Datenschutzrechtliche Prüfung von Websites

Sayeed Klewitz-Hommelsen

Fachbereich Angewandte Informatik, Fachhochschule Bonn-Rhein-Sieg
D-53757 Sankt Augustin, Grantham Allee 20
Sayeed.klewitz-hommelsen@fh-bonn-rhein-sieg.de

Schlagworte: Datenschutz, Selbsttest, p3p, platform for privacy preferences, Suchmaschine, Selbstkontrolle, cookies, JavaScript, SaD, System zur automatisierten Datenschutzprüfung

Abstract: Die Prüfung von Internetseiten durch die für den Datenschutz verantwortlichen Personen hat einen Umfang erreicht, dass er manuell praktisch kaum noch zu leisten ist. Ziel des Projektes SaD (System zur automatisierten Datenschutzprüfung) ist es, die Verantwortlichen bei der Prüfung zu unterstützen und auf automatisiert ermittelbare problematische Seiten hinzuweisen. Dabei verbleibt die abschließende Bewertung beim Benutzer. Das System liefert eine Zusammenstellung von Seiten, die unter verschiedenen Aspekten datenschutzrechtlich problematisch sein können. Außerdem wird eine Statistik erzeugt, die es dem Verantwortlichen ermöglicht, seine Seite im Verhältnis zu anderen zu sehen.

1. Europäisches Datenschutzrecht als Handlungsparameter

Europa hat sich in den letzten Jahren zu einer wesentlichen Triebfeder der Rechtssetzung im Technologierecht entwickelt. Dies gilt im speziellen auch für den Datenschutz. Beginnend mit der Datenschutzrichtlinie von 1995[1] über eine Reihe weiterer Normsetzungen speziell im Datenschutzbereich. Die Bundesrepublik Deutschland hat die Umsetzung der europäischen Vorschriften nicht fristgerecht durchgeführt. Letztlich wurde im Jahr 2001 das neue Bundesdatenschutzgesetz (BDSG 2001) erlassen. Darüber hinaus haben sich einige der spezialgesetzlichen Vorschriften auf Veranlassung europäischer Normsetzungen[2] hin geändert, so zum Beispiel das Teledienstedatenschutzgesetz und der Mediendienstestaatsvertrag.

---

[1] Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr.

[2] Richtlinie 2000/31/EG des Europäischen Parlaments und des Rates vom 8. Juni 2000 über bestimmte rechtliche Aspekte der Dienste der Informationsgesellschaft, insbesondere

Die europäischen Initiativen zielen dabei auf die Harmonisierung des europäischen Rechtsraumes ab[3]. Ob der Weg über die nationalen Umsetzung sehr glücklich gewählt ist, braucht hier nicht problematisiert zu werden.

Der Schutz der persönlichen Daten knüpft an den Begriff der personenbezogenen Daten[4] an: Diese werden definiert als Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (§ 3 Abs. 1 BDSG 2001). Diese weite Definition schließt beinahe alle Kommunikationsdaten von natürlichen Personen ein, da diese heute fast immer auf die ein oder andere Weise durch technische Mittel auf die konkrete Person bezogen werden können.

2. Die Projektidee

Zur Überprüfung der Datenschutzvorschriften sind mehrere Kontrollinstanzen etabliert worden. Im Bereich der öffentlichen Verwaltung in Deutschland hat sich dabei die Institution des Datenschutzbeauftragten bewährt. Der Bundesbeauftragte (§ 24 Abs. 1 BDSG 2001) und die Landesbeauftragten für den Datenschutz (nach jeweiligen Landesdatenschutzgesetzen) überprüfen die Einhaltung von Datenschutzvorschriften aus externer Sicht. Daneben haben öffentliche Stellen eigene Datenschutzbeauftragte einzusetzen (§ 4f Abs. 1 BDSG 2001) deren Aufgabe darin besteht, auf die Einhaltung der Vorschriften über den Datenschutz vor Ort hinzuwirken (§ 4g Abs. 1 BDSG 2001). Dazu gehört heute typischerweise auch die Prüfung der Internetangebote einer Behörde auf Konformität mit dem Datenschutzrecht.

Betrachtet man den Umfang der Internetangebote der Behörden im Zeitalter von E-Government[5], wird schnell deutlich, dass dies praktisch nicht mehr geleistet werden kann. Der Umfang der Internetseiten erreicht schnell Volumina im Umfang mehrstelliger Megabytebeträge[6], die manuell nicht mehr sinnvoll geprüft werden können.

---

des elektronischen Geschäftsverkehrs, im Binnenmarkt (ABl. EG Nr. L 178 S. 1), http://europa.eu.int/eur-lex/pri/de/oj/dat/2000/l_178/l_17820000717de00010016.pdf (10.04.2003).

[3] Nr. 8 der Erwägungsgründe der EU-Richtlinie über den elektronischen Geschäftsverkehr, vgl. Fn. 2.

[4] Art. 2 a) Richtlinie 95/46/EG, Fn 1.

[5] *Von Lucke, Jörn:* Portale für die öffentliche Verwaltung, in: *Reinermann/von Lucke* (Hrsg.): Portale in der öffentlichen Verwaltung, Forschungsinstitut für öffentliche Verwaltung, Speyer 2000, S. 7 ff., 12.

[6] Dies wurde sehr schnell zu einem der Probleme im Projekt, da alle Seiten für eine Analyse vom Server kopiert werden müssen.

Aus dieser Situation heraus entstand die Idee, Internetseiten mittels einer Suchmaschine einer Vorprüfung zu unterziehen, um diejenigen Seiten zu ermitteln, die einer näheren, menschlichen Überprüfung zu unterziehen sind. Aus der Entstehungsgeschichte heraus, sollte das System weniger der externen Datenschutzkontrolle dienen sondern in erster Linie die Selbstkontrolle unterstützen. Die intern Verantwortlichen für den Internetauftritt und die betrieblichen Datenschutzbeauftragen sollten ein Instrument in die Hand bekommen, die Internetangebote zu prüfen, für die Sie verantwortlich sind. Zugleich sollte auf diesem Weg ein Missbrauch des Systems verhindert werden. Das System läuft zur Zeit auf einem Hochschulserver und ist damit auch unverdächtig, heimlich der externen Datenschutzkontrolle zu dienen.

Das System behält dabei die externe Brille auf und überlässt die interne Prüfung den jeweils Verantwortlichen. Hier sind insbesondere die organisatorischen und technischen Anforderungen an eine E-Government Plattform zu beachten[7].

3. Anforderungen an externe Datenschutzprüfung

Die Anforderungen, die von Seiten des Datenschutzes an Internetauftritte gestellt werden, sind vielfältig und können nicht sämtlich automatisiert werden. Insgesamt hat der Umgang mit personenbezogenen Daten rechtmäßig zu erfolgen. Das bedeutet, die Daten müssen rechtmäßig erhoben werden. Es sollen so wenig wie notwendig personenbezogene Daten erfasst werden (Datensparsamkeit). Zugleich sollen die Daten für einen bestimmten Zweck erhoben werden. Die Betroffenen sollen über diese Zwecke und Widerspruchsrechte informiert werden (§ 4 Teledienstedatenschutzgesetz). Die erhebende Stelle hat ihrerseits ihre Identität offen zu legen (§ 6 Teledienstegesetz).

Zugleich haben die Betreiber von Medien- oder Telediensten die notwendige Sicherheit für die Nutzer ihrer Seiten sicher zu stellen (§ 9 BDSG 2001). Dabei sind die interaktiven Schnittstellen zwischen externen Nutzern und dem Anbieter besonders kritisch.

Last but not least sollten öffentliche Stellen beim Datenschutz eine Vorreiterstellung einnehmen und internationale Standards zum Datenschutz

---

[7] *Sonntag, Michael/Wimmer, Maria:* Datenschutzaspekte von e-Government mit besonderem Bezug auf das eGOV-Projekt, in: *Schuber, Sigrid/Reusch, Bernd/Jesse, Norber* (Hrsg.): Informatik bewegt, Lecuture Notes in Informatics – Proceedings Volume P-19, S. 462 ff.

implementieren. Hier ist etwa an das p3p-Protokoll[8] zu denken, das es den Betreibern von Seiten ermöglicht, die eigene Datenschutzpolicy zu veröffentlichen und Nutzern, dies beim Surfen zu berücksichtigen.

4. Möglichkeiten externer Datenschutzprüfung

Aus den vorstehend genannten Anforderungen wurde eine dedizierte Suchmaschine entwickelt, die eine Reihe der genannten Anforderungen, soweit dies technisch möglich erschien, erfüllt.

4.1. Verwendung von Cookies

Sowohl aus technischen Gründen als auch aus Achtlosigkeit, kommen immer wieder Cookies zum Einsatz. Dabei handelt es sich um kleine Dateien, die ein Webserver auf dem lokalen Rechner des Surfers ablegen kann. Dadurch wird dieser Rechner und damit meist auch der jeweilige Nutzer für den Server identifizierbar. Werden die Cookies nach einer Session wieder gelöscht, ist dies regelmäßig unkritisch. Problematisch wird dies, wenn die Cookies längerfristig auf dem Rechner verbleiben, um den jeweiligen Surfer auch nach längerer Zeit wieder zu identifizieren, wie es verschiedene Ad-Server praktizieren[9].

Das System muss deshalb Cookies erkennen und vor allem feststellen, ob Sie zum längeren Verbleib auf dem Rechner gesetzt werden.

4.2. Einsatz von JavaScript

JavaScript ermöglicht es dem Webseitenprogrammierer, auf dem Rechner des Surfers Programme ablaufen zu lassen, typischerweise um die Interaktion mit dem Benutzer komfortabel zu gestalten. Aufgrund der Möglichkeiten von JavaScript wird immer wieder aus Sicherheitsgründen empfohlen, JavaScript zu deaktivieren bzw. nur gezielt einzuschalten.[10] Insbesondere öffentliche Seitenbetreiber sollten hier als Vorbild voranschreiten. Das Deaktivieren von JavaScript bedeutet allerdings meist, dass der Nutzer auf das ein oder andere „Schmankerl" verzichten muss. Im schlimmsten Falle – und solche wurden beobachtet – lässt sich eine solche Seite nicht mehr ansehen.

---

[8] World wide web consortium, Platform for Privacy Preferences Initiative: http://www.w3.org/P3P/ (10.04.2003).

[9] Europäisches Verbrauchschutzzentrum: safer surfen – selbst sichern, http://www.datenschutzzentrum.de/material/themen/safesurf/safer/ (10.04.2003) unter dem Stichwort ‚Cookies'.

[10] Landesbeauftragter für den Datenschutz, Tipps für Bürger JavaScript, http://www.datenschutzzentrum.de/material/themen/safesurf/safer/ (10.04.2003)

Das System muss also feststellen, ob JavaScript zum Einsatz kommt und dies vermelden.

4.3. Sichere Formulare und Datenübertragung

Immer dann, wenn es dem Nutzer möglich sein soll, mit dem Anbieter in Kontakt zu treten, kommen Formulare für die Dateneingabe zum Einsatz. Dabei werden typischerweise Daten des Nutzers übertragen, wie z.B. Adressdaten oder Login-Kennungen.

Das System hat deshalb zu prüfen, ob die Daten – insbesondere beim Login – verschlüsselt übertragen werden. Außerdem weist es alle Stellen nach, an denen Eingaben möglich sind. Es ist allerdings schwierig, sicher festzustellen, um welche inhaltlichen Daten es sich handelt: Name, Adresse, Alter, Einkommen, Emailadresse u.ä. Werden potentiell problematische Abfragen gefunden, wird darauf hingewiesen. Nicht verschlüsselte Logins bzw. Passwortabfragen werden angemerkt. Hier liegt sicher eines der größten Sicherheitsprobleme.

Weiterhin bieten Formulare die Möglichkeit, verdeckt Felder mit Werten zu belegen, die den Nutzer zwar identifizieren können, diesem aber gar nicht bekannt werden. Solche verdeckten Datenfelder werden ebenfalls aufgespürt.

Immer dann, wenn personenbezogene Daten erhoben werden, muss der Betroffene auch über den Zweck der Erhebung, sein Widerspruchsrecht und anderes informiert werden (z.B. § 4 Teledienstedatenschutzgesetz). Das System sucht auf den jeweiligen Seiten nach einschlägigen Schlüsselwörtern, die auf eine solche Information hindeuten. An dieser Stelle sind noch Fortentwicklungen geplant, um die Erkennungsrate zu optimieren.

4.4. Adressangaben auf Internetseiten

Immer wieder stellt man fest, dass auf Internetseiten Adressangaben veröffentlicht werden. Dies kann einer gesetzlichen Verpflichtung entsprechen (Nennung des Verantwortlichen) oder gerade die Absicht des Verfassers darstellen (Nennung von Beratungsstellen und Ansprechpartnern). Es kann aber auch unbeabsichtigt geschehen. Um diese Fälle zu identifizieren, werden die Internetseiten nach Adressangaben und Emailkennungen durchsucht. Dabei erfolgt die Suche mit Hilfe von regulären Ausdrücken.

4.5. Vorhandensein von p3p Informationen

Die platform for privacy preferences soll es einem Informationsanbieter ermöglichen, in standardisierter Weise den Umgang mit personenbe-

zogenen Daten darzulegen, so dass der Browser des Nutzers, diese Angaben lesen und ggf. darauf reagieren kann. Moderne Browser werten die p3p-Informationen.[11] Das Vorhandensein dieser Informationen wird geprüft und vermerkt.

4.6. Untersuchung von Multimedia-Datentypen

Andere Datentypen als Textinformationen können bisher nur mit geringem Erfolg auf ihren Inhalt hin ausgewertet werden. Das System sieht deshalb ganz davon ab, bemerkt solche Datentypen (z.B. Video) aber und gibt dieses im Ergebnisbericht an. Der Grund dafür liegt darin, dass gerade in solchen Multimediadateien personenbezogene Daten vorhanden sein können. Die Prüfung obliegt allerdings dem Verantwortlichen.

4.7. Suche nach Kontaktinformationen

Nach verschiedenen deutschen Vorschriften, die ihrerseits auf den europäischen Vorgaben beruhen[12], müssen sich die Betreiber von Internetseiten identifizieren (z.B. § 10 Mediendienstestaatsvertrag vom 20.12.2001, § 6 Teledienstegesetz 2001). Diese Angaben müssen gut sichtbar und leicht zugänglich sein. Auch hier wird versucht, durch eine Schlüsselwortsuche, diese Informationen zu identifizieren. Besonders problematisch ist Verknüpfung von Kontaktinformationen mit JavaScript, denn in diesem Falle kann ein Nutzer, der aus Sicherheitsgründen JavaScript abgeschaltet hat, nicht auf diese wesentlichen Informationen zugreifen.

4.8. Überprüfung von Email-Adressen zur Kontaktaufnahme

Üblicherweise werden bei Internetangeboten Emailadressen zur Kontaktaufnahme angeboten. Leider muss man bisweilen feststellen, dass diese Adressen nicht bedient werden. Im Probebetrieb wird deshalb ebenfalls getestet, ob die entsprechende Emailadresse auch beantwortet wird. Dieses Feature ist aber noch nicht im laufenden Betrieb in die Auswertung eingebunden.

---

[11] Vgl. p3p Fn. 8; Schleswig-Holsteiner Datenschutzbeauftragter, 24. Tätigkeitsbericht, 2002, Kap. 11. http://www.datenschutzzentrum.de/material/tb/tb24/kap11.htm#Tz11.5 (10.4.2003).

[12] Richtlinie 2000/31/EG vgl. Fn. 2.

5. Ergebnis der Prüfung

Das System zur Automatisierten Datenschutzprüfung erzeugt am Ende einen Bericht, in dem der Auftraggeber über alle gefundenen Punkte, die zuvor genannt worden sind, informiert wird. Der Bericht wird nicht selbst übermittelt sondern es wird ein Link auf den SaD Server geschickt, auf dem das gesamte Ergebnis gespeichert wird. Dabei werden lediglich die Zusammenfassungen aufbewahrt und für statistische Auswertungen vorgehalten.

6. Schutz vor Missbrauch

Ein Problem beim Systemdesign stellte der Schutz vor einer missbräuchlichen Nutzung dar. So listet das System beispielsweise alle gefundenen Emailadressen auf. Um diesem entgegen zu wirken, wurden mehrere Schranken aufgebaut:

Der Scan einer Site kann nur von einer Person gestartet werden, die eine Emailadresse aus dem Adressbereich der zu prüfenden Domain besitzt. Um sicherzustellen, dass diese Person auch tatsächlich die Prüfung angefordert hat, wird zuvor eine Prüfemail an diese Adresse mit einem Aktivierungskey gesendet. Erst wenn dieser eingegeben wurde, wird die Durchsuchung der Site gestartet. Außerdem erhält der admin bzw. der webadmin eine Email, die ihn über den Suchvorgang informiert, damit diese Person nicht durch irgendwelche Ergebnisse überrascht wird.

Schließlich werden bestimmte Adressen von vorn herein gegen eine Blacklist geprüft. So können z.B. Nutzer der großen Freemail-Account Anbieter, den Dienst für ihren Provider nicht aufrufen.

7. Ausblick

Das System befindet sich zur Zeit in einer Testphase, in der jeder auf den Dienst zugreifen kann. In erster Linie wird an der Optimierung des Codes gearbeitet. Die zeitkritischen Passagen, die bisher durch Scriptsprachen realisiert worden waren, werden nun in C umgeschrieben und verfeinert. Auch an der Reduktion der zur Prüfung zu übermittelnden Daten wird gearbeitet. So werden im Augenblick Algorithmen getestet, die für bestimmte Bereiche eines Webangebotes eine Vorhersage treffen sollen, ob es sich um potentiell relevante Bereiche (für eine weitere Suche) handelt. Hier sind noch aufwendigere Tests notwendig.

Soweit im System landessprachliche Besonderheiten genutzt werden (z.B. bei den Wortlisten) bestehen Überlegungen, das System auf andere

europäische Sprachen auszudehnen, da zumindest die rudimentären Kriterien, nach denen das System prüft, für alle EU-Länder gleich sind.

Darüber hinaus wird über den Aufbau eines „Datenschutzindex" nachgedacht. Dieser soll als eine Art Kennzahl die Datenschutzrelevanz eines Internetangebotes bewerten. Die bisherigen statistischen Auswertungen reichen dafür aber noch nicht aus. Hier wird insbesondere eine genauere Differenzierung notwendig sein, die erst nach einem längeren Einsatz und Test, eine sinnvolle Größe darstellen kann.

Der SaD Server ist zur Zeit unter der folgenden Internetadresse erreichbar: http://sad.inf.fh-bonn-rhein-sieg.de/.