist? Und natürlich gibt es die Differenzierungen, die Google anbietet. Aber lesen die anderen Aggregatoren diese Differenzierung auch? Nein. Jeder kann die Meta-Tags und die robots.txt so lesen, wie er will. Sie müssen sich das wie ein großes Stadion vorstellen, vor dem ein Türsteher steht, der darüber bestimmt, wer rein darf und wer nicht. Nun gehen aber manche nicht durch den Haupteingang, sondern durch einen der vielen Seiteneingänge. Manchen ist auch einfach die Anweisung des Türstehers egal. Und selbst wenn man sich jetzt an den großen Suchmaschinen orientiert, die ja mit Variablen wie „nosnippet“-Tags vielleicht eine Differenzierung vorgegeben haben, dann sind, wenn man genauer hinsieht, auch diese Differenzierungsmöglichkeiten nicht ausreichend hilfreich. Sie müssen sich einmal vorstellen, dass Sie jeden einzelnen Crawler von

jedem Aggregator separat instruieren müssen. Sie wissen aber gar nicht, was der Crawler eigentlich mit ihrem Content macht. Sie wissen nicht, für welchen Dienst dieser Crawler eigentlich sucht. Sie wissen nicht, welche Webseiten der Crawler eigentlich bedient. Es gibt viele Crawler, die bedienen aus demselben Index verschiedene Produkte und Webseiten. Dann mögen Sie vielleicht das eine Produkt, weil es ihnen tatsächlich Traffic bringt, das andere aber nicht, weil es einfach nur darauf abzielt, Ihnen Konkurrenz zu machen. Sie können technisch aber nicht sagen: Ja, diesen Crawler möchte ich für den einen Fall zulassen, für den anderen Fall aber nicht. Sie können als kleinsten Adressaten eben nur den Crawler adressieren, wissen aber nicht, was der Crawler mit ihren Inhalten macht.

Als weiteres Beispiel können wir den „nosnippet“-Meta-Tag anschauen, vielleicht auch noch kombiniert für Google News. Google sagt: Sie können doch separat Google News adressieren und dann sogar sagen, nur Snippets sollen dort nicht angezeigt werden. Was dabei aber ausgeblendet wird: Wer nicht bei Google News erscheint, wer also den Googlebot „News“ sperrt, der erscheint auch nicht in den extrem wichtigen News-Universals in der ganz normalen Websuche. Wenn Sie auf www.google.de nach Nachrichten suchen, dann bekommen Sie als erstes einen Universal angezeigt, also einen separaten Block, der mit „News zu“ überschrieben ist. Von da kommt der meiste Traffic, weil Google natürlich damit impliziert, dass man in diesem Block das Wichtigste findet, wenn man nach Nachrichten sucht. Dieser Universal bezieht aber seine Informationen von Google News, also über den Googlebot „News“. Wenn man nun aber Google News sperrt, erscheint man nicht mehr im wichtigsten Block der Google Websuche, also der ganz normalen Suche. Das aber kann sich niemand leisten, weil natürlich die Google Websuche in Relation gesehen viel wichtiger ist, als Google News. Und deshalb sagen alle Verlage, dass sie nicht Google News sperren können, wenn sie dann aus den News-Universals herausfallen.

Oder nehmen Sie das Beispiel der Snippets. Sie sperren einen Snippet. Was passiert? Ihr Ergebnis ist kleiner als das der anderen, weil die anderen größer und schöner positioniert werden. Dann klicken alle auf das größer aussehende, weil alle glauben, das ist relevanter, und schon sagt der Algorithmus: Oh, ein Ergebnis, wo das Snippet fehlt, das kann ja nicht wichtig sein, da hat ja niemand draufgeklickt. Also sagt der Algorithmus: Unwichtig, das Ergebnis nehmen wir wieder weg.

Der wichtigste Aspekt aus meiner Sicht an dieser ganzen Diskussion um robots.txt und Meta-Tags ist also, dass die Aggregatoren es letztlich selbst in der Hand haben, was sie aus der Kommunikation machen. Selbstverständlich können die Aggregatoren jede Sperre von Inhalten zum Anlass nehmen, um zu sagen: Na gut, du magst uns nicht, wir mögen dich auch nicht. Also werten wir eine Sperre von Inhalten so, dass deine Webseite einfach irrelevant ist. Und dann wird einfach massiv damit bestraft, dass man den Seiten die Qualität abspricht, die sie eigentlich haben.

Nehmen Sie einmal das Extrembeispiel: Sie haben die hochwertigste Verlagsseite mit den tollsten Artikeln und benutzen robots.txt. Wird man Sie jetzt finden? Nein. Selbst wenn die Verlagsseite das Verlinken zulässt, aber keine Snippets anzeigen lässt: Wird man Sie finden? Nein. Die Aggregatoren werden ihre Wertung daraus machen, Inhalte hinter einem Link, auf den keiner klickt, können nicht viel wert sein, die

zeigen wir nicht an. Und was entsteht dadurch? Ein Druck! Die Verleger werden hier untereinander ausgespielt. Derjenige, der Inhalte sperrt, wird automatisch gegenüber denen, die nicht sperren, benachteiligt. Damit spielen natürlich die Aggregatoren und erzeugen einen Druck: Machst du nicht mit, zeige ich den Snippet bzw. das Ergebnis der Konkurrenz an. Hier wird also die Konkurrenz der Verlage untereinander dazu ausgenutzt, die Verlage letztlich unter Druck zu setzen, möglichst viele Inhalte zur Verfügung zu stellen und eben nicht zu sperren. Die Aussage ist: Wenn du sperrst, zeigen wir dich nicht mehr an. Deswegen kann das System in dieser Form auf keinen Fall die Lösung sein und man muss sich klar von der Vorstellung verabschieden, jetzt ein wenig an robots.txt herumzufeilen und vielleicht ein paar weitere Meta-Tags zu integrieren und schon sei das Problem gelöst. Diese Technologie liefert nur ein Instrument, um das grundsätzliche Recht, das man erst einmal schaffen muss, nachher umzusetzen. Ohne das Recht nützt aber die Technik erst einmal überhaupt nichts, sondern sie läuft leer. Das heißt, Sie brauchen sich nicht Gedanken über die Technik zu machen, bevor Sie nicht das Leistungsschutzrecht haben. Wenn aber das Leistungsschutzrecht da ist, steht die Technik bereit, um es dann auch umzusetzen. Ich glaube, das reicht erst einmal. Vielen Dank.

Der Vorsitzende: Vielen Dank, Herr Dr. Höppner. Nicht dass wir vorher nicht wussten, dass es sehr kontroverse Positionen gibt, aber das ist, glaube ich, eben noch einmal ganz gut herausgearbeitet worden. Deswegen freuen wir uns jetzt auf die Antworten von Herrn Prof. Dr. Lewandowski. Bitte schön!

Prof. Dr. Dirk Lewandowski (Hochschule für Angewandte Wissenschaften, Hamburg): Auch ich möchte mich dafür bedanken, dass ich heute hier sprechen darf und auf die beiden Fragen in aller gebotenen Kürze eingehen. Frau Dr. Sitte hatte gefragt, wie es aussehe mit den robots.txt-Dateien. Sie sind zwar nicht verpflichtend, aber in der Tat werden sie von den großen Suchmaschinen beachtet. Das heißt aber nicht, dass auch andere Suchmaschinen sie beachten müssen. Herr Dr. Höppner hat ja bereits angesprochen, dass es sozusagen die Bösen gibt, die robots.txt missachten. Es gibt aber in der Tat auch Dienste, die speziell darauf abstellen, zu schauen, wo etwas verboten ist, um dann gezielt diese Inhalte aufzugreifen. Das ist aus meiner Sicht gar nicht das allzu große Problem. Wenn Sie heute eine Suchmaschine aufbauen, dann müssen Sie sich schon sehr gut überlegen, ob Sie diese robots.txt befolgen wollen. Denn bei robots.txt haben Sie als Webseitenbetreiber nicht nur die Möglichkeit, etwas zu verbieten, sondern auch explizit etwas zu erlauben. Und, was man in der Vergangenheit auch schon gesehen hat, ist, dass Webseitenanbieter durchaus gut gemeint sagen: Dem guten Google-Bot erlaube ich alles und den anderen erlaube ich nichts. Das heißt, Sie werden von wesentlichen Inhalten unter Umständen ausgeschlossen. Die Frage ist insofern berechtigt, ob es für Suchmaschinen sinnvoll sein könnte, robots.txt vollständig zu beachten. Die Großen tun es und sie tun es auch zu Recht.

Der zweite Teil der Frage hob darauf ab, ob man mit einer Erweiterung von robots.txt das Problem lösen könne. Es wäre sicherlich sinnvoll, robots.txt weiter auszufüllen und weitere Steuerungsmöglichkeiten zu ermöglichen. Es wurde am Anfang auch gesagt, dass es sich um eine Anweisungssprache handelt, die sehr alt ist, wenig verändert wurde und leider nur eine recht ungenaue Steuerung ermöglicht. Ich halte es schon für sinnvoll, weiteres hinzuzufügen. Ob damit aber die hier angesprochenen Probleme gelöst wären,

das ist dann doch eine andere Frage. Zu dem Beispiel, das Herr Dr. Holfelder in Bezug auf ein gut aussehendes, die Suchmaschine verpflichtend anzuzeigendes Snippet gebracht hat, kann ich nur sagen, dass es nicht besonders sinnvoll ist, den Suchmaschinen vorzugeben, welche Snippets sie anzuzeigen haben. Der Fall wirkt auf mich ein wenig konstruiert, es gebe ein Snippet, das gut aussehe, aber es verberge sich dahinter eine Website, die Spam sei, was man nicht ohne weiteres feststellen könne, weshalb man es in der Suchergebnisliste anzeigen müsse. Suchmaschinen wie Google sind ganz gut in der Lage, von vornherein die Dokumente entsprechend zu ranken oder gar von dem Index auszuschließen, so dass dieser Fall in der Praxis kaum eintreten dürfte.

Dann gab es die Frage von Herrn Dr. von Notz nach der automatischen Erkennung von Presseerzeugnissen. Hier gibt es ganz unterschiedliche Aspekte, die zu klären wären. Zum einen muss man klären, ob die Kollektionen zu erkennen sind, das heißt die einzelnen Verlagsangebote, und die zweite Ebene ist die der Artikel. Wenn wir uns heute Google News ansehen, dann ist es ja keine Kollektion die maschinell zusammengestellt ist, sondern bei Google haben sich Menschen dafür entschieden, was sie als Nachrichtenquelle ansehen. Insofern ließe sich auch entsprechend umsetzen, dass Nachrichtenquellen gemeldet und dann in den Index aufgenommen werden. Auf der Ebene der einzelnen Artikel festzustellen, ob es jetzt eine journalistische Kommunikation oder eine interessen geleitete ist, scheint mir nur schwer möglich zu sein. Also, eigentlich gibt es ein Einerseits-andererseits-Urteil. Auf der einen Seite haben wir die journalistischen Angebote. Wenn wir uns das auf Google News ansehen, dann stehen etwa 1.000 deutsche Nachrichtenquellen dahinter. Diese zu erkennen, das sollte nicht das Problem sein. Sie auf der Ebene der Artikel zu erkennen, das dürfte dagegen äußerst schwierig sein.

Der Vorsitzende: Vielen Dank, Herr Prof. Dr. Lewandowski. Wir kommen dann zu Herrn Steidl. Da gab es eine Frage des Kollegen Heveling. Bitte schön!

Michael Steidl (International Press Telecommunications Council, London): Ich danke Ihnen für die Möglichkeit, hier vor den Mitgliedern des Deutschen Bundestages zu sprechen, insbesondere als Vertreter einer internationalen Organisation. Die Frage von Herrn Heveling lautete, ob der Standard robots.txt ausreiche, um sich aus der Verarbeitung durch Suchmaschinen abzumelden. Ich freue mich, dass gerade an mich eine Frage über einen Standard gestellt wurde, weil die IPTC letzten Endes die Standardisierungsorganisation der Nachrichtenbranche ist. Wir wurden 1965 gegründet, und seit damals machen wir die technischen Standards international für den Nachrichtenaustausch. Praktisch jede Nachricht, die Sie heute von einer Nachrichtenagentur weltweit erhalten, folgt einem IPTC-Standard. Dazu möchte ich auch anmerken, dass unsere Standards in der Regel immer so konstruiert sind - in unserem Namen schwingt ja Telecommunication mit - dass wir uns der Tatsache bewusst sind, dass sie eine Kommunikation abdecken. Das heißt, sie haben einen Absender, einen Empfänger und werden verarbeitet. Deshalb ist ein ganz wesentlicher Faktor bei der Erstellung unserer Standards, dass wir uns darum kümmern, dass die Bedürfnisse der Vorgenannten Berücksichtigung finden.

Wir haben gerade damit begonnen, Formate für den Meldungsaustausch zu entwerfen. Uns ist dabei bewusst geworden, dass Rechten und dem formalen Festhalten an ihnen eine immer größere Bedeutung zukommt. Deshalb setzen wir uns seit den 2000er Jahren auch intensiv damit auseinander, wie diese Rechte maschinenlesbar ausgedrückt werden können. Die Beteiligung etwa an der Erarbeitung des ICAP-Standards war ein Beitrag dazu. Der zweite, jüngere Beitrag ist, dass wir eine Weiterentwicklung von ICAP in Form der sogenannten RightsML-Standards begonnen haben und jetzt so weit sind, dass wir eine umfassende Spezifikation vorstellen können.

Das ist der Hintergrund, vor dem ich Ihre Frage höre und eine Antwort versuche, also auch die Frage, was robots.txt für ein Standard ist. Der Umgang mit der Terminologie ist leider oft diffus. Es ist von robots.txt die Rede, aber man meint meistens eine Mischung aus Rechtsausdrücken, die in einer tatsächlich vorhandenen Datei namens robots.txt auf Webseiten enthalten sind, aber genauso rasch wird die Möglichkeit inkludiert, Meta-Tags einzusetzen, die auch über diese mögliche Verwendung von Webseiten Auskunft geben. Es hat sich dazu auch der Begriff Robots Exclusion Protocol-Standard als umfassender Begriff eingebürgert und ich erlaube mir, diese Kombination von robots.txt und Meta-Tags anzusprechen.

Was ist das Kriterium für einen Standard, auf den man vertrauen kann? Das Erste ist die Klärung, wer sich ggf. überhaupt an einen Standard hält. Wenn ich diesen Standard anwende, von wem kann ich erwarten, dass er diesen Standard ebenfalls anwendet? Dazu gebe ich den Hinweis, dass auf dem Google Webmaster Central Blog eine Mitteilung vom Juni 2008 verbreitet wurde, die über das Robots Exclusion Protocol Auskunft gibt und in der es heißt, alle folgenden Features von REP würden derzeit von Google, Microsoft und Yahoo umgesetzt. Ich habe versucht, das zeitlich nachzuvollziehen, und bin auf eine identische Mitteilung auf der Webseite von Google Developers im Unterbereich Webmaster gestoßen, letztes Update vom Februar 2012, worin es heißt, dass die Dokumente Auskunft darüber geben, wie Google mit dem Protokoll umgeht. Es muss also beachtet werden, dass es hier offensichtlich auch Unterschiede gibt, zwischen jenen, die sich zu festgeschriebenen Definitionen bekennen.

Das Zweite ist folglich, wie diese Definitionen aussehen. Es gibt eine Unterscheidung nach Agenten, das heißt nach Software, die auf Webseiten zugreift, was eine rein kategorische Unterscheidung ist. Es ist nirgendwo definiert, dass Bestandteil des Standards ist, welcher Agent wohin gehört, und man als Anwender unbedingt beobachten muss, mit wem man überhaupt spricht bzw. wer auf die Webseite kommt und etwas abfragt. Weil es ja, nüchtern gesagt, so ist, dass eine Abfrage über das Web ein rein technischer Vorgang ist und man keine Visitenkarte hineinstecken und sagen muss: Hallo, ich bin der Webcrawler von da und dort, sondern dass es letzten Endes genügt, eine IP-Adresse anzugeben, also eine Kennzahl des Webs, und schon kann dahinter ein solcher Web-Suchagent verborgen sein, ohne dass man je erfährt, für wen er konkret arbeitet. Das heißt, der Agent ist zwar eine Kategorisierung, hilft aber nicht weiter, um zu identifizieren, wer mit meiner Webseite spricht.

Bezüglich der Definitionen gibt es allerdings auch noch Unterschiede sowohl zum Inhalt der robots.txt-Datei als auch zu den Meta-Tags, die hier verwendet werden. Grundsätzlich ist dazu zu sagen,

dass Meta-Tags letzten Endes nur ein Stichwort abgeben und es dazu lediglich sehr kurze Beschreibungen gibt, was die Anwendung bedeutet. Diese Beschreibungen, auch Meanings genannt, sind in der Regel gerade mal einen Satz lang und sollen einigermaßen, so hofft man, transparent machen, was damit gemeint ist. Ich zähle hier die für die Meta-Tags definierten Elemente oder Stichworte aus dem Robots Exclusion Protocol auf, die da lauten: All, noindex, nofollow, non, noarchive, nosnippet, noodp, notranslate, noimageindex und unavailableafter. Hier endet die Liste im Google-Dokument. Mir ist nicht aufgefallen, dass das Description-Element hier angeführt wäre.

Der Vorsitzende: Herr Steidl, ich muss einmal kurz nachfragen, oder auch Herrn Heveling, von ihm ging die Frage ja aus, passt das, was Sie vortragen, noch zusammen mit der Beantwortung. Es ist zwar alles interessant, aber ich muss auch sicherstellen, dass sich hier Absender und Empfänger treffen. Fahren Sie also bitte fort, Herr Steidl.

Michael Steidl (International Press Telecommunications Council, London): Danke für die Klärung. Für die Begriffe der IPTC sind dies sehr kurze Beschreibungen, und aus unserer Erfahrung mit Standards, die wir seit den 70er Jahren kreieren, ist es so, dass so kurze Beschreibungen sehr oft unterschiedlich gedeutet oder interpretiert werden.

Bei der Frage, ob robots.txt ein Standard ist, muss man zunächst zum Hintergrund und zum rechtlichen Träger des Standards etwas sagen. Es gibt nämlich keine registrierte Organisation, keine Gesellschaft und keinen Verein, der Ansprechpartner und verantwortlich für die Pflege von robots.txt ist. Auf den verschiedensten Webseiten der Suchmaschinen findet man dazu immer wieder Stellungnahmen, aus denen hervorgeht, wie robots.txt letztlich interpretiert wird. Das heißt, dass letzten Endes der Anwender, also derjenige, der mit den Suchmaschinen kommuniziert, von Webseite zu Webseite gehen muss, um die Informationen zusammenzutragen, wie unterschiedlich die Suchmaschinen robots.txt handhaben. Authentische Aussagen dazu, wie robots.txt zu deuten und zu verwenden ist, gibt es insofern nicht. Es gibt keine Trägerorganisation für robots.txt, das muss auch einmal ganz klar festgehalten werden.

Das sind wohl eher formale Aspekte zu robots.txt. Ein praktischer Aspekte hingegen ist, ob es sinnvoll ist, robots.txt zu verwenden, um sich abzumelden. Vor zehn Jahren mag es genügt haben, robots.txt zu verwenden, aber heute genügt es nicht mehr, denn es wird immer wieder das Wort Granularität erwähnt. Ein wesentlicher Aspekt der Granularität ist, dass es unterschiedliche Inhalte auf einer Seite gibt. Das Robots Exclusion Protocol in beiden Varianten, sowohl robots.txt als auch Meta-Tags, kann einzig und allein mit der ganzen Seite und nicht mit den einzelnen Abschnitten einer solchen umgehen.

Wenn Sie sich jetzt auf eine typische Nachrichten-Übersichtsseite einer Zeitung begeben, werden Sie dort immer auf der Homepage und auch auf ressortmäßigen Unterseiten eine Nachrichtenübersicht finden. Das heißt, dort, wo sehr viele Meldungen, in der Größenordnung von 20 und mehr, in Form ihrer Überschrift, aber auch in Form einer Zusammenfassung des Inhaltes zu finden sind, kann man lediglich die ganze Nachrichten-Übersichtsseite ausschließen, aber keine Teile davon. Es ist mithin mit Mitteln von

robots.txt und Meta-Tags unmöglich zu sagen, die ersten zehn Meldungen, die sind ohnehin sehr allgemein gültig, die können durchaus verwendet werden, aber dann habe ich weiter unten drei Hintergrundstorys, die ich als Nachrichtenagentur aufwendig erarbeitet habe und deshalb nicht möchte, dass diese weiter indiziert werden. Diese Unterscheidung lässt sich einfach nicht treffen. Letztendlich gibt es an der Stelle nur ein Hop oder Top, ein Ja oder Nein und es gibt hier keine wirkliche Granularität.

Der Vorsitzende: Gut, Herr Steidl, ich muss noch einmal dazwischen gehen. Denn ich merke, dass sich bei allen Fraktionen etwas Unruhe breitmacht. Deswegen bitte ich um Nachsicht, wir wollen schon darauf achten, dass auch noch andere Aspekte angesprochen werden. Die Antwort lautete: Ja, aber es kommt darauf an. Das wäre die Kurzfassung auf die Frage von Herrn Heveling. Herr Heveling hatte mir auch signalisiert, dass er mit Ihrer Antwort soweit arbeiten kann. Ich schlage deshalb vor, dass wir die zweite Fragerunde aufrufen. Für die CDU/CSU-Fraktion Herr Kollege Thomas Jarzombek, bitte.

Abg. Thomas Jarzombek (CDU/CSU): Meine Frage geht sowohl an Herrn Dr. Holfelder als auch an Herrn Steidl. Auf der deutschen Google-Seite findet man ein ganz interessantes Dokument mit der Überschrift „Google blockieren“. Es wird dann erläutert, dass es wichtig sei, selbst wenn man mit Hilfe einer robots.txt-Datei das Crawlen des Contents einer Website durch Spider unterbinden wolle, dass Google den Content auf andere Weise ermitteln und ihn dem Index hinzufügen könne. Das heißt für mich, dass robots.txt bei Google nicht allein komplett sperrend wirkt. Ich würde gern von Ihnen wissen, ob Sie das bestätigen können und wie sich das auf Google News bezieht, ob dort eine Unterscheidung gemacht wird.

Der Vorsitzende: Für die Fraktion der SPD, Herr Dörmann, bitte.

Abg. Martin Dörmann (SPD): Ich habe eine Frage an Herrn Dr. Holfelder und meine Kollegin Zypries will auch eine Frage stellen. Herr Dr. Holfelder, wenn ich noch einmal auf die von Herrn Dr. Höppner dargelegten Zusammenhänge zurückkomme, wie das mit dem Traffic genau ist, so interessiert mich, ob es Erkenntnisse darüber gibt, welcher Traffic von Google auf die Angebote der Verleger geht bzw. User von diesen abhält? Welche objektivierbaren Erkenntnisse können Sie diesbezüglich vorbringen? Und Herr Höppner hat ja auch gesagt, dass es bei Sperrungen eben bestimmte Minuslistungen gibt. Wie ist das Verhältnis an der Stelle?

Abg. Brigitte Zypries (SPD): Ich habe eine Frage an Herrn Prof. Dr. Lewandowski. Herr Dr. Holfelder sagte vorhin, wenn das Leistungsschutzrecht komme, müsse eine Verlagsliste an Google geliefert werden, damit man überhaupt weiß, wer von dem Verbot umfasst ist. Würden Sie das ebenso einschätzen, und wenn ja, würden Sie meinen, dass das aus verfassungsrechtlichen Gründen notwendig wäre im Hinblick auf das Bestimmtheitsgebot und deshalb eigentlich in das Gesetz aufgenommen werden müsste?

Der Vorsitzende: Wir fahren fort mit Herrn Schulz.

Abg. Jimmy Schulz (FDP): Ich habe eine Frage an Herrn Dr. Holfelder und eine Frage an Herrn Dr. Höppner. Soeben haben wir gehört, dass der Standard robots.txt nicht von einer bestimmten Instanz oder einem Gremium gemacht wird, wie übrigens meiner Erinnerung nach alle Standards im Internet eher nach dem Best-Practice-Prinzip funktionieren und gleichwohl sehr exzellent funktionieren. Wir haben gehört, dass es die Option gibt, sich nicht an die robots.txt zu halten. Wäre es insofern zum Beispiel denkbar, Herr Dr. Holfelder, dass man solche Standards, wie wir sie gerade diskutiert haben, die auch Sie vorgeschlagen und angedeutet haben, wie Verlage damit umgehen könnten, wäre es also denkbar, dass man diese Standards in Deutschland mit einem Schutz versehen könnte, um damit einen Teil der Problematik möglicherweise zu beheben?

Herr Dr. Höppner, Sie erwähnten in Ihrem Beitrag, die Problematik liege darin, dass Suchmaschinen den Verlagen einen wesentlichen Teil des Traffics wegnehmen. Ist das wirklich so? Lässt sich belegen, dass Google Ihnen User wegnimmt bzw. dadurch, dass Sie von Google mit diesen Snippets verlinkt werden, weniger User haben, als wenn nicht verlinkt würde? Das würde mich noch einmal interessieren.

Der Vorsitzende: Für die Fraktion DIE LINKE., Herr Behrens, bitte.

Abg. Herbert Behrens (DIE LINKE.): Welcher Aufwand entsteht eigentlich bei Suchmaschinen oder ähnlichen Diensten, würde ich gern Herrn Dr. Holfelder fragen, wenn man, aus welchen Gründen auch immer, keine Lizenzen abschließen mag? Wir gehen davon aus, auch in Anbetracht der Ergebnisse der Anhörung im Rechtsausschuss, dass sich, zumindest für den Anfang, bei einer Vielzahl von Verträgen Rechteinhabern einfach nicht sofort melden werden. Uns interessiert, was das für diejenigen, die sich nicht rechtzeitig melden, bedeuten würde. Wie ist das technisch realisiert und in welchen Bereich ist das eigentlich qualitativ einzustufen? Ich stelle diese Fragen auch vor dem Hintergrund, dass wir es ja mit einer sehr vielfältigen Medienbranche in Deutschland zu tun haben. Können insbesondere überhaupt Kleinverlage diesen Aufwand leisten, oder besteht in diesem Zusammenhang nicht auch die große Gefahr, dass eine weitere Medienkonzentration eintritt?

Abg. Dr. Petra Sitte (DIE LINKE.): Als Herr Steidl eben seine Erläuterung gab, hat Herr Dr. Holfelder mehrfach mit dem Kopf geschüttelt. Mich würde erstens interessieren, warum er den Kopf geschüttelt hat und zweitens reden wir immer über robots.txt, aber es gibt ja durchaus gerade vor dem Hintergrund anderer innovativer Lösungen suchmaschinenähnliche Dienste bzw. Suchmaschinen, die RSS-Feeds nutzen, die die Verlage auch anbieten. In welchem Verhältnis steht das denn dann zueinander?

Der Vorsitzende: Es kommt noch Herr Dr. von Notz an die Reihe, und dann schließen wir die zweite Fragerunde ab. Herr Dr. von Notz, bitte schön.

Abg. Dr. Konstantin von Notz (BÜNDNIS 90/DIE GRÜNEN): Vielen Dank, Herr Vorsitzender. Meine erste Frage geht an Herrn Dr. Höppner. Es ist zugegebenermaßen keine vollkommen technische, aber Ihre Ausführungen, Herr Dr. Höppner, waren ja auch nicht nur rein technisch, wenn ich an die

Köstlichkeiten aller Welt in einem Taxi denke. Ich habe Ihre Ausführungen wie der Kollege Jimmy Schulz auch im Grunde genommen so verstanden, dass es den Verlagen ohne Suchmaschinen besser ginge, als mit. Mich interessiert, ob das wirklich Ihre These ist.

Die zweite Frage richtet sich an Herrn Prof. Dr. Lewandowski. Der Regierungsentwurf selbst zielt ja lediglich auf gewerbliche Anbieter, was Suchmaschinen betrifft und auf gewerbliche Anbieter von Diensten, die ihre Inhalte entsprechend aufbereiten. Nach Auffassung einiger Sachverständiger in der Anhörung des Rechtsausschusses am 30. Januar 2013 fallen darunter auch Dienste wie Twitter und Facebook. Deshalb lautet meine Frage, wie das dort eigentlich technisch umgesetzt werden soll. Die Dringlichkeit dieser Frage liegt darin, dass das Gesetz, über welches wir hier noch von technischer Seite reden, bereits am Freitag zur Abstimmung auf der Tagesordnung des Deutschen Bundestags steht. Deswegen ist es wichtig zu wissen, ob derart wesentliche Fragen schon geklärt sind.

Der Vorsitzende: Mit der Bitte um prägnante, präzise, kurze Antworten kommen wir dann zur Antwortrunde und wir beginnen wieder mit Herrn Dr. Holfelder:

Dr. Wieland Holfelder (Technikdirektor, Google Germany GmbH): Danke. Herr Jarzombek, Sie haben gefragt, wie sich dieser Widerspruch auflösen lässt. Man muss grundsätzlich verstehen, dass es zwei Dinge gibt. Es gibt einmal das Crawlen und dann noch das Indizieren. Crawlen bedeutet, dass ich eine Webseite öffne und dann in dieser weitere Links finden kann, die ich dann dazu verwende, um das gesamte Netz zu erfassen. Indizieren bedeutet, den Inhalt heranzuziehen und ihn in den Index aufzunehmen.

Was wir auf dieser angesprochenen Seite sagen, ist, dass es durchaus sein kann, dass es selbst wenn Sie Ihre Seite nicht crawlbar machen, für uns Möglichkeiten gibt, diese zu finden, weil beispielsweise, der Link zu Ihrer Webseite auf einer anderen Webseite steht. Das ist nicht ausgeschlossen. Um zu vermeiden, dass die Inhalte einer Webseite tatsächlich angezeigt werden, müssen Sie in dieser speziellen Seite ein entsprechendes Tag einfügen, um auszuschließen, dass die Seite angezeigt wird.

Abg. Martin Dörmann (SPD): Können Sie anhand Ihrer Zahlen das Argument nachvollziehen, dass Google unter dem Strich dazu führt, dass bei den Presseverlagen weniger Nachrichten direkt nachgefragt werden?

Dr. Wieland Holfelder (Technikdirektor, Google Germany GmbH): Natürlich haben wir unsere eigenen Quellen und auch Zahlen veröffentlicht, dass wir Klicks in einer Größenordnung von sechs Milliarden an die Verlagswebseiten weiterleiten bzw. im vergangenen Jahr über AdSense-Produkte sieben Milliarden Dollar ausgeschüttet haben. Ich glaube aber, viel beeindruckender ist, dass Herr Döpfner in einem Interview, das vor wenigen Tagen, am 29. Januar, im „Wall Street Journal Deutschland“ veröffentlicht wurde, sagte: „Das gleiche gilt für Google, wir sind Google unendlich dankbar, dass teilweise bis zu 80 Prozent unserer Leser über Google unsere Artikel finden. Wir sind dankbar dafür, dass sie mit dieser

genialen Suchmaschine das Internet wirklich massenmarktfähig gemacht haben. Gleichzeitig streiten wir aber mit Google über Fragen piratenartiger Geschäftsmodelle, oder anders formuliert, über eine faire Architektur beim Benutzen und Vermarkten unserer Inhalte.“ Das heißt: Natürlich wollen wir all die Bösen nicht haben, aber auf die Argumentation, dass die Verlage nicht von Google profitierten, aufzubauen, macht natürlich keinen Sinn, wenn das selbst der Vorstandsvorsitzende von Springer anders sieht.

Herr Schulz hatte gefragt, ob es für robots.txt nicht eine Instanz, bzw. ein Gremium gebe. Wie Sie schon richtig gesagt haben, gilt das für viele andere der Web-Standards auch und robots.txt ist ja durchaus auch von HTML übernommen worden. Deshalb habe ich vorhin, das war einer meiner Kopfschüttler, auch gesagt, dass das Description-Tag Teil des HTML-Standards ist, aber von den Suchmaschinen an der Stelle akzeptiert wird und auch wieder darauf verwiesen wird. Die Internetstandards sind ja alle miteinander verwoben. Auch TCP/IP, auf dem das Internet aufgebaut ist, ist inzwischen über 30, 40 Jahre alt und funktioniert nach wie vor bestens. Deshalb kann die Argumentation, es ist 20 Jahre alt, deswegen funktioniert es nicht mehr, auch angezweifelt werden. Ansonsten würde das Internet, so wie wir es heute alle genießen, gar nicht funktionieren. Natürlich ist es denkbar, über Erweiterungen zu reden und zu diskutieren. Ich bin mir nicht sicher, ob der Ansatz, einen internationalen Standard in deutsches Recht zu überführen, der Richtige ist, denn wir wollen ja in Deutschland eigentlich keinen Alleingang machen. Von daher müssen wir aufpassen, welche Rolle wir spielen wollen oder können. Aber gemeinsam daran arbeiten, dass robots.txt oder Robots Exclusion Protocol besser werden, dem steht überhaupt nichts entgegen. Wir haben immer gesagt, dass wir da mit den entsprechenden Parteien reden und das werden wir auch weiterführen.

Herr Behrens hatte nach dem entstehenden Aufwand gefragt und was es für kleinere Verlage bedeuten würde, die vorgeschriebenen Mechanismen umzusetzen. Ich denke, da gilt es zwei Seiten zu beachten. Nämlich zum einen: Können es die kleinen Verlage leisten? Das andere ist: Wenn neue Dienste auf den Markt kommen oder jemand wie Google, Bing oder Yahoo, werden die dann alle Tausend oder wie viele Verlage auch immer tatsächlich auf der Liste haben, Verträge abschließen oder wird man sich auf die ersten 20 konzentrieren? Das heißt, es geht gar nicht darum, ob sich die Kleinen das leisten können, sondern ob sie überhaupt adressiert werden. Speziell, wenn es um Innovationen geht, wenn ich ein neues Start-Up-Unternehmen bin und sage, dass ich einen genialen neuen Dienst habe, der auf Snippets zugreift, die nach wie vor rechtlich in Ordnung sind, aber eine Vereinbarung benötigen, dann würde ich mich als junges Start-up-Unternehmen davor hüten, in Deutschland anzufangen und meinen Dienst hier anzubieten, sondern würde wahrscheinlich irgendwo im Ausland mein Kapital und meinen Firmensitz suchen und dort erst einmal die Innovationen durchführen. Selbst wenn sie dann in den deutschen Markt kommen, ist die Frage, ob sie es sich leisten können, Tausend Verlagswebseiten unter Vertrag zu nehmen. Das heißt, wir sperren Innovation an dieser Stelle aus und stellen uns selbst ein bisschen ins Abseits. Ich bekomme immer wieder Anfragen, auch von der politischen Seite, was man denn tun kann, damit das nächste Google, Facebook oder Apple in Deutschland passiert. Die Antwort ist ganz einfach: Kein Leistungsschutzrecht ist ein sehr guter Anfang.

Frau Sitte hatte wegen des Kopfschüttelns gefragt. Da war noch ein anderes Kopfschütteln, bei dem es um die Seitengranularität ging, wo Sie gesagt haben, dass man wirklich nur die ganze Seite sperren könne. Ich kann innerhalb einer Seite sogenannte I-Frames einbetten. Technisch gesehen sind das im Prinzip eigene Seiten und ich kann innerhalb dieser dann auch wieder das Robots Exclusion Protocol anwenden. Das kann ich natürlich, wenn ich es auf die Spitze treiben will, beliebig rekursiv machen. Das ist in der Praxis aber schlicht und ergreifend nicht notwendig, denn wenn ich nur eine Übersichtsseite habe und dort keine Inhalte anzeige, dann reicht es sicherlich aus, wenn ich es auf dieser Ebene einschränke und wenn ich dann sage, ich klicke mich zu dem einzelnen Artikel durch, dann habe ich die Möglichkeit, es genau auf diese Artikelwebseite einzuschränken. Ich sehe da also in der Praxis keine wirkliche Relevanz. Wenn ich wollte, könnte ich es tun, aber in der Praxis ist es nicht erforderlich.

Der Vorsitzende: Herr Dr. Höppner ist als Nächster dran. Da gab es eine Frage des Kollegen Schulz.

Dr. Thomas Höppner (Rechtsanwalt, Vertreter des Bundesverbands Deutscher Zeitungsverleger):

Es gab noch eine zweite Frage, die sich aber letztlich auf dasselbe bezog, nämlich ob die These stimme, Suchmaschinen würden nichts bringen, sondern man stünde ohne sie besser da. Hier muss klar unterschieden werden, über was für eine Suchmaschine wir denn reden: Es gibt horizontale Suchmaschinen, die das ganze Web abdecken. Da hat Google einen Marktanteil von über 90 Prozent in Europa. Aus gutem Grund. Horizontale Suchmaschinen sind aber nicht das, über was wir hier im Kern reden, sondern daneben gibt es, erstens, viele vertikale Suchmaschinen, das heißt Suchmaschinen, die sich auf ganz speziellen Content konzentrieren, in unserem Fall also Nachrichten. Daneben gibt es Aggregatoren, die mit einer Suchanfrage noch weniger zu tun haben, weil man noch nicht einmal eine Suchanfrage stellen muss, sondern die Aggregatoren selbst voreinstellen, was den Leser interessieren könnte, oder in weiteren Fällen, und davon gibt es ja auch viele Anbieter, ohne gefragt zu werden, aktiv, das heißt im Push-Modus, an den User die Nachrichtenübersichten schicken. Damit verabschieden wir uns völlig von einer Suche. Wir dürfen hier also nicht Äpfel mit Birnen vergleichen.

Kommen wir also zurück zur horizontalen Suche. Niemand bestreitet, auch die Verleger nicht, dass die horizontale Suche essentiell wichtig für das Funktionieren des Internets ist. Natürlich hat die horizontale Suche einen echten Mehrwert für Leute, die nach ganz speziellen Themen suchen und spezielle Fragen haben und dazu einen Vorschlag haben wollen. Aber darum, wie gesagt, geht es ja nicht, wenn man aktiv ein Portal aufbaut, bei dem man gar keine Frage erwartet und herauskitzeln will, sondern gleich die Antworten gibt. Aber zu den Vertikalen komme ich gleich noch.

Für die Verlage ist es dasselbe. Die Verlage verdienen daran, wenn sie Aufmerksamkeit bekommen. Aufmerksamkeit ist das Ziel. Wenn nun aber die Aggregatoren die Aufmerksamkeit für sich bündeln, weil sie eben die Nachfrage, nämlich nach aktuellen Nachrichten, selbst bedienen und selbst die Lösungen geben, dann ziehen sie die Aufmerksamkeit auf sich selbst und dann bleibt sie nicht mehr bei den Verlagen, die eigentlich die Informationen geliefert haben.

Wenn wir uns also Google News anschauen, Sie hatten konkret danach gefragt, ob es da Belege gibt, dann muss ich sagen, dass es äußerst schwierig ist für die Verlage, das zu prüfen, weil wir ja nicht Google News selbst kontrollieren. Die einzigen Studien, die wir dazu kennen, kommen zu dem Ergebnis, dass nahezu die Hälfte der Leser eben nicht durchklicken, sondern ihnen die Übersichten auf Google News reichen. Dann kann man sich doch schon einmal fragen, warum sie sich die News auf Google News holen müssen, wenn doch alle Verlagsseiten ähnlich aufgebaute Übersichten aufweisen – sie würden natürlich sagen, höherwertige. Natürlich fehlt in einem solchen Fall der Traffic, der bei Google verbleibt und der nicht weitergeleitet wird, irgendwo an anderer Stelle. Die entsprechenden Studien wurden bislang jedenfalls nicht widerlegt. Wir haben eben auch noch nicht die Antwort gehört, obwohl der Kollege gefragt wurde, wie denn das Verhältnis aussehe zwischen den Nutzern und denjenigen, die durchklicken. Dazu haben wir nichts vernehmen können. Ich glaube, es ist ziemlich plausibel, anzunehmen, dass bei Diensten wie Google News eben nicht alle durchklicken. Und selbst wenn wir einen Dienst hätten, der 100 Prozent der Leser durchleitet, dann würde dieser Dienst, dieser Aggregator, doch immer mindestens genauso viel Traffic und genauso viel Aufmerksamkeit auf sich bündeln, wie er zur Quellseite weggibt, weil jeder Klick auf die Quelle in diesem Fall ja vom Aggregator kommt. Also zunächst einmal muss der Nutzer zum Aggregator kommen. Und wie kommt er da hin? Er wird angelockt mit den Leistungen, die der Aggregator sich woanders genommen hat. Das heißt, die Aggregatoren bedienen sich und sagen: Tolle Nachrichten, die könnt ihr bei uns alle sehen. Und natürlich ziehen sie damit Aufmerksamkeit auf sich, die dann woanders fehlt. Ich denke, es gibt ein berechtigtes Interesse, dass man etwas dagegen hat und das nicht beliefern will. Wenn es schon keine technischen Instrumente gibt, sich dagegen adäquat zu schützen, dann finde ich, ist es berechtigt, dass man danach fragt, ob man da nicht einen rechtlichen Schutz bekommen kann.

Der Vorsitzende: Danke, Herr Dr. Höppner. Waren das die Antworten soweit? Ja, dann fahren wir fort mit Herrn Prof. Dr. Lewandowski. Es gab Fragen von Frau Zypries, wenn ich es richtig sehe, und von Herrn Dr. von Notz. Bitte sehr, Herr Prof. Dr. Lewandowski.

Prof. Dr. Dirk Lewandowski (Hochschule für Angewandte Wissenschaften, Hamburg): Zuerst möchte ich die Frage von Frau Zypries aufgreifen, die zwei Teile hatte. Einmal ging es um die Verlagslisten und den Aspekt, ob das technisch zumutbar sei für Google und andere Dienste. Man muss dazu sagen, dass Google, vielleicht entgegen der landläufigen Annahme, bereits jetzt schon manuell gepflegte Listen verwendet. Das betrifft Dienste wie Google News, wo die Pressequellen manuell bestimmt werden, es betrifft auch Dienste wie Google Scholar, eine wissenschaftliche Suchmaschine, aber auch die Adressverifikation in den Branchenbucheinträgen in der lokalen Suche. Also, es wird durchaus schon ein solcher Aufwand betrieben und ich würde das nicht als allzu große Schwierigkeit sehen wollen. Der zweite Teil Ihrer Frage betraf einen rechtlichen Aspekt. Diesen Teil kann ich schlichtweg nicht beantworten.

Es gab dann die Fragen von Herrn Dr. von Notz. Eine zielte auf die gewerblichen Anbieter im Gesetzentwurf und die zweite darauf, was es mit Twitter und Facebook auf sich hat, wie man die Dinge dort technisch umsetzen kann. Hierbei haben wir es mit einer anderen Art der Verlinkung zu tun, denn es

wird manuell von Menschen verlinkt. Die Snippets werden im Prinzip erst einmal selbst geschrieben, außer dass ein Snippet oder ein entsprechender Textteil herauskopiert werden kann. Grundsätzlich sehe ich weniger das Problem der technischen Kontrolle. Letztendlich kann man natürlich Leute haben, die einen kompletten Text herauskopieren, dann ist der Fall relativ klar. Und dann gibt es den Fall, wo man das Snippet selbst schreiben kann, dann geht es nur noch um den Link. Und dazwischen wäre die Kopie der Snippets und man hätte dann einen ähnlichen Fall, wenn ich das richtig verstanden habe.

Der Vorsitzende: Gut, dann fahren wir fort mit Herrn Steidl. Da gab es Fragen von Herrn Jarzombek.

Michael Steidl (International Press Telecommunications Council, London): Es geht dabei um die Frage, ob Inhalte auch auf andere Weise in den Besitz einer Suchmaschine kommen können als durch direkten Zugriff auf eine Seite. Dazu kann ich nur sagen, es ist vermutlich und teilweise so dokumentiert, zumindest finde ich auf der Developer-Information von Google den Hinweis: „Suchmaschinen-Algorithmen können entscheiden, ob sie eine Seite in ein Suchergebnis einfügen, obwohl sie nie auf die Seite, die angezeigt wird, jemals direkt zugegriffen haben.“ Das heißt, es ist deshalb stark zu vermuten, die Algorithmen von Google sind ein strenges Geschäftsgeheimnis, dass von Verweisen auf Inhalte anderer Seiten ausgegangen wird. Wenn also auf Seite A steht, auf Seite B gibt es Interessantes zu einem bestimmten Thema zu finden, und ich suche dann nach diesem Thema, dann wurde zwar nie auf die Seite B zugegriffen, aber durch diesen Verweis auf Seite A wird daraus geschlossen, dass auf Seite B ein solcher Inhalt zu finden ist und es wird dann der Verweis angezeigt. Das wirft für mich die Frage auf, wenn nie auf die Seite zugegriffen wurde, wie wertet man dann Meta-Tags dieser Seite aus, die verbieten könnten, dass die Seite indiziert wird?

Zu Meta-Tags kann ich grundsätzlich anmerken, dass die Beschreibung Meta-Tag-Description definitiv nicht Teil des HTML-Standards ist, sondern Meta-Tags als solche ein Standard sind. Sie sind vor allen Dingen anonyme Platzhalter, wo man Beliebiges vorhalten kann. Im HTML-Standard ist spezifiziert, dass jemand etwas in die Description schreiben kann. Das ist zwar, wie Herr Schulz gesagt hat, ein Stück weit Best Practice, aber wenn wir nach Standards fragen, müssen wir sagen: Nein, das ist kein Standard.

Der Vorsitzende: Danke für die Beantwortung der Fragen, die in der zweiten Fragerunde an Sie gestellt wurden. In Anbetracht der noch zur Verfügung stehenden Zeit schlage ich vor, dass wir uns in dieser letzten Fragerunde auf eine Frage pro Fraktion beschränken. Wir haben dann sicher noch ausreichend Zeit für die Beantwortung. Wir beginnen wieder mit der Fraktion der CDU/CSU, Herr Dr. Brandl, bitte.

Abg. Dr. Reinhard Brandl (CDU/CSU): Ich habe zwei Fragen, wobei mir aufgefallen ist, dass es zu einem Aspekt zwei unterschiedliche Antworten gab. Wie kommt die Google-News-Liste nun eigentlich zustande? Kommt sie automatisch zustande über Heuristiken oder kommt sie manuell zustande, indem jemand im Hintergrund die Liste der Verlags- oder Webseiten pflegt? Und dann habe ich noch eine Frage die Technik betreffend. Kann ein Webseitenbetreiber differenzieren zwischen dem Bot von Google, der normalen Suchmaschine und dem Bot von Google News? Kann er zum Beispiel Google News aussperren, aber den

normalen Google Bot zulassen. In Ihrem Handout, Herr Dr. Holfelder, ist lediglich von einem Google Bot die Rede und es wird auf den mich interessierenden Punkt nicht näher eingegangen.

Der Vorsitzende: An Herrn Dr. Holfelder gehen die beiden Fragen. Wir fahren fort mit der SPD-Fraktion. Abg. Klingbeil, bitte.

Abg. Lars Klingbeil (SPD): Ich habe noch eine Frage an Herrn Dr. Holfelder. Es liegt jetzt ja der Gesetzentwurf der Bundesregierung vor und es sieht ganz danach aus, dass er am Freitag dieser Woche in der 2. und 3. Lesung verabschiedet werden soll, obwohl jetzt wieder viele verfassungsrechtliche Fragen in den Raum geworfen wurden. Mich interessiert, wie Google all die Vorgaben, die der Gesetzentwurf technischer und auch rechtlicher Art enthält, umsetzen würde. Sehen Sie Möglichkeiten dazu und wie würden diese aussehen?

Der Vorsitzende: Und für die FDP-Fraktion Abg. Schulz, bitte.

Abg. Jimmy Schulz (FDP): Ich habe eine Frage an Herrn Dr. Höppner. Es wird ja immer wieder diskutiert, dass ein deskriptiver Teil unter anderem für die Verlinkung frei vom Leistungsschutzrecht bleiben soll. Es handelt sich dabei um einen Minimalteil, vielleicht ein Snippet. Wie definieren Sie diesen Teil, der frei von einem Leistungsschutzrecht bleiben soll? Sind das eine Art Zeichenanzahl, ein Satzbaustein, wie viele Wörter, wie wollen Sie das definieren?

Der Vorsitzende: Abg. Behrens, bitte.

Abg. Herbert Behrens (DIE LINKE.): Ich habe eine Frage an Herrn Prof. Dr. Lewandowski. Wir hatten in einer Kleinen Anfrage an die Bundesregierung danach gefragt, wie es um die Betreiber von sozialen Netzwerken steht. Prof. Dr. Lewandowski hatte eben darauf Bezug genommen. Ich würde gern wissen, ob die Betreiber solcher Dienste, wenn sie sich nicht immer lizenzieren wollen, künftig technische Sperren gegenüber den deutschen Nutzern einrichten. Also, hieße das, dass parallel zu den Streitigkeiten zwischen Youtube und GEMA in Frankreich, dass ein Tweet bzw. ein Eintrag nicht in jedem Land verfügbar ist, auch hier so etwas auftreten würde? Und, vor allem, mit welchem Aufwand wäre das zu realisieren?

Der Vorsitzende: Abg. Dr. von Notz, bitte.

Abg. Dr. Konstantin von Notz (BÜNDNIS 90/DIE GRÜNEN): Von Herrn Dr. Holfelder bzw. Herrn Dr. Haller möchte ich gern wissen, ob denn die Einführung des Gesetzes, wenn es so kommen sollte, wie es uns vorliegt, dazu führen würde, dass Google seine Dienste einfach abschaltet. Was wäre für Google eigentlich die Konsequenz aus der Einführung des Gesetzes gesehen in einem Zeitfenster von sechs Monaten?

Der Vorsitzende: Wir kommen dann wieder zur Beantwortung. Herr Dr. Holfelder, sie sind als Erster an der Reihe. Bitte schön!

Dr. Wieland Holfelder (Technikdirektor Google Germany GmbH): Herr Dr. Brandl hatte danach gefragt, wo die Google-News-Liste herkommt, ob sie auf Heuristik, einem Algorithmus oder auf manueller Erstellung basiert. Die Antwort lautet: Beides. Es gibt eine gewisse Heuristik, die feststellt, dass es eine Newsseite sein könnte. Daneben gibt es aber auch einen manuellen Pflegeaufwand, nachzusehen, ob die Treffer tatsächlich in die Kategorie passen. Der gravierende Unterschied ist, wenn es ein Verbotrecht gibt, dann reicht uns das in der Regel nicht, sondern wir benötigen eine verbindliche Liste. Darin liegt der Unterschied.

Dann hatten Sie danach gefragt, ob ein Webseitenbetreiber zwischen Google News und Google unterscheiden könne. Ja, das kann er. Was Sie in dem Snippet, in dem Meta-Tag gesehen haben, ist die Entscheidung, lediglich den Google Bot, der für die Websuche zuständig ist, anzusprechen. Es gibt aber auch einen Google Bot, der für die Newssuche zuständig ist und einen, der für die Bildsuche zuständig ist, und die werden mit entsprechend anderen User-agent-strings identifiziert. Das heißt, man könnte hier hineinschreiben, Googlebot „News“ darf nicht, aber Googlebot darf, und dann haben Sie diese Unterscheidung getroffen.

Herr Klingbeil, Sie hatten nach verfassungsrechtlichen Aspekten gefragt. Diesbezüglich würde ich gern die „Juristenkarte“ ziehen. Denn das ist nicht mein Fachgebiet, dafür ist Herr Dr. Haller zuständig.

Der Vorsitzende: Also, die Frage war – ich interpretiere das jetzt etwas weiter gehend, Kollege Dr. von Notz hat das ja eben auch getan –, ob es aufgrund des Leistungsschutzrechts zu einer Abschaltung als rein technischem Vorgang kommen würde. Das war die Frage nach den möglichen Konsequenzen, und ich glaube, Herr Dr. Haller wird darauf antworten. Bitte schön!

Dr. Arnd Haller (Justiziar, Google Germany GmbH): Vielen Dank. An dieser Stelle hängen ja Recht und Technik eng miteinander zusammen. Und wir haben viel über Standards diskutiert, robots.txt zu ACAP, und welches der bessere Standard ist. Aber darum geht es nach meinem Dafürhalten überhaupt nicht bei der Einführung eines Leistungsschutzrechts, denn hier stellt sich die technische Frage, wie man das Gesetz, das hier vorliegt, als Suchmaschine wie Google oder auch in sozialen Netzwerken implementieren kann. Insofern müssen Techniker und Juristen relativ eng zusammenarbeiten. Als Jurist muss ich unsere Techniker dahin gehend beraten, wie man das Gesetz implementieren kann. Und die Antwort lautet: Man kann es nicht implementieren. So einfach ist das. Warum kann man es nicht implementieren und warum gibt es verfassungsrechtliche Bedenken, die von den beiden Professoren Dr. Blankenagel und Dr. Spoerr angemeldet wurden, die zu dem Ergebnis gelangten, die Informationsfreiheit sei gefährdet und das Gesetz schlichtweg verfassungswidrig?

Der Vorsitzende: Das ist bekannt, Herr Dr. Haller. Das ist soweit auch uns nicht neu.

Dr. Arnd Haller (Justiziar, Google Germany GmbH): Sicher, aber das hat eng mit der Frage zu tun, wie man es implementieren kann. Denn, wir müssen das Gesetz dann ja auch anwenden. Der Suchroboter muss verstehen, was ein Presseverleger im Sinne des Gesetzes sein soll. Eben wurde zwar gesagt, dass man das schon erkennen könne. Es mag sein, dass man es möglicherweise bei Google News erkennen kann, wenn man unterstellt, dass in Google News nur Nachrichten enthalten sind, wobei diese Unterstellung schon falsch ist. Aber in der Suche – und über die reden wir hier ja auch, wir reden nicht nur über Newsaggregatoren, sondern wir reden auch über die Websuche –, da kann ein Bot nicht entscheiden, ob ein Text von einem Presseverleger stammt oder identisch – wortgleich möglicherweise – auf der Seite eines Journalisten liegt. Ein Bot kann nicht unterscheiden und verstehen, was ein „journalistischer Beitrag“ ist, was eine „periodisch veröffentlichte Sammlung“ ist. Ein Bot kann technisch nicht auslesen, was bei „Würdigung der Gesamumstände als überwiegend verlagstechnisch“ anzusehen ist. Derartige Fragen, und das sind rechtliche Fragen, die aber die Technik beantworten muss, lassen sich gegenwärtig außerhalb von News schlichtweg nicht umsetzen. Und das beantwortet so ein wenig auch die Frage, wie wir mit dem Gesetz umgehen werden. Es wird sich zeigen, wenn das Gesetz in Kraft tritt. Erstens wurde gesagt, es sei verfassungswidrig und zweitens wurde gesagt, es verstoße gegen internationale Abkommen. Und drittens wurde es bislang nicht notifiziert bei der EU mit der Konsequenz, dass das Gesetz möglicherweise vollständig leerläuft, zwar in Kraft tritt, aber vollständig leerläuft, weil es keine Anwendung bei Suchmaschinen oder Aggregatoren findet. Also, insofern müssen wir erst einmal abwarten, bis das Gesetz in Kraft tritt und wie der Wortlaut dann konkret aussieht. Ich sehe im Moment eher schwarz, dass man das Gesetz anwenden kann, ohne permanent in die Falle zu tappen, dass man doch auf einzelne Artikel verlinkt, die unter das Leistungsschutzrecht fallen, bei denen sich Suchmaschinen aber vorher von dem einzelnen Anbieter nicht die Rechte haben einräumen lassen.

Gestatten Sie mir noch eine kurze Bemerkung zu der europäischen Perspektive. Wir wollen deutsche Verlage schützen, müssen aber zugleich noch alle europäischen Verlage im EWR schützen, da die nämlich auch alle mit drin wären, was heißen würde, dass jede deutsche Suchmaschine und jeder deutsche Newsaggregator nicht nur bei jedem deutschen Verlag nachfragen müsste, sondern auch noch bei jedem anderen kleinen Verlag in Europa. Das hieße, dass es viel mehr sein werden, als die Zahl, die bislang im Raum steht.

Der Vorsitzende: Wir fahren fort mit Herrn Dr. Höppner. Bitte schön.

Dr. Thomas Höppner (Rechtsanwalt, Vertreter des Bundesverbands Deutscher Zeitungsverleger): Die Frage nach zulässigen Links spielt ja letztlich auf ein ähnliches Thema an, nämlich, ob die Informationsfreiheit beschränkt wird, indem man Links zulässt und erst einmal klären muss, was eigentlich noch zu einem Link gehört. Der Gesetzentwurf hat ja die Verlinkung ausdrücklich vom Schutzrecht ausgenommen. Das bedeutet, dass Links nach wie vor angezeigt werden dürfen. Deshalb verstehe ich die Diskussion, die hier gerade angesprochen wurde in Bezug auf die angebliche Beschränkung der Informationsfreiheit nicht. Natürlich können User nach wie vor Informationen uneingeschränkt im Internet finden. Sie können entweder direkt auf den Verlagsseiten oder nach Verlagsseiten suchen. Man bekommt

dann den Link und kann auf diesen klicken, woraufhin man dann bei der Verlagsseite landet und die Inhalte weiter konsumieren kann.

Ich würde es für die Informationsfreiheit als viel schwerwiegender empfinden, wenn Verlage als Reaktion auf einen mangelnden Leistungsschutz gezwungen würden, weniger Inhalte ins Netz oder sie alle nur noch gegen ein Entgelt zur Verfügung zu stellen oder, viel schlimmer noch, mit robots.txt zu sperren, so dass sie gar nicht mehr von den Usern gefunden werden können. Das wären doch die Alternativen. Sieht man sich die Alternativen einmal genauer an, dann ist der Hinweis, dass man durch die Beschränkung von Suchergebnissen auf reine Links, wie es für viele Jahre Standard war, die Informationsfreiheit beschränken würde, nicht mehr wirklich plausibel. Um aber zu einem zulässigen Link zu kommen, ist ja erneut eine technische Frage, nämlich die nach der Begründung und wieweit man hier die Verlinkung zieht, relevant. Fest steht jedenfalls, dass die Anzeige der URL immer zulässig sein wird. Die URL nämlich als Link. Es ist heute doch schon Standard, dass bereits in der URL auch die Überschrift des Beitrags enthalten ist. Das heißt, dass man anhand der URL auch den Inhalt des Beitrags kursorisch lesen kann. Man wird sich ggf. fragen, ob es noch zulässig ist, wenn sowieso auf die Meta-Tags abgestellt wird, den Titel in den Meta-Tag mit aufzunehmen. Das wäre nämlich dann eine komplette Überschrift. Man hätte dann die Überschrift und die URL, was für die Information und für das Aufsuchen von Nachrichten völlig ausreichend wäre. Warum muss man denn schon auf der Ergebnisseite einer Suche die Antwort bringen? Warum muss man da denn schon riesige Texte anbieten? Das ist doch eigentlich gar nicht die Funktion der Suche, sondern, und das hat in den ersten Jahren bestens funktioniert und von daher kommen die Suchmaschinen ja auch, dass sie Links anzeigen. Und dem steht nun einmal das Leistungsschutzrecht nicht entgegen.

Der Vorsitzende: Meine Herren, vielen Dank, dass Sie sich die Zeit genommen haben, für die Beantwortung unserer Fragen zur Verfügung zu stehen. Einige Punkte sind sicherlich noch nicht abschließend geklärt, weshalb es an uns liegt, damit weiter beratend umzugehen. Ich danke auch meinen Kolleginnen und Kollegen, dass sie sich hier mit Fragen an der Beratung beteiligt haben und schließe hiermit die Sitzung des Unterausschusses Neue Medien für heute.

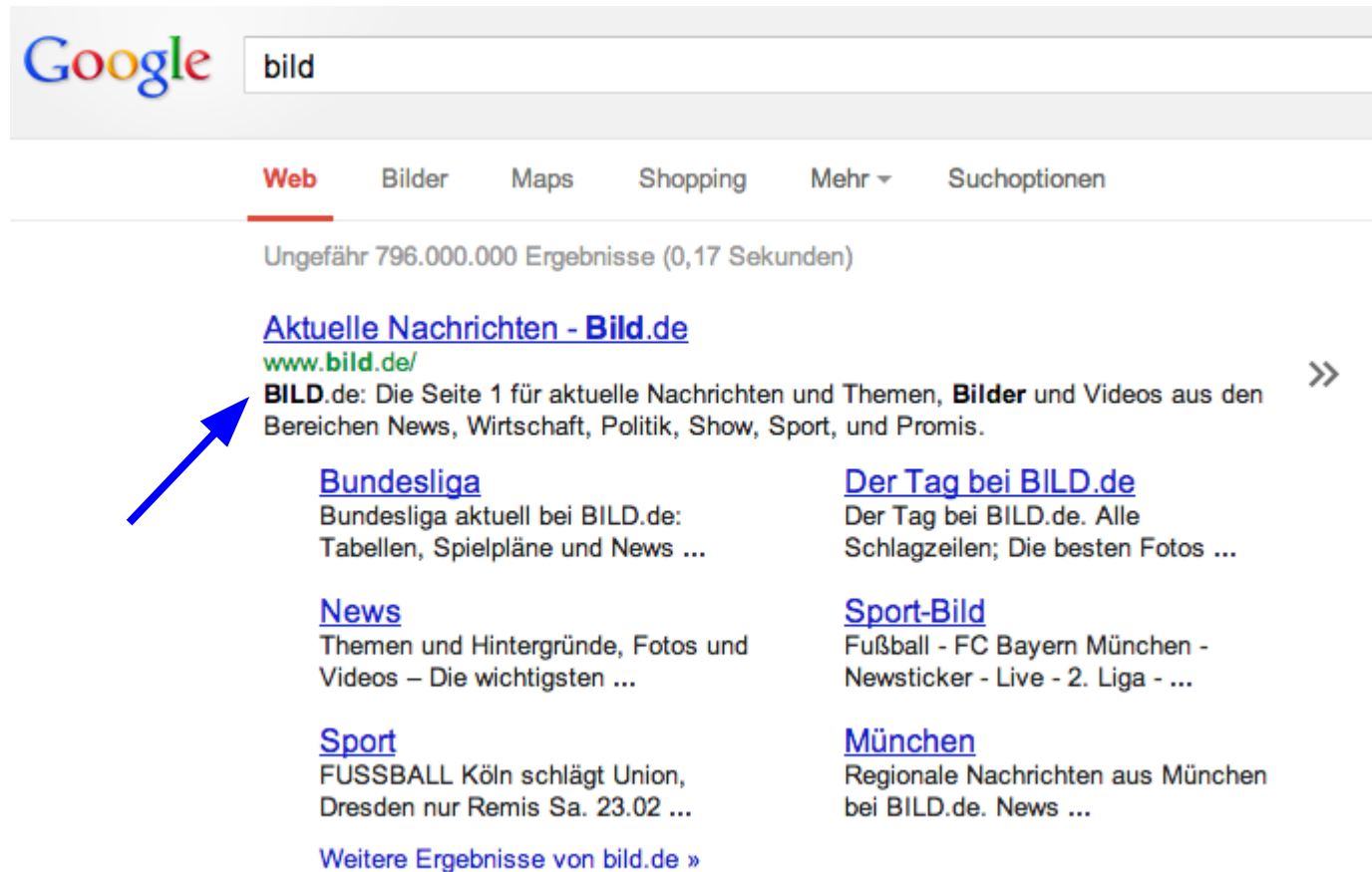
Schluss der Sitzung: 14:35 Uhr

Sebastian Blumenthal, MdB
Vorsitzender

Beispiele zu Snippets, Robots.txt und Robots Excursion Protocol (REP)

**Als Beilage zur öffentlichen Anhörung zum "Entwurf eines Siebenten Gesetzes
zur Änderung des Urheberrechtsgesetzes - technische Aspekte" am 25. Februar 2013**

Beispiel 1: Websuche auf www. google.de nach "bild"



Google bild

Web Bilder Maps Shopping Mehr ▾ Suchoptionen

Ungefähr 796.000.000 Ergebnisse (0,17 Sekunden)

[Aktuelle Nachrichten - Bild.de](#)
www.bild.de/ >>

BILD.de: Die Seite 1 für aktuelle Nachrichten und Themen, **Bilder** und Videos aus den Bereichen News, Wirtschaft, Politik, Show, Sport, und Promis.

[Bundesliga](#)
Bundesliga aktuell bei BILD.de:
Tabellen, Spielpläne und News ...

[Der Tag bei BILD.de](#)
Der Tag bei BILD.de. Alle
Schlagzeilen; Die besten Fotos ...

[News](#)
Themen und Hintergründe, Fotos und
Videos – Die wichtigsten ...

[Sport-Bild](#)
Fußball - FC Bayern München -
Newsticker - Live - 2. Liga - ...

[Sport](#)
FUSSBALL Köln schlägt Union,
Dresden nur Remis Sa. 23.02 ...

[München](#)
Regionale Nachrichten aus München
bei BILD.de. News ...

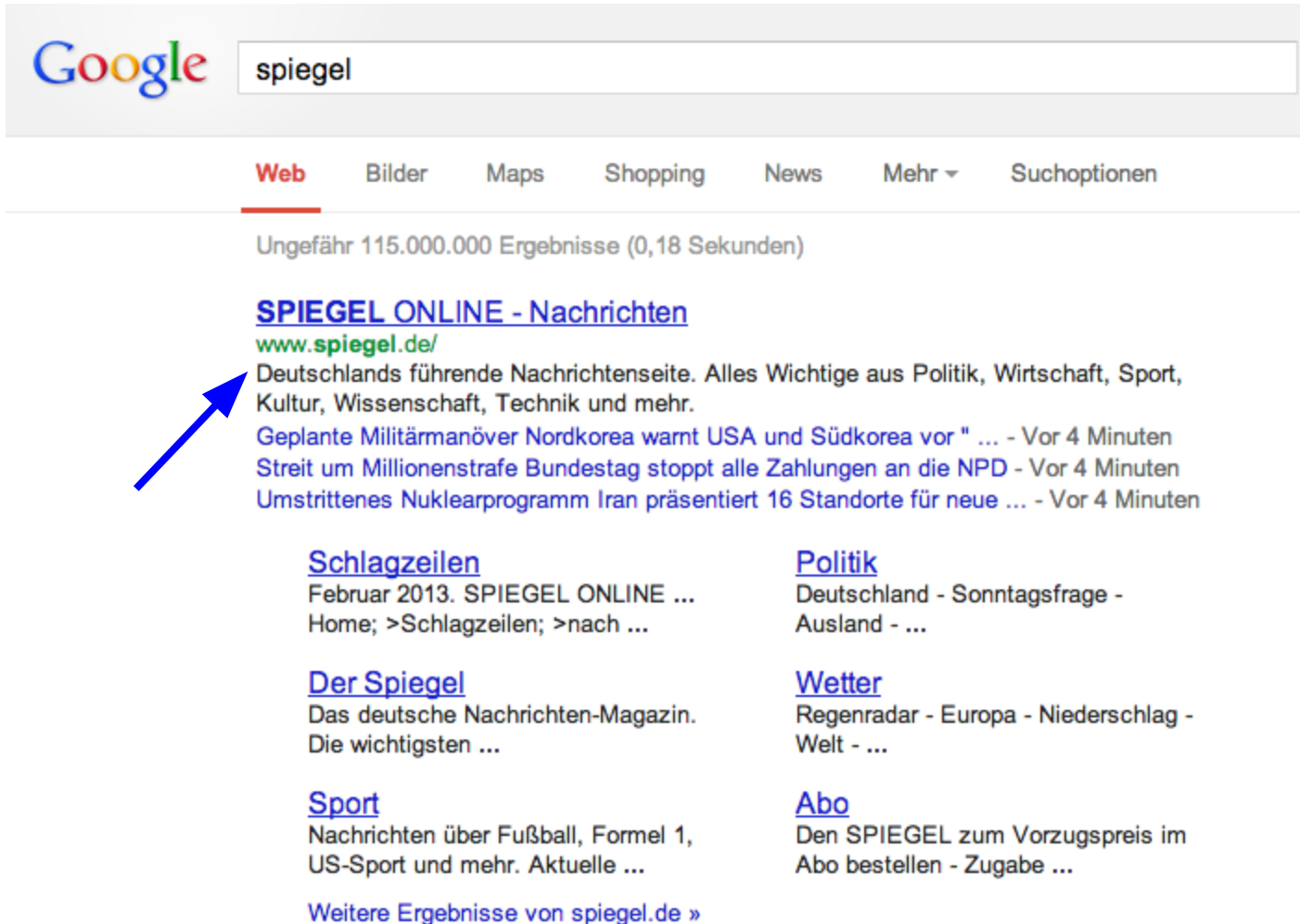
[Weitere Ergebnisse von bild.de »](#)

Beispiel 1: Ausschnitt aus dem HTML Quelltextes von www.bild.de

```
1 <!DOCTYPE html>
2 <html lang="de" class="no-js">
3 <!--[16237890/3051/1300378794000]-->
4 <head>
5 <meta charset="utf-8" />
6 <title>
7     Aktuelle Nachrichten -
8     Bild.de</title>
9 <meta name="keywords" content="Bild-Channel Home" />
10 <meta name="description" content="BILD.de: Die Seite 1 für aktuelle Nachrichten und Themen,
11 Bilder und Videos aus den Bereichen News, Wirtschaft, Politik, Show, Sport, und Promis." />
12 <meta name="robots" content="index, follow, noarchive, noodp" />
13 <link rel="search" title="Bild.de" href="/opensearch.xml"
14 type="application/opensearchdescription+xml" />
15 <link rel="canonical" href="http://www.bild.de/" />
16 <link rel="publisher" href="https://plus.google.com/114953790992839012570/" />
17 <link rel="shortcut icon" href="http://bilder.bild.de/fotos/favicon-28103320/Bild/1.bild.ico">
18 <link rel="apple-touch-icon" href="http://bilder.bild.de/fotos/apple-touch-icon-144px-
19 28104066/Bild/2.bild.png">
20 <link rel="alternate" title="Bild.de" href="http://www.bild.de/rss-feeds/rss-
21 16725492,feed=home.bild.html" type="application/rss+xml" />
22 <meta property="fb:app_id" content="78193590860"/>
23 <meta property="fb:admins" content="501487474, 100000905792367"/>
24 <meta property="og:title" content="BILD.de"/>
25 <meta property="og:type" content="website"/>
26 <meta property="og:url" content="http://www.bild.de/" />
27 <meta property="og:description" content="BILD.de: Die Seite 1
28 für aktuelle Nachrichten und Themen, Bilder und Videos aus den Bereichen News, Wirtschaft,
29 Politik, Show, Sport, und Promis." />
```

Hiermit teilt www.bild.de Suchmaschinen mit, welchen snippet-Text sie gerne dem Nutzer zeigen würden. Das ist auch der Text, den die Google Suche für diese Website anzeigt.

Beispiel 2: Websuche auf www. google.de nach "spiegel"



Google

Web Bilder Maps Shopping News Mehr ▾ Suchoptionen

Ungefähr 115.000.000 Ergebnisse (0,18 Sekunden)

SPIEGEL ONLINE - Nachrichten
www.spiegel.de/
Deutschlands führende Nachrichtenseite. Alles Wichtige aus Politik, Wirtschaft, Sport, Kultur, Wissenschaft, Technik und mehr.
Geplante Militärmanöver Nordkorea warnt USA und Südkorea vor " ... - Vor 4 Minuten
Streit um Millionenstrafe Bundestag stoppt alle Zahlungen an die NPD - Vor 4 Minuten
Umstrittenes Nuklearprogramm Iran präsentiert 16 Standorte für neue ... - Vor 4 Minuten

Schlagzeilen
Februar 2013. SPIEGEL ONLINE ...
Home; >Schlagzeilen; >nach ...

Der Spiegel
Das deutsche Nachrichten-Magazin.
Die wichtigsten ...

Sport
Nachrichten über Fußball, Formel 1,
US-Sport und mehr. Aktuelle ...

Politik
Deutschland - Sonntagsfrage -
Ausland - ...

Wetter
Regenradar - Europa - Niederschlag -
Welt - ...

Abo
Den SPIEGEL zum Vorzugspreis im
Abo bestellen - Zugabe ...

[Weitere Ergebnisse von spiegel.de »](#)

Beispiel 2: Ausschnitt aus dem HTML Quelltextes von www.spiegel.de

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
  "http://www.w3.org/TR/html4/loose.dtd"><html>
2 <head>
3     <title>SPIEGEL ONLINE - Nachrichten</title>
4     <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
5 <meta http-equiv="X-UA-Compatible" content="IE=edge" />
6 <meta name="author" content="SPIEGEL ONLINE, Hamburg, Germany" />
7 <meta name="copyright" content="SPIEGEL ONLINE, Hamburg, Germany" />
8 <meta name="email" content="spiegel_online@spiegel.de" />
9 <meta name="MSSmartTagsPreventParsing" content="true" />
10 <meta http-equiv="imagetoolbar" content="no" />
11 <meta name="keywords" content="SPIEGEL ONLINE, DER SPIEGEL, Nachrichten, News,Home" />
12 <meta name="description" content="Deutschlands führende Nachrichtenseite. Alles
    Wichtige aus Politik, Wirtschaft, Sport, Kultur, Wissenschaft, Technik und mehr."
```

Hiermit teilt www.spiegel.de Suchmaschinen mit, welchen snippet-Text sie gerne dem Nutzer zeigen würden. Das ist auch der Text, den die Google Suche für diese Website anzeigt.

Beispiel 3: Google Web Suche auf www.google.de nach "Nordkorea warnt USA"

Google Nordkorea warnt USA

Web Bilder Maps Shopping News Mehr ▾ Suchoptionen

Ungefähr 317.000 Ergebnisse (0,45 Sekunden)

News zu Nordkorea warnt USA

Geplante Militärmanöver: **Nordkorea warnt USA** und Südkorea vor "Aggressionskrieg"
Spiegel Online - vor 22 Minuten
Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer Befehlshaber hat einen **US-General** telefonisch vor einer Eskalation der Lage ...

USA und Japan warnen **Nordkorea** wegen Atombombentests
DiePresse.com - vor 19 Stunden
Nach Raketestest: **Experte warnt: Nordkoreas** Atomprogramm macht große Fortsch...
FOCUS Online - vor 6 Tagen

Nordkorea warnt USA und Südkorea vor "Aggressionskrieg"
ml.spiegel.de/article.do?id=885191
vor 31 Minuten – Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer Befehlshaber hat einen **US-General** telefonisch vor einer Eskalation der Lage ...

SPIEGEL ONLINE - Nachrichten
www.spiegel.de/
Geplante Militärmanöver: **Nordkorea warnt USA** und Südkorea vor "Aggressionskrieg" - Fotos. Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer ...

Nordkorea warnt die USA
german.irib.ir/nachrichten/.../217529-nordkorea-warnt-die-us...
vor 5 Stunden – Pjongjang (nzz.ch) - **Nordkorea** hat die **USA** aufgefordert, ein für März geplantes Militärmanöver mit Südkorea abzusagen. Die jährliche ...

Nordkorea warnt die USA - NZZ.ch, 23.02.2013
www.nzz.ch/aktuell/.../nordkorea-warnt-die-usa-1.18019768
vor 8 Stunden – Die jährliche Militärübung sei ein Beleg für die Feindseligkeit der **USA** gegenüber **Nordkorea**, meldeten staatliche Medien.

Nordkorea warnt die USA vor einem Nuklearkrieg - Die Welt
www.welt.de > Politik > Ausland
02.10.2012 – Nur ein Funke genüge, um einen Nuklearkrieg auszulösen, **warnt Nordkoreas** Vizeaußenminister Pak vor der UN. Die **USA** hätten die ...

USA und Japan warnen **Nordkorea** wegen Atombombentests ...
diepresse.com > Politik > Außenpolitik
vor 20 Stunden – Beide Länder seien zu "starkem Handeln" entschlossen, sagt **US-Präsident Obama** nach einem Treffen mit Japans Regierungschef Shinzo ...


Es wird von Google **keine** Werbung eingeblendet da es keine relevante Werbung zu dieser Suchanfrage gibt.


Beispiel 3: Google News Suche auf news.google.de nach "Nordkorea warnt USA"


Google Nordkorea warnt USA

Web Bilder Maps Shopping **News** Mehr ▾ Suchoptionen


Ungefähr 2.700 Ergebnisse (0,38 Sekunden)


 **Geplante Militärmanöver: Nordkorea warnt USA und Südkorea vor ...**
Spiegel Online - vor 20 Minuten
Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer Befehlshaber hat einen **US-General** telefonisch vor einer Eskalation der Lage ...


 **Nordkorea warnt die USA**
WZZ Online - vor 7 Stunden
Alle 5 Nachrichtenquellen »

 **USA und Japan warnen Nordkorea wegen Atombombentests**
DiePresse.com - vor 19 Stunden
Nach dem jüngsten **nordkoreanischen** Atomtest haben die **USA** und Japan eine scharfe Warnung an Pjöngjang gerichtet. Angesichts der ...

Nach Raketentest: Experte warnt: Nordkoreas Atomprogramm macht ...
FOCUS Online - 17.02.2013
Zu Recht, wie Experte Siegfried Hecker jetzt im FOCUS **warnt**. Der **US**-Atomphysiker Siegfried Hecker sieht **Nordkorea** auf dem Weg zur ...

 **USA warnen Nordkorea vor weiteren Provokationen**
Derwesten.de - 16.02.2013
Die **USA** haben **Nordkorea** vor weiteren Atomtests gewarnt. Die Regierung in Pjöngjang müsse von "zusätzlichen Provokationen" Abstand ...
+ Mehr anzeigen

 **Atomstreit mit Nordkorea: Südkorea warnt vor neuen Atomtests**
STERN.DE - 05.02.2013
Die **USA** und ihre Verbündeten hoffen, durch den bevorstehenden **nordkoreanischen** Atomtest Erkenntnisse über den aktuellen Stand des ...

 **Nach dem neuen Atomtest: Nordkorea-Debatte setzt China unter ...**
Handelsblatt - 13.02.2013
UND DANN WARNT NORDKOREA ÜBER HINTERTÜREN DEN USA VON DER ABSICHT ATOMTESTS MACHEN ZU WOLLEN...WOHL AUS ...
+ Mehr anzeigen

Lawrow: Kernwaffentests in Nordkorea kein Anlass für militärische ...
RIA Novosti - 22.02.2013
Moskau **warnt** laut dem russischen Außenminister Sergej Lawrow vor einem Nutzen der Situation um den Kernwaffentest in **Nordkorea** als ...

Auf Google News wird von Google grundsätzlich keine Werbung eingeblendet.

Beispiel 3: Der Artikel auf www.spiegel.de

Gewinner in der Kategorie
„Bester Sound und beste Beats“

SPIEGEL ONLINE POLITIK

NACHRICHTEN VIDEO THEMEN FORUM ENGLISH DER SPIEGEL SPIEGEL TV ABO SHOP Energie


Home Politik Wirtschaft Panorama Sport Kultur Netzwelt Wissenschaft Gesundheit einestages Karriere Uni Schule Reise Auto

Nachrichten > Politik > Ausland > Nordkorea > Militärmanöver: Nordkorea droht USA und Südkorea Login | Registrierung

Geplante Militärmanöver: Nordkorea warnt USA und Südkorea vor "Aggressionskrieg"

Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer Befehlshaber hat einen US-General telefonisch vor einer Eskalation der Lage gewarnt. Anlass sind geplante militärische Großmanöver der Amerikaner zusammen mit Südkorea.

ANZEIG Was ist in der Pipeline für die Energiezukunft Deutschlands? Stakol



Fotos ▶

REUTERS/ KCNA

Seoul - Vor Beginn zweier Großmanöver der US-Streitkräfte mit Südkorea hat das kommunistische Nordkorea erneut scharfe Töne angeschlagen. Die Volksarmee unterstellte beiden Ländern am Samstag, einen "Aggressionskrieg" vom Zaun brechen zu wollen.

Diejenigen, die einen Krieg auslösten, seien "zu einer elenden Zerstörung bestimmt", drohte der Militärvertreter Pak Rim Su im Grenzort Panmunjom in einer Botschaft an den Befehlshaber der US-Streitkräfte in Südkorea, James Thurman. Die Botschaft sei telefonisch übermittelt worden, berichteten die Staatsmedien.

Die Lage auf der

ANZEIG Die Worte zum Sonntag. IVO



Das neue
HTC ONE

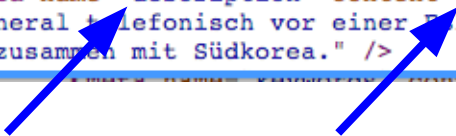
Nur bei O₂
inkl. gratis Beats™
by Dr. Dre Kopfhörer



Möglichkeiten der Refinanzierung

Beispiel 3: Ausschnitt des HTML Quelltextes des Artikels auf www.spiegel.de

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/loose.dtd"><html>
2 <head>
3     <title>Militärmanöver: Nordkorea droht USA und Südkorea - SPIEGEL ONLINE</title>
4     <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
5 <meta http-equiv="X-UA-Compatible" content="IE=edge" />
6 <meta name="author" content="SPIEGEL ONLINE, Hamburg, Germany" />
7 <meta name="copyright" content="SPIEGEL ONLINE, Hamburg, Germany" />
8 <meta name="email" content="spiegel_online@spiegel.de" />
9 <meta name="MSSmartTagsPreventParsing" content="true" />
10 <meta http-equiv="imagetoolbar" content="no" />
11 <meta name="date" content="2013-02-23T16:39:00+0100" />
12     <meta name="last-modified" content="2013-02-23T17:23:03+0100" />
13     <meta property="og:site_name" content="SPIEGEL ONLINE" />
14     <meta property="og:url" content="http://www.spiegel.de/politik/ausland/militaermanoever-nordkorea-
15 droht-usa-und-suedkorea-a-885191.html" />
16     <link rel="canonical" href="http://www.spiegel.de/politik/ausland/militaermanoever-nordkorea-droht-usa-
17 und-suedkorea-a-885191.html" />
18     <meta property="og:image" content="http://cdn2.spiegel.de/images/image-464895-thumb302-
19 tmsf.jpg" /><meta name="twitter:card" value="summary" />
20     <meta name="twitter:site" value="@SPIEGELONLINE" />
21     <meta property="og:type" content="article" />
22     <meta property="og:title" content="Geplante Militärmanöver: Nordkorea warnt USA und Südkorea vor
23 &quot;Aggressionskrieg&quot; - SPIEGEL ONLINE" />
24     <meta property="og:description" content="Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer
25 Befehlshaber hat einen US-General telefonisch vor einer Eskalation der Lage gewarnt. Anlass sind geplante militärische
26 Großmanöver der Amerikaner zusammen mit Südkorea." />
27     <meta name="description" content="Der Ton wird rauer im Korea-Konflikt: Ein nordkoreanischer Befehlshaber hat
28 einen US-General telefonisch vor einer Eskalation der Lage gewarnt. Anlass sind geplante militärische Großmanöver der
29 Amerikaner zusammen mit Südkorea." />
```



Hiermit teilt spiegel.de Suchmaschinen mit, welchen snippet-Text sie gerne dem Nutzer zeigen würden. Etwas gekürzt ist das auch der Text, den die Google Suche und Google News für diesen Artikel angezeigt haben.

Beispiel 4: robots.txt Datei von www.spiegel.de

```
User-agent: WebReaper  
Disallow: /  
User-agent: Slurp  
Crawl-delay: 18
```

Der User-agent ist die eindeutige "Visitenkarte" eines Crawlers, mit der er dem Webseiten-Betreiber mitteilt, wer er ist und was er macht. Der Webseiten-Betreiber kann dann entscheiden und in der Datei robots.txt oder per REP festlegen, für welche "User-agents" welche Regeln gelten sollen.

Die Zeilen 1 und 2 legen fest, dass alle Crawler, außer dem Crawler mit dem User-agent *WebReaper*, die Seite www.spiegel.de durchsuchen dürfen.

In Zeilen 3 und 4 wird definiert, dass der Crawler mit dem User-agent *Slurp* nur maximal alle 18 Sekunden auf www.spiegel.de zugreifen darf.

Beispiel 5: REP Meta-Tags

Die Nutzung von REP ist denkbar simpel. Wenn zum Beispiel ein Verlag nicht möchte, dass zusätzlich zu einem Link auch ein Snippet seines Artikels in der allgemeinen Google Suche erscheint, genügt folgende kurze Eingabe auf der entsprechenden Webseite:

```
<meta name="googlebot" content="nosnippet">
```