Ontologiebasierte Unterstützung zur Modellierung datenschutzrechtskonformer Web Service

Paul Peitz, Christian Baumann, Richard Wacker, Oliver Raabe

Paul.Peitz@sap.com, Ch.Baumann@sap.com, Wacker@kit.edu, Raabe@kit.edu

Abstract: In der Vision des Internets der Dienste werden Daten zwischen verteilten Softwarekomponenten hochautomatisiert Übermittelt. Dies wirft neben den bekannten technischen Problemen auch rechtliche Probleme auf. Insbesondere der Umgang mit personenbezogenen Daten wird vom Endanwender kritisch betrachtet. Der Gesetzgeber regelt daher den Umgang mit diesen Daten im Bundesdatenschutzgesetz, welches in seiner Komplexität jedoch vom Serviceentwickler ohne zusätzliche Unterstützung praktisch nicht umgesetzt werden kann. In dieser Arbeit wird ein Ansatz gezeigt, der diese Unterstützung, basieren auf einer Formalisierung des BDSG in einer Ontologie sowie einer semantischen Servicebeschreibung, bereitstellt.

1 Einführung und Problemstellung

In der Vision des Internets der Dienste[JRS08] [HAW08, S. 100] dient die heutige Infrastruktur des Internets als Grundlage für die Beauftragung, aber auch die Erbringung von Dienstleistungen. Dabei wird durch eine Vereinheitlichung der Schnittstellen ein hohes Maß an Automatisierung erreicht, wodurch sich Web Services auf der einen Seite leicht in bestehende Abläufe integrieren lassen, was aber auch neue Probleme aufwirft.

So sind aus rechtlicher Sicht insbesondere die Komposition von Web Services [Bau08] sowie der Datenschutz kritische Punkte bei einer weitgehend automatisierten Dienstausführung. Insbesondere der unkontrollierte Fluss personenbezogener Daten über Unternehmensgrenzen hinweg durch Web Services könnte zu einem Vertrauensverlust bei den Endanwendern führen, der einer weiteren Verbreitung des Internets der Dienste im Wege stehen kann.

Diese Problematik wurde auch vom Gesetzgeber erkannt, der mit dem Bundesdatenschutzgesetz (BDSG)[TEG04] ein umfassendes Regelwerk geschaffen hat, das den Umgang mit personenbezogenen Daten regelt. Web Services werden jedoch in der Regel durch Serviceentwickler erstellt, die typischerweise ein hohes Maß an technischem und fachlichem Wissen besitzen, in Bezug auf diese rechtlichen Rahmenbedingungen aber Laien sind, was zu unbeabsichtigten Verstößen gegen geltendes Recht führen kann.

In der vorliegenden Arbeit wird ein Konzept vorgestellt, welches den Serviceentwickler bei der Entwicklung datenschutzrechtskonformer Web Services unterstützt, indem die relevanten rechtlichen Grundlagen bedarfsgerecht aufbereitet werden und mit zusätzlichen Informationen ergänzt werden, um so eine fundierte Entscheidung zu ermöglichen.

Die hier vorgestellte Arbeit gliedert sich in drei Teile: In Kapitel 2 werden die Grundlagen der Formalisierung des Rechts sowie eine Datenschutzontologie vorgestellt. Diese dient als Grundlage für die im Kapitel 3 beschriebene Unterstützung des Serviceentwicklers durch eine verständliche Erklärung der relevanten Rechtsbegriffe. Abschließend werden die Ergebnisse in Kapitel 4 zusammengefasst sowie die nächsten möglichen Schritte zur Erweiterung des Ansatzes skizziert.

2 Formalisierung des Rechts

Eine Grundvoraussetzung, um die in der Einführung genannten Ziele zu erreichen ist es, dass die zu beachtenden Gesetze in einer Form vorliegen, die eine Verarbeitung durch den Computer ermöglicht. Diese formalisierte Darstellung muss aus den Gesetzen, die in natürlicher Sprache vorliegen, abgeleitet werden. Dazu gibt es zwei grundsätzliche Ansätze: Eine automatische Formalisierung mit Hilfe von NLP Technologien sowie eine manuelle Übertragung. Eine zweite Dimension bei der Formalisierung stellt die Sprache dar, in die der Gesetzestext transformiert wird. Hier stehen z.B. klassische Ansätze wie Business Rules oder aber semantische Sprachen wie OWL oder F-Logic zur Verfügung. Einen Überblick, insbesondere über die semantischen Ansätze, findet sich in [BCBG05]

Für den eingangs erläuterten Fall des Datenschutzes wurden im Rahmen des TEXO Projekts ausgewählte Paragraphen des BDSG manuell in F-Logic Regeln übertragen. Diese Datenschutzontologie soll im Folgenden vorgestellt werden.

2.1 Datenschutzontologie

Ziel der Datenschutzontologie war es, in der Ontologie nicht nur die Beziehungen zwischen Rechtsbegriffen auszudrücken, sondern die im Gesetzt gegebenen Regelungen an sich. Daher eignete sich F-Logic durch seine Regelbasiertheit besser für diese Aufgabe besser als OWL basierte Ansätze, juristische Normen lassen sich vielfach direkt in F-Logic Regeln übertragen. Des Weiteren verfügt F-Logic mit Ontobroker[1] über einen ausgereiften Reasoner sowie mit dem NeOn Toolkit[2] über eine frei verfügbare Modellierungsumgebung.

Die Übertragung des Gesetzestextes erfolgte manuell, da automatisierte Ansätze eine zusätzliche Fehlerquelle darstellen würden, die zu einer nicht akzeptablen Qualitätsminderung der Ontologie führen würden.

Abschließend wird als Beispiel die Formalisierung des 3 Abs. 7 BDSG gezeigt, die im Gesetzestext wie folgt lautet:

> Verantwortliche Stelle ist jede Person oder Stelle, die personenbezogene Daten für sich selbst erhebt, verarbeitet oder nutzt oder dies durch andere im Auftrag

[1] `http://www.ontorprise.com`
[2] `http://www.neon-toolkit.org`

vornehmen lässt.

Nach der Übertragung in F-Logic wird daraus folgendes:

```
FORALL anUmgang,aPerson
aVerantwStelle(anUmgang,aPerson):#verantwStelle AND
aVerantwStelle(anUmgang,aPerson)[#imKontextVon->anUmgang]
AND aPerson[#spielt->aVerantwStelle(anUmgang,aPerson)]
<-
aPerson:#Person AND
aPerson[#fuehrtDurch->anUmgang] AND
aPerson[#laesstDurchfuehren->anUmgang] AND
anUmgang:#Umgang.
```

3 Unterstützung des Serviceentwicklers

Basierend auf der im vorhergehenden Kapitel eingeführten Datenschutzontologie wird in diesem Kapitel gezeigt, wie sich in Anlehnung an die juristische Methodik[Lar83] eine Unterstützung des Serviceentwicklers realisieren lässt. Dabei wird analog der juristischen Methodik in zwei Schritten vorgegangen:

1. Im ersten Schritt müssen aus dem Gesetzestext alle relevanten Begriffe ermittelt werden, die als Vorbedingung der gewünschten Rechtsfolge notwendig sein. Dieser Schritt wird in Abschnitt 3.1 beschrieben.
2. Die Elemente der Servicebeschreibung müssen unter diese Begriffe subsumiert werden. Dazu ist einer weitergehende Auslegung des Gesetzestextes mit Hilfe externer Quellen notwendig.

Abbildung 1 zeigt zusammenfassend die beiden Teile.

3.1 Der Normgraph

Ausgehend von der gewünschten Rechtsfolge müssen im ersten Schritt die dazu notwendigen Voraussetzungen ermittelt werden. Diese finden sich an verschiedenen Stellen im Gesetz, da die Begriffe in der Definition der Rechtsfolge typischerweise durch weitere Rechtssätze definiert werden. Dieses Vorgehen muss rekursiv erfolgen, bis eine Ebene mit atomaren Begriffen erreicht wird, die durch Gesetzestexte nicht weiter erläutert werden. Durch dieses Vorgehen entsteht ein sogenannter Normgraph, dessen Knoten die auszulegenden Begriffe im Gesetzestext sind. Ein Beispiel ist in Abb. 1 gezeigt.

Technisch erfolgt die Umsetzung dieses Verfahren, indem die vorhandenen F-Logic Regeln ausgewertet werden. Diese bestehen aus einem Regelkopf (Head) und einen Regelrumpf

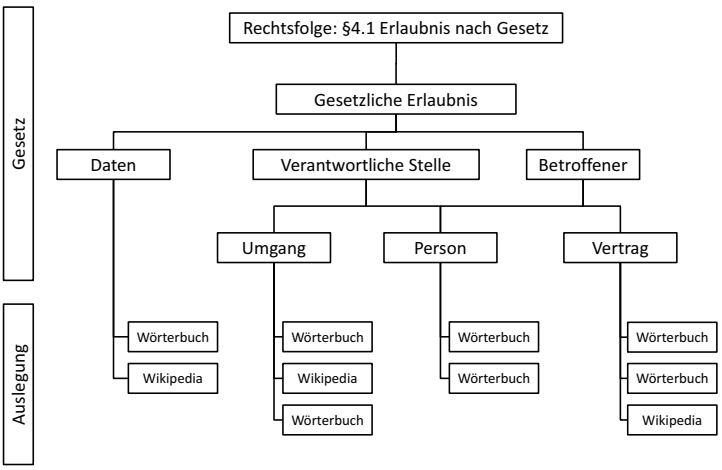

Figure 1: Normgraph und externe Quellen

(Body). Der Regelkopf beschreibt entweder die Rechtsfolge oder den zu definierenden Begriff, während die Attribute und Relationen im Rumpf die Voraussetzungen bzw. die Definition enthalten. Um den Normgraph zu erstellen werden rekursiv, ausgehen von einer gewünschten Rechtsfolge, alle Regeln gesucht, die ein Element der Rumpfes im Kopf aufweisen.

Ein Beispiel findet sich im oberen Teil der Abb. 1.

3.2 Integration externer Quellen

Selbst die Bergriffe, die sich auf der untersten Ebene des Normgraphen befinden, bedürfen häufig noch einer Auslegung, um sie mit den Sachverhaltselementen eines realen Falles in Übereinstimmung zu bringen.

Dieses Unterordnen eines konkreten Sachverhaltselements, in unserem Fall also eines Elements der Servicebeschreibung, unten einen abstrakten Begriff im Gesetz bezeichnet man als Subsumtion[Lar83]. Um diese durchführen zu können sind häufig weitere Informationsquellen notwendig. In erster Linie werden hierfür Wörterbücher sowie Enzyklopädien verwendet, die Wortdefinitionen und sprachliche Erklärungen enthalten. Typische Beispiele, die auch in dieser Arbeit verwendet werden, sind das Wortschatz-Portal der Universität Leipzig[3] oder die Wikipedia[4]. Im jetzigen Stand der Arbeit werden diese Quellen manuell in die Ontologie integriert. Dabei wurde darauf geachtet, dass die Texte allgemeinverständlich sind und somit ohne juristisches Vorwissen verwendet werden können.

[3] http://wortschatz.uni-leipzig.de/
[4] www.wikipedia.de

Figure 2: Screenshot: Auslegungshilfen

3.3 Prototyp

Um die Umsetzbarkeit der hier vorgestellten Methode zu verifizieren wurde ein Prototyp implementiert. Dieser erlaubt es, eine Rechtsfolge auszuwählen und zeigt den sich daraus ergebenden Normgraphen an. Nach der Selektion eines Begriffes auf der untersten Ebene werden die zusätzlichen Erklärungen eingeblendet (Abb. 2)

Die Implementierung basiert auf Ontobroker, einem Reasoner, der sich durch die besonders gute Unterstützung von F-Logic auszeichnet. Als graphisches Framework zur Visualisierung des Normgraphen wurde das Open Source Framework jgraphX ausgewählt. Der Prototyp wurde so konzipiert, dass sich auch zukünftige Erweiterungen, wie sie im abschließenden Kapitel beschreiben werden, integrieren lassen.

4 Zusammenfassung und Ausblick

Die Komplexität des Rechts und seiner Auslegung setzt automatisierten Ansätzen enge Grenzen. Der vorliegende Ansatz erlaubt es dennoch auch einem juristisch nicht ausgebildetem Serviceentwickler, einen Überblick über die Regelungen des BDSG im Hinblick auf den von ihm entwickelten Service zu erhalten. Ausgehen von der gewünschten Rechtsfolge, typischerweise der Erlaubnis aufgrund gesetzlicher Regelungen, werden ihm die Voraussetzungen für den von ihm entwickeltem Service erläutert. Um dies zu erreichten wurde ein Teil des BDSG mit Hilfe von F-Logic formalisiert und um externe Quellen erweitert.

Bisher erfolgt die Beurteilung der Service Elemente rein manuell durch den Serviceentwickler. Der nächste logische Schritt ist es, diese Subsumtion nicht nur durch externe Textquellen zu unterstützen, sondern auf Basis der vorhandenen Servicebeschreibung eine

semiautomatische Subsumtion durchzuführen.

Um dieses Ziel zu erreichen muss eine Servicebeschreibung vorliegen, aus der sich alle benötigten Informationen ableiten lassen. Dazu eigenen sich semantische Servicebeschreibungen, die auf Service Ontologien basieren und somit bereits einen hohen Grad an Formalisierung besitzen. Ein Beispiel für eine solche Ontologie ist die im Rahmen des TEXO Projekts entwickelte Service Ontology[OBB+09], die als Grundlage für die Fortführung dieser Arbeit dient. Wenn diese Ontologie um juristische Begriffe erweitert wird kann die bereits bestehende Einordnung der Servicebeschreibungselemente verwendet werden, um diese Begriffe den Definitionen in der Datenschutzontologie zuzuordnen.

References

[Bau08] Christian Baumann. Contracting and Copyright Issues for Composite Semantic Services. In *ISWC '08: Proceedings of the 7th International Conference on The Semantic Web*, pages 895–900, Berlin, 2008. Springer.

[BCBG05] V. RIchard Benjamins, Pompeu Casanovas, Joost Breuker, and Aldo Gangemi, editors. *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*. Springer-Verlag, Heidelberg, 2005.

[HAW08] Lutz Heuser, Claudia Alsdorf, and Dan Woods. *International Research Forum 2007*. Evolved Technologists Press, New York, 2008.

[JRS08] Christian Janiesch, Rainer Ruggaber, and York Sure. Eine Infrastruktur für das Internet der Dienste. *HMD-Praxis der Wirtschaftsinformatik*, 45(261):71–79, 2008.

[Lar83] Karl Larenz. *Methodenlehre der Rechtswissenschaft*. Springer, Berlin, 5. edition, 1983.

[OBB+09] Daniel Oberle, Nadeem Bhatti, Saartje Brockmans, Michael Niemann, and Christian Janiesch. Countering Service Information Challenges in the Internet of Services. *Business & Information Systems Engineering*, Volume 1, Number 5 / Oktober 2009:370–390, 2009.

[TEG04] Marie-Theres Tinnefeld, Eugen Ehmann, and Rainer W. Gerling. *Einführung in das Datenschutzrecht*. Oldenbourg Wissensch.Vlg, 4 edition, 2004.