Auf dem Weg zum Begriff

Vom Rechtswort zur Rechtsontologie: Automatisierte Verfahren zur semantischen Erschließung von Texten

Doris Liebwald
d@liebwald.com

Abstract: *Die Komplexität und Begrifflichkeit des Rechts stellen besondere Herausforderungen an die Automatisierung und Wissensrepräsentation im Recht. Dieser Beitrag gibt einen Überblick über die Landschaft der Verfahren zur semantischen Erschließung von Rechtstexten und behandelt die sich hierbei aus den Eigenheiten des Rechts und der Rechtssprache ergebenden Problemstellungen. Zentrale Themen sind Dokumentstrukturen, Metadaten, Thesauri, Ontologien und das Textmining.*

1. Einleitung

Die Frage der besten und zweckmäßigsten Formalisierung des Rechts zur computergestützten Verarbeitung, sei es mit logischen, begrifflichen oder anderen insb. sprachbezogenen Formalisierungen, ist ein wesentlicher Forschungsgegenstand der Rechtsinformatik. Zentrales Thema ist die Entwicklung einer maschinenverarbeitbaren Semantik zur Beschreibung der normativen und der realen Welt in maschinenausführbarer Sprache. Genau betrachtet handelt es sich bei diesen Arbeiten um maschinenverarbeitbare Spezifikationen der Semantik, die Maschine kann folglich nicht dynamisch und zur Laufzeit die Bedeutung von Rechtstexten tatsächlich "verstehen", vielmehr sind die Bedeutungen von Begriffen und deren Verwendung wie auch Lernprozesse vom Menschen vorgegeben und die Maschine kann nur jenes Wissen verarbeiten, dass ihr explizit gemacht wurde.

Die verfolgten Ziele können grob in zwei Kategorien aufgeteilt werden: einerseits die benutzerorientierte Aufbereitung von Rechtsinformation um Fachkreisen genauso wie Rechtsunterworfenen besseren Zugang zu der wachsenden Menge an juristischen Texten zu bieten, andererseits die maschinenausführbare Repräsentation von Rechtsnormen zur Schaffung von Systemen, die Rechtsregeln unmittelbar anwenden oder den Menschen in der Anwendung solcher Regeln unterstützen. Hinzu treten Aspekte der Interoperabilität und der Automatisierung von Arbeitsprozessen.

Ontologien als formale Wissensmodelle zur Beschreibung der Bedeutung von Informationen und deren Kontext werden gegenwärtig als Schlüsselinstrument zur expliziten Beschreibung von Konzepten der Domäne Recht betrachtet und stehen im Mittelpunkt der Forschung. Gruninger und Lee unterscheiden drei Hauptanwendungsgebiete für Ontologien: Organisation und Wiederverwendung von Wissen, automatisches Schließen sowie Kommunikation unter und zwischen Menschen und Anwendungen,[1] womit Ontologien auch für die Rechtsinformatik als Modelle der Wissensrepräsentation, für die Metabeschreibung des Rechtssystems und auch für die Automatisierung von juristischen Entscheidungen von höchstem Interesse sind. Valente sieht vier potentielle Anwendungsbereiche für Rechtsontologien: die Organisation und Strukturierung von Information, Schlussfolgerung und Problemlösung, semantische Indexierung und Suche sowie

[1] Gruninger/Lee, *Ontology Applications and Design, 2002 (40).*

die semantische Integration bzw. Interoperation.[2]

Dieser Beitrag konzentriert sich auf automatisierte Verfahren zur semantischen Repräsentation und zum semantischen Retrieval textueller Rechtsinformation, die dem Nutzer das Auffinden der für ihn relevanten Rechtstexte und Informationen erleichtern sollen. Im Zentrum der Betrachtungen stehen Rechtsontologien und verwandte Modelle der Wissensrepräsentation, es wird jedoch auch auf linguistische und statistische Methoden eingegangen, die neben ihren selbständigen Anliegen sowohl die Erstellung von Ontologien als auch die Verbindung zwischen Ontologie und Textkorpus wesentlich unterstützen können.

1.1. Warum semantische Repräsentation?

Die syntaktische Repräsentation vermag im juristischen Information Retrieval nur in beschränkem Maße semantische und pragmatische Bedeutungen und somit semantische Beziehungen zwischen Informationsbedarf und Dokumentinhalt herzustellen.[3] Die abstrakte Rechtsfrage, die juristischen Konzepte und Problemstellungen müssen in eine technische Suchanfrage übersetzt werden, die Suchterme werden mit dem Dokumentenindex abgeglichen. Hinzu kommt die mangelnde Behandlung linguistischer Problemstellungen, die sich in der deutschen Sprache in besonderem Maße stellen; Veränderungen der Wortform durch Konjugation, Deklination etc. bereiten dem Informationssuchenden nach wie vor Kopfzerbrechen. Heute übliche Systeme gehen selten über die Freitextsuche im Volltextindex, eine Suche in den grundlegenden Dokument-Metadaten wie Dokumenttyp, Autor, Ausgabedatum etc. und eine einfache semantische Beschreibung des Inhalts über grobe Klassifikationen, Verschlagwortung oder eine Normierung der verwendbaren Begriffe mittels eines Thesaurus hinaus.[4] Die Retrievalfunktionen sind jedoch theoretischen Beschränkungen unterworfen, die Matthijssen wie folgt isoliert: (1) der Index kann den Informationsgehalt eines Dokuments nur teilweise beschreiben, (2) bei Formulierung der Suchanfrage kann der Informationsbedarf nur unvollständig beschrieben werden, (3) die Matching-Funktion ist eine grob heuristische und beschränkt auf ein enges System von Vermutungen, (4) es existiert eine konzeptionellen Lücke, d.h. eine Diskrepanz zwischen subjektiver Sichtweise des Informationssuchenden bezüglich des Informationsgehaltes von Dokumenten und der reduzierten formalen Sichtweise auf Dokumente, die das Information Retrieval System bietet.[5] In Anbetracht der Rechtsinformationsflut auf nationaler und auch europäischer Ebene ist eine rein syntaktische Suche nicht mehr adäquat, es bedarf einer semantischen Suche, die das juristische Strukturwissen und Begriffsdenken repräsentiert. Rechner benötigen jedoch umfassende Informationen um syntaktisch ungleiche Strukturen als sinngleich erkennen und Worten, Phrasen und Sätzen ihre sich aus dem Kontext ergebende Bedeutung zuordnen zu können.[6] Die Rechtssprache stellt an die letztendlich binär arbeitende Maschine besondere Herausforderungen. Wie Haft treffend formuliert

---

[2] Valente, *Types and Roles of Legal Ontologies, 2005.*

[3] *Zum Information Retrieval immer noch aktuell* Salton/McGill*, Information Retrieval, 1987. Grundlagenwerke im Bereich des juristischen Information Retrieval sind* Bing *(Hrsg.), Handbook of Legal Information Retrieval, 1984 und* Rose, *A Symbolic and Connectionist Approach To Legal Information Retrieval, 1994.*

[4] *Siehe* Liebwald*, Evaluierung juristischer Datenbanken, 2003.*

[5] Matthijssen, *Interfacing between Lawyers and Computers, 1999 (29).*

[6] *Eine umfassende und eindrucksvolle Abhandlung der Probleme der semantischen Repräsentation natürlichsprachlich gegebenen Wissens gibt* Helbig, *Die semantische Struktur natürlicher Sprache, 2001.*

sind Rechtsbegriffe *"nicht durch begriffliches Ja-Nein-Denken, sondern durch typologisches Mehr-oder-Minder Denken zu erfassen."*[7] Vage und offene Rechtsbegriffe, die Systematik und Komplexität der Begrifflichkeiten, das hohe Abstraktionsniveau der Rechtssprache, die verschiedenen Sprachstile und Dokumenttypen, unpersönlicher Stil sowie die Neigung der Juristen zu Substantivierungen, Komposita, Partizipien statt Nebensätzen und komplexen Sätzen erschweren die automatische semantische Analyse von Rechtstexten.[8]

Die Forschung im Bereich der künstlichen Intelligenz im Recht hat klar gezeigt, dass nicht die logische Subsumtion das Kernproblem der Formalisierung des Rechts darstellt, sondern die Interpretation und Anwendung von Rechtsbegriffen.[9] Unbestimmte Rechtsbegriffe, die dem Recht innewohnende Dynamik, systematischer Zusammenhang und syntaktische Mehrdeutigkeiten zeigen sich als äußerst problematisch. Recht basiert auf Text und Sprache, Sprache ist interpretationsbedürftig. Das Wort gewinnt erst durch die ihm beigemessene Semantik Inhalt. Semantik, der Bedeutungsgehalt, der bestimmten Zeichen, insb. Worten, Sätzen zugeschrieben wird, ist nicht in der Syntax enthalten, sie kommt von "außen", wird durch individuelle, mentale Modelle gebildet. Sie macht das Wort zum kontextabhängigen Begriff.[10] Wort bezieht sich somit auf die äußere Form, Begriff hingegen auf den Bedeutungsinhalt. Selbst wenn sich der Gesetzgeber um größtmögliche Exaktheit von Rechtstexten und Rechtsbegriffen bemühen wollte, kann, da dies die Unschärfe und Schranken der Sprache selbst nicht erlauben, absolute Genauigkeit nicht erreicht werden. Umgekehrt können unscharfe Rechtsbegriffe als Antwort auf den Mangel an Exaktheit der Realität betrachtet werden. Darüber hinaus ist es nicht ausreichend, alleine die Rechtsbegrifflichkeiten und das Rechtswissen darzustellen, auch das Weltwissen, in das das Rechtswissen eingebettet ist, muss zumindest in Teilen abgebildet werden. Eine Formalisierung kann somit, zumindest aus heutiger Sicht, nie eine vollständige sein, sie ist eine Annäherung, ein Modell mit variablem Abstraktionsniveau und bedeutet immer einen Informationsverlust.

2. Dokumentstrukturen und Metadaten

Viele Rechtsdokumente verfügen über eine ausgeprägte Struktur, die bei entsprechender Auszeichnung das Information Retrieval wesentlich unterstützen kann. Idealer Weise erfolgt eine solche Annotierung bereits im elektronischen Dokumenterzeugungsprozess bei gleichzeitiger Erfassung aller relevanten Dokument-Metadaten. Metadaten sind Daten, die ausgewählte Aspekte anderer Daten beschreiben, also Daten über Daten. Metadatenindizes oder Metadatenontologien können für den Menschen lesbar und die Maschine

---

[7] Haft, *Recht und Sprache, 1994 (281).*

[8] *Eine umfassende Abhandlung bietet* Bydlinski, *Juristische Methodenlehre und Rechtsbegriff*[2]*, 1991; siehe auch* Müller *et al., Rechtstext und Textarbeit, 1997.*

[9] *Z.B. die Umsetzung des Latent Damage Acts mittels Prädikatenlogik im Latent Damage Advisor,* Susskind/Capper, *Latent Damage Law: The Expert System, 1988 oder die Projekte TAXMAN I und II zu einem beratenden System zur Steueroptimierung bei Unternehmensumgründung*, McCarty, *Reflections on "Taxman", 1977.*

[10] *Gemäß ISO 1087 ist unter Begriff (concept) "a unit of thought constituted through abstraction on the basis of properties common to a set of objects" zu verstehen, wobei angemerkt wird: "concepts are not bound to particular languages; they are, however, influenced by the social or cultural background" (vgl. auch DIN 2342, welche den Begriff als "Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird" definiert).*

verarbeitbar gestaltet werden und bieten den Vorteil einer einfacheren Integration von heterogenen Datenquellen, z.B. auch von Bild- oder Tonmaterial. Bei der Informationssuche kann dann auf diese Dokumentstrukturen, z.B. den Dokumenttitel oder den Tenor einer Entscheidung, oder Metadaten wie Schlagworte, Autor, Dokumenttyp, Informationen zum Dokumenterstellungsprozess, Verweisungen etc. zurückgegriffen und die Suchanfrage somit besser konkretisiert und sprachlich kontrolliert werden. Liegen Dokumente im Wesentlichen unstrukturiert und unbeschrieben vor, ist die Dokumentenkollektion aber in sich ausreichend homogen, so kann der Einsatz von Textmining-Methoden zur automatischen oder semiautomatischen Generierung von Metadaten und zur Erkennung von Textmustern sinnvoll sein, anderenfalls bleibt nur die manuelle Erschließung oder die Arbeit auf unstrukturiertem Text.[11]

In den letzen Jahren haben sich für Rechtdokumente die Textauszeichnungssprache XML (Extensible Markup Language), eine Metasprache zur hierarchischen Strukturierung von Texten, und RDF (Resource Description Framework) oder XML Topic Maps zur Bereitstellung von Metadaten durchgesetzt.[12] Diese Methoden erlauben zudem Logiknotationen, automatische Verknüpfungen, die Darstellung auch komplexerer Beziehungen zwischen den einzelnen Informationsressourcen und vereinfachte Visualisierung. Die Wahl der Mittel zur Dokumentstrukturierung und Ablage der Metadaten, ob Feldstrukturen, XML oder andere Methoden, ist aus Sicht des Information Retrieval von untergeordneter Bedeutung, die Verwendung offener XML- oder RDF-Standards[13] bietet jedoch den Vorteil höherer Interoperabilität. Beginnend mit dem Saarbrücker Standard für Gerichtsentscheidungen[14] entstanden zahlreiche Initiativen zur Schaffung gemeinsamer XML-Strukturen für typische Rechtsdokumente wie Normen und Entscheidungen. Die primären Ziele dieser Bestrebungen sind der einfachere interne und externe Informationsaustausch, die bessere Erschließbarkeit homogen aufgebauter Dokumente und einheitlicher Metadatenstrukturen sowie die einheitliche Suche über verteilte Quellen.[15] Hervorzuheben sind MetaLex,[16] ein offenes Format, das einen generischen und erweiterbaren Rahmen für die XML- und RDF-Auszeichnung von Strukturen und Inhalten juristischer Dokumente bietet, das italienische NormeInRete Projekt,[17] das europäische Netzwerk LEXML,[18] das sich die Unterstützung des weltweiten automatisierten Austausch juristischer Informationen auf XML- und RDF-Basis zum Ziel gemacht hat, das neue ONE-LEX Projekt

---

[11] *Siehe z.B.* Francesconi/Peruginelli, *Searching and Retrieving Legal Literature through Automated Semantic Indexing, 2007.*

[12] *Für eine Einführung in diese Technologien siehe insb.* Gaevic *et al., Model Driven Architecture and Ontology Development, 2006. Einen guten Einblick geben auch* Eckstein/Eckstein*, XML und Datenmodellierung, 2003.*

[13] *Zu den Spezifikationen des W3C siehe* http://www.w3.org/XML/ *und* http://www.w3.org/RDF/.

[14] *Siehe* Ebenhoch/Gantner, *Der Saarbrücker Standard für Gerichtsentscheidungen, 2001.*

[15] *Für einen Überblick siehe den dreiteiligen Artikel von* Notholt, *Das Semantic Web: Schritte auf dem Weg zum juristischen Einsatz (Teil 1), Die Standards des Semantic Web (Teil 2), Die Zukunft des Semantic Web (Teil 3), 2005. Eine Übersicht über gegenwärtige Initiativen und laufende Projekte geben* Biagoli *et al. (Hrsg.), Proceedings of the V Legislative XML Workshop (Florenz 2006), 2007.*

[16] *MetaLex ist aus dem niederländischen E-Power Projekt hervorgegangen, siehe* http://www.metalex.eu/.

[17] *Siehe* http://www.normeinrete.it *und insb.* Biagioli, C.*, NIR Editor, A XML Specific Environment for Legislative Drafting, 2003.*

[18] *Siehe* http://www.lexml.de.

(ONtologies for European Laws in EXecutable format) des European University Institute Florenz[19] und das ebenfalls junge ESTRELLA Projekt (European project for Standardized Transparent Representations in order to Extend Legal Accessibility), das ein Legal Knowledge Interchange Format (LKIF) aufbauend auf XML-basierte Standards des Semantic Webs inklusive RDF und OWL entwickeln will.[20]

Sowohl Dokumentstrukturen als auch Metadaten bleiben allerdings primär dem Bereich der Wissensorganisation, also der Ordnung von Wissen und der Syntax verhaftet und sind nur beschränkt zur Darstellung von Semantik tauglich. Dies soll jedoch nicht das Potential und die Wichtigkeit dieser Methoden schmälern, eine gute Dokumentstruktur und ein kluges System für Metadaten können das notwendige Ausmaß der zu modellierenden maschinenverarbeitbaren Semantik wesentlich reduzieren und stellen in vielen Fällen ganz einfach die "machbarere" Alternative zur vollständigen Formalisierung dar.

3. Semantische Modelle zur Wissensrepräsentation[21]

Aufgrund der assoziativen Begriffsstrukturen, der mehrdimensionalen Systematik und der Abstraktheit ist dem Recht eine natürliche Taxonomie zur schlichten hierarchischen Klassifizierung nach logischer Zusammengehörigkeit von Begriffen nicht inhärent, verwendete Klassifikationsschemata wie z.B. der Index des Bundesrechts in Österreich[22] oder der Fundstellennachweis des geltenden Gemeinschaftsrechts[23] sind von außen herangetragen und bedürfen eines Lernprozesses. Verschlagwortungen sind für den Informationssuchenden intuitiver verwendbar, bedürfen aber konsequenter Vergabe der Schlagwörter und gehen für große Dokumentsammlungen zumeist entweder nicht ausreichend in die Tiefe oder werden für den Nutzer zu umfangreich und unübersichtlich. Wichtige und traditionelle Behelfe zur juristischen Informationsrecherche sind fachspezifische Indizes, Glossare, Lexika, Thesauri und insb. die bislang allerdings primär in Printform vorliegenden hoch entwickelten und allgemein akzeptierten Systeme und Kommentare. Auch einfache Inhaltsverzeichnisse und Abstracts bieten dem Juristen gute Hilfen zur raschen Inhaltsbewertung. Es besteht hier somit ein großer Pool an Wissen und Erfahrung, auf den moderne Methoden der Wissensrepräsentation zurückgreifen können.

3.1. Thesauri[24]

Ein Thesaurus ist eine geordnete Zusammenstellung von Begriffen einer Wissensdomäne, bietet ein kontrolliertes Vokabular mit eindeutigen Deskriptoren zur Indexierung und Wiederauffindung von Dokumenten und erlaubt die Erfassung von Äquivalenzrelationen (insb. zur Behandlung von Synonymen, Homonymen und Polysemen), (poly-)hierarchischen Relationen und Begriffsverwandtschaften

---

[19] *Siehe* Sartor, *The ONE-LEX project and the informational unification of the laws of Europe, 2005.*

[20] *Siehe* http://www.estrellaproject.org/lkif-core/ *(siehe auch Kapitel 4.1).*

[21] *Ausführlich* Schweighofer, *Rechtsinformatik und Wissensrepräsentation, 1999.*

[22] *Ein jährlich vom Bundeskanzleramt/Verfassungsdienst herausgegebenes systematisches Verzeichnis des geltenden österr. Bundesrechts.*

[23] *Zugänglich z.B. unter* http://eur-lex.europa.eu/de/repert/index.htm.

[24] *Einen Überblick gibt* Foskett, *Thesaurus, 1997; detaillierter* Lancaster, *Vocabulary Control for Information Retrieval, 1986.*

(Assoziationsrelationen).[25] Ein Thesaurus ist ein mächtiges Tool zur Indexierung von Dokumenten, zur Selektion indizierter Dokumente und zur Erschließung weiterer Informationen auf Grundlage der vom Thesaurus abgebildeten Relationen, er reicht jedoch nicht aus, um wirkliche Bedeutungszusammenhänge darzustellen und die im jeweiligen Kontext relevanten Informationen effizient zu selektieren und kann auch das auf ein String-Matching ausgelegte Volltext-Retrieval nicht ausreichend unterstützen.[26]

3.2. Semantische Netze und Topic Maps[27]

Semantische Netze sind formale Modelle von Begriffen und ihren Relationen und gehen insoweit über Taxonomie oder Thesaurus hinaus, als sie eine umfassendere Berücksichtigung von Kontext durch die Darstellung von beliebigen Zusammenhängen erlauben und eine spezielle Grammatik für die Typisierung und Verwendung von Begriffen und Beziehungen verwendet wird. Die Mehrdimensionalität semantischer Netze erlaubt dem Nutzer assoziative Herangehensweisen und kontextabhängige Sichten. Semantische Netze können z.B. mit Topic Maps[28] modelliert werden. Topic Maps dienen der Beschreibung von Wissensstrukturen und verknüpfen diese mit Informationsquellen. Sie bestehen aus einer Sammlung von Themen (*topics*), wobei *topics* beliebige Dinge, Personen, Gegenstände, Orte, Ereignisse etc. sein können, ihren Beziehungen untereinander (*associations*) und den sog. *occurrences,* welche die jeweiligen *topics* mit ihren Vorkommen z.B. in Textdokumenten verknüpfen. Hierbei kann ein *topic* pro *association* mit unterschiedlichen Rollen (*roles*) belegt sein, sog. *scopes* erlauben eine kontextbezogene Einschränkung des Gültigkeitsbereichs von *topic characteristics*, *associations* und *occurrences*. Topic Maps und verwandte Methoden eigenen sich daher sehr gut zur Darstellung und Navigation von juristischem Strukturwissen und insb. der im Recht so wichtigen Verweisungen, wie dies z.B. in Systemen oder Kommentaren abgebildet ist.

3.3. Ontologien[29]

Ontologien sind nun wie semantische Netze formale Modelle von Begriffen und ihren Relationen, stellen jedoch mächtigere Modellierungsmöglichkeiten zur Verfügung, erlauben durch Inferenzmöglichkeiten von Logik eine höhere Ausdruckstärke und bieten den Vorteil, Funktionen von Indizes, Glossaren, Taxonomien, Thesauri, semantischen Netzen und Topic Maps in sich vereinigen zu können. Der "moderne" Begriff der Ontologie wird jedoch unscharf verwendet,

---

[25] *Standardisiert in ISO 2788, Guidelines for the establishment and development of monolingual thesauri (1986), entspricht DIN 1463-1, Erstellung und Weiterentwicklung von Thesauri; Einsprachige Thesauri (1987/11).*

[26] *Siehe auch Liebwald, Interfacing between Different Legal Systems, 2008.*

[27] *Zur Einführung siehe* Sowa *(Hrsg.), Principles of Semantic Networks: Explorations in the Representation of Knowledge, 1991. Siehe auch* Turtle/Croft, *Inference Networks for Document Retrieval, 1990 und insb.* Levine *et al., Toward Connectionist Representation of Legal Knowledge, 1994.*

[28] *Standardisiert in ISO/IEC 13250, Topic Maps (2nd edition 2002); formuliert in XML durch die TopicMaps.Org Authoring Group, XML Topic Maps (XTM) 1.0, Topics Maps.Org Specification (2001).*

[29] *Als Standardwerk kann* Staab/Studer *(Hrsg.), Handbook on Ontologies, 2004 empfohlen werden; ausführlicher* Sharman *et al., Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems, 2007.*

wodurch sich in Anlehnung and die Terminologie der W3C Recommendation OWL die Unterbegriffe *heavyweight ontologies* für verstärkt axiomatisierte Ontologien und *lightweight ontologies* für weniger ausdruckskräftige Ontologien, wobei jener auch auf Thesauri oder Topic Maps Anwendung findet, herausgebildet haben.[30] Die Grenze dieses inflationären Gebrauchs scheint der Übergang von der Wissensrepräsentation zur Wissensorganisation zu bilden, der allerdings ebenfalls ein fließender ist.

Ontologien existieren entsprechend ihren Aufgabestellungen in hoher Typenvielfalt, die grob in Metadaten-Ontologien, allgemeine Ontologien zur Darstellung des Weltwissens, spezifische Domainontologien, methoden- und aufgabenorientierte Ontologien sowie repräsentative Ontologien, die nur den Rahmen, die Repräsentation definieren, unterteilt werden können. Hierarchisch betrachtet kann an oberster Stelle eine sog. *Upper* oder *Foundation Ontology*[31] stehen, eine generische Ontologie, die nicht domainspezifische sondern allgemeingültige Entitäten beschreibt, darunter eine *Core Ontology*, eine Basis- oder Kernontologie, die aus den minimalen Konzepten besteht, die zum Verständnis der weiteren Konzepte, nämlich die der darunter liegenden *Domain Ontologies*, notwendig sind. Die Methoden sind ebenso breit gefächert und reichen von psycholinguistisch inspirierten WordNet-Methoden[32] mit natürlichsprachlichen Begriffserklärungen, die ohne formale Sprache zur Definition der Semantik auskommen und zur Erstellung semantischer Lexika herangezogen werden, bis hin zu regelbasierten Systemen mit hohem Formalisierungsgrad wie Cyc[33], das sich Millionen von logischen Axiomen, Regeln und Aussagen bedient, die die Bedingungen der einzelnen Objekte und Klassen spezifizieren und der Inferenzmaschine auf Alltagswissen basierende logische Schlussfolgerungen erlauben. Dementsprechend umfangreich sind auch die Einsatzmöglichkeiten von Ontologien, die vom derzeitigen Hauptanwendungsgebiet des Dokumentenmanagements über Entscheidungsunterstützungssysteme bis hin zu Anwendungen der künstlichen Intelligenz reichen.

4. Ontologien im Recht[34]

Der Begriff der Ontologie ist aus der griechischen Philosophie entlehnt und geht auf die "Lehre vom Seienden" des Philosophen Parmenides von Elea (um 515-445 AC) zurück.[35] In der Informatik und in den Informationswissenschaften wird unter

---

[30] *W3C Recommendation, OWL Web Ontology Language (2004). OWL ist eine auf RDF beruhende Ontologie-Modellierungssprache. Diese Empfehlung verwendet zusätzlich den Begriff* full *für eine* heavyweight ontology *mit maximaler Ausdrucksstärke und syntaktischer Freiheit von RDF, allerdings ohne Gewähr der vollständigen Verarbeitbarkeit. Zu den W3C Spezifikationen siehe http://www.w3.org/2004/OWL/.*

[31] *Prominente* Upper Ontologies *sind der Dublin Core (http://dublincore.org/), Open Cyc (http://www.opencyc.org/), SUMO (http://www.ontologyportal.org/) und DOLCE (http://www.loa-cnr.it/DOLCE.html).*

[32] *Siehe http://wordnet.princeton.edu/ und insb.* Fellbaum *(Hrsg.), WordNet: An Electronic Lexical Database, 1998.*

[33] *Siehe hierzu die ausführlichen Informationen auf den Pages der Cycorp, Inc. auf http://www.cyc.com/ (und dort insb. die Publikationsliste unter http://www.cyc.com/cyc/technology/pubs).*

[34] *Einen Überblick über gegenwärtige Anwendungen und Entwicklungen geben insb.* Benjamins *et al. (Hrsg.), Law and the Semantic Web, 2005 und* Casanovas, *Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007).*

[35] Ontos*, griech. "das Seiende";* logos*, griech. "Wort".*

einer Ontologie, der gebräuchlichen Definition Grubers folgend, eine explizite formale Spezifikation einer Konzeptualisierung verstanden.[36] Mittels Ontologien kann Semantik präzisiert werden, sie setzen allerdings die Einigung auf eine standardisierte Terminologie, Beziehungen und Regeln voraus, die von allen Akteuren, Menschen wie Maschinen, geteilt werden muss. Der Wert einer Ontologie steigt und fällt mit ihrer Anerkennung in der Fachwelt. Dem entspricht die Definition einer Ontologie von Uschold und Gruninger als "*an shared understanding of some domain of interest.*"[37] Auch Gruber präzisiert: *"A specification of a representational vocabulary for a shared domain of discourse – definitions of classes, relations, functions, and other objects – is called an ontology."* Für den erfolgreichen Einsatz einer Ontologie müssen die Begrifflichkeiten und die Repräsentation des Strukturwissens den Erfordernissen des jeweiligen Anwendungsgebietes und der jeweiligen Anwendergruppen entsprechen. Und auch oder gerade im juristischen Bereich besteht eine Vielfalt von semantischen Räumen, womit bei realistischer Betrachtung ein Konsens nur entweder auf einer abstrakteren Ebene oder für sehr spezifische Anwendungen stattfinden kann. Nichtsdestotrotz wird jede Ontologie unvermeidlich zu einem gewissen Maße die subjektive Betrachtungsweise ihres oder ihrer eigentlichen Ersteller repräsentieren.

Letztendlich sind aber auch Ontologien formale Modelle, die abstrahieren und die Realität vereinfachen. Sie sind nicht in der Lage, Kontext vollständig abzubilden, implizites Wissen, das berücksichtigt werden soll, muss explizit gemacht werden. Mit Blick auf die Seinslehre existiert für die Maschine nur das, was repräsentiert werden kann. Ontologien beschränken die Anzahl der Interpretationsmöglichkeiten von Begriffen und ihren Relationen und reduzieren Wissen auf den größten gemeinsamen Nenner, ermöglichen aber gerade dadurch den präziseren Wissensaustausch zwischen den Akteuren, die Menschen oder Maschinen sein können. Daher besteht auch eine gewisse Gefahr, bei der Formalisierung bereits Interpretationen und Wertungen vorwegzunehmen, die eigentlich dem Benutzer, dem Juristen überlassen sein sollten. Eine zu starke Festlegung, Standardisierung, Terminologienormierung kann sich schließlich auch als eine Art semantische Fessel[38] entpuppen und die Weiterentwicklung und Vielfältigkeit von Wissen und Sprache beinträchtigen.

Zur Ontologieerststellung bieten sich prinzipiell drei Verfahren an, automatisierte statistische Verfahren, deren Aufwand am geringsten ist, die aber eine gewisse Unschärfe bergen, linguistische Verfahren, die insb. in Kombination mit anderen Verfahren die im Information Retrieval leidlich bekannten Probleme wie allgemeinsprachliche Synonymie, Veränderung der Wortform durch Konjugation, Deklination abfangen können, und manuelle Verfahren, die – entsprechende Motivation, Konsequenz und Genauigkeit vorausgesetzt – die besten Ergebnisse liefern.[39] Bei breiten Ontologien wird zumeist auf eine semiautomatische Produktion gesetzt, wobei potentiell relevante Begriffe automatisch extrahiert, manuell integriert und durch einfache linguistische Tools ergänzt werden, grundsätzlich ist die Wahl der Methoden und Verfahren aber von den gegebenen Voraussetzungen und den konkreten Anforderungsspezifikationen abhängig. Ausgangspunkt können z.B. auch einschlägige elektronische Thesauri oder Lexika sein, wobei jedoch evtl.

---

[36] *Im Original: "*An ontology is an explicit specification of a conceptualization.*" Siehe* Gruber, *A Translation Approach to Portable Ontology Specifications, 1993 (199).*

[37] Uschold/Gruninger, *Ontologies: Principles, Methods and Applications, 1996.*

[38] *Der Begriff der semantischen Fessel ist entlehnt von* Schieffer/Muñoz, *Vom Change Management zum Change Meaning, 2003.*

[39] *Zu Verfahren und Methoden siehe insb.* Gómez-Pérez *et al., Ontology Engineering, 2004 und* Uschold/Gruniger*, Ontologies: Principles, Methods and Applications, 1996. Siehe auch die Literaturhinweise in FN 28.*

unterschiedliche Zielsetzungen der importierten Quellen und die daraus für die eigene Anwendung resultierenden Konzeptualisierungsdefizite nicht übersehen werden dürfen.[40] Äußerst schwierig gestaltet sich die Formalisierung von implizitem Wissen und macht eine enge Zusammenarbeit von Experten und Anwendern des zu formalisierenden Domainwissens, Informatikern und Softwareentwicklern notwendig. Anwendungsorientierte, spezifische Domainontologien sind heute machbar und im Einsatz, Probleme stellen jedoch nach wie vor die spätere Erweiterung und Anpassung, die Verknüpfung verschiedener Domainmodelle sowie die Vernetzung der Begriffswelten des Weltwissens (Weltmodell) und der Domainmodelle.

Bereits Aufbau und Pflege eines Thesaurus können sehr aufwendig sein, umso arbeits- und kostenintensiver ist die Erstellung und Wartung einer Ontologie mit höherem Formalisierungsgrad. Ontologien verlangen Expertenwissen, konsequentes Arbeiten und Konsistenz, die Qualität der Konzeptualisierungen ist letztlich für den Wert des Endproduktes ausschlaggebend. Die Wiederverwendung und die gemeinsame Nutzung von Ontologien werden zwar schon zwecks Reduktion der Kosten propagiert, sind praktisch jedoch schwierig. Kompatible Ontologien verlangen hohe Qualität, gemeinsames Design und interoberable Technologie, selbst dann kann es schwierig sein, die der übernommenen Ontologie zugrundeliegenden Prämissen nachzuvollziehen und die Ontologie mit den spezifischen eigenen Anforderungen in Einklang zu bringen. Später möglicher Weise notwendig werdende Erweiterungen, Aktualisierungen und Anpassungsmöglichkeiten müssen, soweit möglich, bereits in der Entwicklungsphase berücksichtigt werden; insb. bei umfänglicheren Ontologien können nachträgliche Validierungen, Umstrukturierungen oder umfassendere nichtmonotone Änderungen nahezu unmöglich werden. Dieser Aspekt ist für Rechtsontologien besonders relevant: das Recht ist dynamisch und besteht aus verschiedenen, veränderlichen semantischen Räumen, Rechtsontologien müssen daher, um nicht durch die Entwicklung des Rechts und der Rechtssprache obsolet zu werden, entsprechend flexibel, dynamisch oder abstrakt gestaltet sein und eine adäquate Darstellung der spezifischen juristischen Zeitschichten und dynamischen Prozesse erlauben.

4.1. Beispiele für Rechtsontologien

Rechtsontologien haben bereits lange Tradition,[41] in der Rechtsinformatik gelangten allerdings erst mit entsprechendem Fortschritt der technologischen Entwicklung um die 1990er McCarty mit seiner (formalen) Language for Legal Discourse LLD[42] und Stamper mit NORMA,[43] einer im theoretischen bleibenden Logik für Normen und Affordanzen, zu erster Prominenz. Zentrale Bedeutung erlangten schließlich die Frame Based Ontologie FBO, eine repräsentative Ontologie entwickelt von van Kralingen und Visser,[44] sowie die Functional Ontology of Law FOLaw von Valente,[45] beide mit eher epistemischen Ansatz. Die FBO ist als allgemeine und wiederverwertbare juristische Ontologie konzipiert und bietet drei

---

[40] *Siehe auch* Hirst*, Ontology and the Lexicon, 2004.*

[41] *Siehe z.B.* Hohfeld, *Fundamental Legal Conceptions as Applied in Judicial Reasoning, 1917; aber auch Hart, Kelsen etc.*

[42] *Siehe* McCarty*, A Language for Legal Discourse, 1989. Siehe auch den Rückgriff auf die LLD in* McCarty, *Deep Semantic Interpretations of Legal Texts, 2007.*

[43] *Siehe* Stamper*, The Role of Semantics in Legal Expert Systems and Legal Reasoning, 1991.*

[44] *Siehe insb.* Kralingen*, Frame-based Conceptual Models of Statute Law, 1995 und* Visser, *Knowledge Specification for Multiple Legal Tasks, 1995.*

[45] Valente, *Legal knowledge engineering: A modelling approach, 1995.*

Klassen (Norm, Aktion, Begriff) von Modellierungsprimitiven, wobei für jede Einheit eine Framestruktur mit allen relevanten Attributen definiert ist und die normspezifische Ontologie für jede Sub-Domäne neu angelegt werden muss. Die Core-Ontologie FOLaw hingegen enthält sechs Grundkategorien des Rechtswissens: normatives Wissen, Weltwissen, Haftungswissen, Sanktionswissen, Rechtschöpfungswissen und Metawissen. FOLaw wurde in mehreren Projekten insb. zur juristischen Falllösung (ON-LINE, CLIME/MILE, PROSA) eingesetzt, als Kernproblem erwies sich hierbei die Modellierung des Weltwissens. Die aus FOLaw gewonnenen Erkenntnisse führten schließlich zur Entwicklung des etwa 200 Begriffe umfassenden LRI-Core, einer Legal Core Ontology, die ursprünglich im Rahmen des Projektes E-Court zur Unterstützung eines flexiblen, multilingualen Information Retrievals über heterogene Quellen (Audio, Video, Text) im Bereich Strafprozess dienen sollte, jedoch auch in den Projekten E-Power und DIRECT eingesetzt wurde.[46] Der LRI Core ist auch Basis der derzeit im Rahmen des ESTRELLA-Projekts (Standardized Transparent Representations in order to Extend Legal Accessibility) entstehenden LKIF Core Ontology. Wie bei der LRI Core und der FOLaw handelt es sich hier um eine sog. Kern- oder Basisontologie, die nur die grundlegenden, allen Rechtsdomänen gemeinsamen Konzepte, Beziehungen und Eigenschaften repräsentiert und primär der Organisation und Indexierung von Bibliotheken von darunter liegenden Domainontologien und der Unterstützung der Wissensakquisition für die Konstruktion neuer Ontologien dient. Ziel des Estrella Projektes ist die Entwicklung eines auf XML, RDF und OWL basierenden Legal Knowledge Interchange Format LKIF, also eines maschinenlesbaren Formates zum Austausch von Rechtswissen zwischen unterschiedlichen Programmen, und geeigneter Programmierschnittstellen (APIs) zur Interaktion mit juristischen wissensbasierten Systemen.[47]

Einen anderen Ansatz wählten das italienische JurWordNet JWN,[48] ein semantisches, ontologiebasiertes juristisches Lexikon, und das darauf aufbauende multilinguale Lexikon LOIS (Lexical Ontologies for legal Information Sharing). Das JWN wurde in Erweiterung der italienischen EuroWordNet-Initiative EWN[49] entwickelt und besteht aus etwa 1700 lexikalischen Begriffen. JWN soll das Information Retrieval unterstützen und als Ressource an Metadaten für die automatische Informationsextraktion, Klassifikation und semantische Auszeichnung von Rechtsdokumenten dienen. Für die Erstellung des JWN wurden die Suchanfragen an die Rechtsinformationssysteme Progetto N.I.R. und ITALGIURE, insb. die UND und ODER Verknüpfungen, zur Identifizierung relevanter Wörter ausgewertet und um Begriffe und Definitionen aus einschlägigen Handbüchern, Wörterbüchern, Lexika und Enzyklopädien erweitert. Über dem JWN liegt die Descriptive Ontology for Linguistic and Cognitive Engineering DOLCE[50] und die Core Legal Ontology CLO.[51]

---

[46] *Einen Überblick zu FOLaw und LRI Core mit weiteren Verweisen geben* Breuker, *et al., Use and Reuse of Legal Ontologies in Knowledge Engineering and Information Management, 2005.*

[47] *Der LKIF Core ist auf der ESTRELLA Homepage unter http://www.estrellaproject.org/lkif-core/ hervorragend dokumentiert. Siehe auch* Hoekstra *et al., The LKIF Core Ontology of Basic Legal Concepts, 2007.*

[48] *Siehe insb.* Gangemi *et al., Jur-Wordnet, a Source of Metadata for Content Description in Legal Information, 2003 und* Sagri/Tiscornia*, Semantic Lexicons for Accessing Legal Information, 2004.*

[49] *EuroWordNet ist ein multilinguales WordNet in neun Sprachen und beruht wiederum auf dem Princeton WordNet (siehe FN 31). Das 1999 abgeschlossene Projekt ist auf http://www.illc.uva.nl/EuroWordNet/ ausführlich dokumentiert.*

[50] *DOLCE selbst wurde im WonderWeb Projekt entwickelt und ist auf http://www.loa-cnr.it/DOLCE.html gut dokumentiert. Siehe auch* Gangemi et al.*, Sweetening Ontologies with*

Das LOIS WordNet[52] entspricht dem Grundgedanken des JWN, soll aber das multilinguale Information Retrieval von Rechtsinformationen über Ländergrenzen hinweg unterstützen und umfasst dazu für jede der sechs beteiligten Sprachen etwa 5000 Begriffe und natürlichsprachliche Definitionen, die über Äquivalenzrelationen mit einem auf dem JurWordNet beruhenden interlingualen Index miteinander verbunden sind. Die Begriffe innerhalb einer Sprache sind über lexikalische (insb. Synonymie/Antonymie) and taxonomische (insb. Hyponymie/Hyperonymie) Relationen miteinander verbunden. Die LOIS-Datenbank enthält sowohl lexikalische Begriffe, die auf dem JWN beruhen und mit verschiedenen Methoden erweitert und ergänzt wurden, als auch aus EU Richtlinien extrahierte Legaldefinitionen. LOIS kämpft daher mit besonderen Herausforderungen, denn Rechtsontologien sind nicht unabhängig von Sprache und Rechtssystem. Rechtsbegriffe können in den verschiedenen Rechtssystemen sehr unterschiedlich interpretiert werden und in der jeweiligen Begriffsstruktur an völlig anderen Plätzen auftreten. Hinzu kommt, dass für ein Volltext-Retrieval Wortvarianten und Ausdrucksvarianten in allen beteiligten Sprachen ausreichend abgedeckt werden müssten.[53]

4. Automatische Textanalyse

Eine Ontologie alleine löst nun noch kein Rechtsproblem. Soll sie der Erschließung von Textdokumenten dienen, etwa durch intelligente Navigation[54] oder semantische Interpretation von Suchanfrage und Texten, so muss immer noch die Lücke zwischen Konzeptualisierung der Information und gespeicherter Form überwunden werden. Die Ontologie muss mit der Textkollektion derart in Übereinstimmung gebracht werden, dass die Maschine den Bezug zwischen Ontologie und Wissensbestandteilen herstellen kann. Dies kann manuell oder semiautomatisch bereits bei Erstellung der Ontologie bzw. der Textdokumente durch Bildung entsprechender Relationen erfolgen. Gängig weil leichter praktikabel ist die Modellierung über Metadaten. Bei großen und wenig strukturiert vorliegenden, rasch wachsenden oder inhomogenen Textkorpora sind automatische Verfahren gefragt, wobei hier auf Erfahrungen und Techniken aus dem Bereich der automatischen Textanalyse, insb. der automatischen Klassifikation, der Informationsextraktion und dem Document Clustering, zurückgegriffen werden kann. Generell können statistische Techniken zur Konsistenz der Begriffe der Ontologie und jener der Textkollektion beitragen. Hierzu bietet sich die Kombination verschiedener linguistischer und statistischer Techniken in einem sog. Textmining-Verfahren an. Durch Textmining können oder sollen durch maschinelle Analyse relevante

---

DOLCE, 2002 und Gangemi *et al., Sweetening WORDNET with DOLCE, 2003.*

[51] Gangemi *et al., A Constructive Framework for Legal Ontologies, 2005 und* Gangemi, A.*, Design Patterns for Legal Ontology Construction, 2007.*

[52] *Siehe insb.* Peters *et al., The Structuring of Legal Knowledge in LOIS, 2007;* Schweighofer/Liebwald*, LOIS: Juristische Ontologien und Thesauri, 2005 und* Dini, L. *et al., Cross-lingual Legal Information Retrieval Using a WordNet Architecture, 2005.*

[53] *Siehe zu dieser Problematik* Liebwald, *Semantic Spaces and Multilingualism in the Law: The Challenge of Legal Knowledge Management, 2007 und* Müller/Burr *(Hrsg.), Rechtssprache Europas. Reflexion der Praxis von Sprache und Mehrsprachigkeit im supranationalen Recht, 2005 und Liebwald, Interfacing between Different Legal Systems, 2008.*

[54] *Siehe z.B.* Zhang/Koppaka, *Semantics-Based Legal Citation Network, 2007 oder* Schweighofer/Liebwald, *Advanced Lexical Ontologies and Hybrid Knowledge Based Systems: First Steps to a Dynamic Legal Electronic Commentary, 2007.*

Informationen aus Textdaten gewonnen, "herausgeschürft" werden.[55]

An erster Stelle steht im Textmining eine linguistische Analyse der Textdaten, um automatisch oder semiautomatisch repräsentative Merkmale zu extrahieren und eine Datenstruktur zu konstruieren. Hierbei können morphosyntaktische Lexika zur Bestimmung von Wortformen und Wortklassen, semantische Lexika mit Regeln zur Disambiguierung und Verarbeitung mehrwortiger Begriffe, Listen zur Erkennung besonderer Elemente wie Eigennamen oder Abkürzungen und schließlich Fachlexika zur Identifizierung von spezifischen Bedeutungen im Kontext eines Fachgebiets und spezifischen semantischen Relationen zwischen Fachbegriffen eingesetzt werden. Darauf setzen die eigentlichen Textmining-Verfahren auf. Diese können z.B. Termhäufigkeiten berechnen und gewichten und das gemeinsame Auftreten von Termen auswerten, die Wahrscheinlichkeit des Vorkommens von Bedeutungsvarianten im Kontext berechnen, sich wiederholende Textstrukturen erkennen, Texte oder Textelemente aufgrund ihrer neuralen Muster einer Klassifikation zuordnen oder Cluster ähnlicher Dokumente bilden. Somit können auch mit Textmining-Verfahren implizite Information und multidimensionale Zusammenhänge zwischen diesen Informationen explizit gemacht werden, wobei allerdings zumindest bei Rechtstexten ein Trainingsprozess auf den Referenzdokumenten notwendig ist.

Zur Anwendung des Textmining auf juristische Dokumente müssen jedoch die Algorithmen herkömmlicher Textmining-Software angepasst werden. Rechtstexte folgen nicht den Regeln, die für eine allgemeine Informationsrecherche im Internet oder für journalistische Texte gelten,[56] Rechtstexte sind komplexer, bedienen sich einer abstrakten Sprache und verschachtelter Sätze. Relevante Schlüsselbegriffe müssen nicht im Titel oder in den ersten Sätzen stecken und werden üblicher Weise auch nicht regelmäßig wiederholt, einer Begründung kann eine das Ergebnis des Textmining wesentlich beeinflussende umfangreiche Abwägung der verschiedenen Argrumente vorausgehen und selbst geringfügige Abweichungen in der Syntax können mitunter mit einem groben Bedeutungswandel einhergehen. Textmining-Verfahren sind, wie dies z.B. die Projekte HYPO[57] und CATO[58] zeigen, bei der Berechnung von Sachverhaltsähnlichkeiten erfolgreicher als bei der Isolierung und Berechnung der Ähnlichkeit der konkreten Rechtsfrage.[59] Bei entsprechender Anpassung an die Erfordernisse juristischer Informationen können jedoch gute Ergebnisse erzielt werden. Hervorzuheben sind in diesem Bereich vor allem die Projekte KONTERM[60], FLEXICON (Fast Legal Expert Information CONsultant)[61],

---

[55] *Zu den Verfahren siehe z.B.* Ferber, *Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web, 2003 oder* Han/Kamber, *Data Mining: Concepts and Techniques, 2001. Für die Anwendung auch auf Rechtstexte siehe insb.* Moens, *Automatic Indexing and Abstracting of Documents Texts, 2000.*

[56] *Siehe hierzu* Moens *et al., Abstracting of Legal Cases: The SALOMON Experience, 1997 und* Moens/Dumortier*, Automatic Abstracting of Magazine Articles: The Creation of 'Highlight' Abstracts, 1998.*

[57] *Eine detaillierte Beschreibung des Computerprogrammes HYPO gibt* Ashley*, Modeling Legal Arguments: Reasoning with Cases and Hypotheticals, 1991.*

[58] *CATO übernimmt und erweitert das dem Programm HYPO zugrundeliegende Modell der fallbeispielbasierten Argumentation. Siehe insb.* Aleven/Ashley*, How Different Is Different? Arguing about the Significance of Similarities and Differences, 1996.*

[59] *Einen Überblick zum automatisierten Fallvergleich gibt* Ashley, *Case-Based Reasoning, 2006.*

[60] *Siehe* Schweighofer/Winiwarter*, Legal Expert System KONTERM, 1993 und* Schweighofer, *Rechtsinformatik und Wissensrepräsentation, 1999.*

[61] Smith *et al., Artificial Intelligence and Legal Discourse: The Flexlaw Legal Text Management System, 1995.*

SALOMON[62] und SMILE (Smart Index LEarner).[63]

Das Projekt KONTERM III soll hier als Anschauungsbeispiel dienen. KONTERM verwendete ein TFxIDF Vektorraummodell[64] zur Repräsentation der Dokumente und sollte durch die automatische Extraktion von Attributwerten den Textkorpus strukturieren, klassifizieren und beschreiben. Das System bestand aus vier Modulen, der selbstorganisierenden Karte SOM (Self-Organizing Map), der LabelSOM zur Beschreibung bzw. zum "Labeln" der Gemeinsamkeiten eines Clusters, der selbstorganisierenden GHSOM (Growing Hierarchical SOM) zur automatischen hierarchischen Organisation und Repräsentation sowie einem Data Enrichment Tool zur Adaptierung der Vektoren. Trainiert wurde auf verschiedenen Textkorpora mit jeweils einigen hundert Dokumenten aus dem Internationalen Recht und dem Europarecht, wobei hier neben deutschen auch englische und französische Textkorpora herangezogen wurden. Die linguistische Behandlung beschränkte sich allerdings auf einen einfachen, für die englische Sprache entwickelten Stemmingparser. Ziel der Bemühungen war, den Besonderheiten der juristischen Sprache durch angepasste Vektorrepräsentation und Textsegmentierung gerecht zu werden. KONTERM erwies sich als zufriedenstellend bei der Klassifikation von Dokumenten, zeigte aber Defizite bei der Bildung der Labels, die den Inhalt der jeweiligen Cluster beschreiben sollten, wobei die Labels mitunter die Gemeinsamkeiten zwischen Dokumenten sehr gut beschrieben, hierbei jedoch nicht jene Begriffe verwendeten, die ein Jurist erwarten würde. Auch die Annahme, dass statistische Ähnlichkeiten zwischen gleichen Begriffen bzw. Begriffsklassen in verschiedenen Sprachen bestehen, zeigte sich zumindest unter den verwendeten Methoden nicht zielführend. Zusammenfassend konnte festgestellt werden, dass die Textanalyse bei höherer Ähnlichkeit der Texte der jeweiligen Dokumentsammlung wesentlich besser funktioniert und es zur eindeutigeren Beschreibung der Texte einer Segmentierung der Dokumente oder der Generierung von Merkmalsvektoren bedarf. Überraschende Ergebnisse erzielte jedoch das Data Enrichment Tool, das einen kleinen, manuell und auf Basis der Referenzdokumente erstellten Thesaurus integrierte und zu einer signifikanten Verbesserung der Qualität der Cluster und Label führte.[65]

5. Fazit

Für die automatische Erschließung von juristischen Texten stehen zahlreiche Methoden zur Verfügung, solche die sich mehr der Syntax und solche die sich mehr der Semantik verschrieben haben, solche die nur einen geringen Grad der Formalisierung und solche die eine hohe Formalisierung erlauben. Manche Verfahren ermöglichen weitgehende Automatisierung, andere bedürfen eines verstärkten ergänzenden manuellen und intellektuellen Aufwandes, eine volle Automatisierung ist aus derzeitiger Perspektive für Rechtstexte nicht realisierbar. Allen Methoden ist gemeinsam, dass sie der Maschine die Bedeutung von Texten

---

[62] Moens *et al., Abstracting of Legal Cases: The SALOMON Experience, 1997.*

[63] Brüninghaus/Ashley, *Improving the Representation of Legal Case Texts with Information Extraction Methods, 2001.*

[64] *In einem Vektorraummodell können Dokumente als Vektoren in einem n-dimensionalen Vektorraum dargestellt und insb. auf Basis von Termhäufigkeiten (Term Frequency/Inverse Document Frequency) oder gewichteten Häufigkeiten verglichen werden. Die Dimension des Vektorraums entspricht hierbei dem Vokabular der Textkollektion, die Dimension des Vektors eines Dokuments den in diesem Dokument enthaltenen repräsentativen Termen.*

[65] *Siehe* Schweighofer *et al., Improvement of Vector Representation of Legal Documents with Legal Ontologies, 2002.*

und Sinneszusammenhängen vermitteln wollen, wobei derzeit jene Modelle präferiert werden, die das assoziative Denken des Menschen nachahmen. Hierzu muss maschinengestaltete Information in maschineninterpretierbare Form gebracht werden, es bedarf aber noch eines weiteren wichtigen Schrittes, die Ausgabe muss letztendlich wieder der menschlichen Informationsverarbeitung zugänglich sein, also dem Menschen das Wissen brauchbar repräsentieren. Im Ergebnis muss das geschaffene System dem Menschen neues Wissen verschaffen bzw. zur Validierung seiner Hypothesen dienen können.

Welche Verfahren und Methoden erfolgreicher sind bzw. welcher Formalisierungsgrad angestrebt werden soll kann nicht generell festgestellt werden und hängt nicht nur davon ab, welches Ziel konkret erreicht werden soll, sondern auch von den Umgebungs- und insb. den Startbedingungen, etwa in welcher Form die Dokumente vorliegen, wie diese auch sprachlich gestaltet und strukturiert sind, ob die Textsammlung insgesamt homogen ist, ob auf bestehende Begriffslexika, Erfahrung und erprobte Software zurückgegriffen werden kann etc. Noch stellt sich in der Praxis jedoch in vielen Fällen weniger die Frage der technischen Machbarkeit, die, wie dieser Beitrag aufzeigt, bei entsprechender Akribie bereits sehr hoch ist, als die der wirtschaftlichen Gangbarkeit.

Literaturverzeichnis

Aleven/Ashley, K.D., *How Different Is Different? Arguing about the Significance of Similarities and Differences; in: Advances in Case-Based Reasoning: Proceedings of the EWCBR-96, Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, 1996, 1-15*

Ashley, K.D., *Modeling Legal Arguments: Reasoning with Cases and Hypotheticals; MIT Press, MA, 1991*

Ashley, K.D., *Case-Based Reasoning, in: Lodder, R./Oskamp, A. (Hrsg.), Information Technology & Lawyers, Advanced technology in the legal domain, from challenges to daily routine; Springer, Dodrecht, NL, 2006, 23-60*

Benjamins, V.R. *et al. (Hrsg.), Law and the Semantic Web, Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications; Springer, Berlin et al., 2005*

Biagioli, C. *et al., NIR Editor, A XML Specific Environment for Legislative Drafting; in: Proceedings of the JURIX 2003 Workshop on the Development of Standards for Describing Legal Documents, elektronische Publikation unter* http://www.lri.jur.uva.nl/standards2003/

Biagoli, C. *et al. (Hrsg.), Proceedings of the V Legislative XML Workshop (Florenz 2006), European Press Academic Publishing, Florenz, 2007*

Bing, J. *(Hrsg.), Handbook of Legal Information Retrieval, Elsevier, New York, 1984*

Breuker, J.A. *et al., Use and Reuse of Legal Ontologies in Knowledge Engineering and Information Management, in: Benjamins, V.R. et al. (Hrsg.), Law and the Semantic Web, 2005 (a.a.O.), 36-64*

Brüninghaus, S./Ashley, K.D., *Improving the Representation of Legal Case Texts with Information Extraction Methods, in: Proceedings of the 8th ICAIL 2001, ACM Press, New York, 42-51*

Bydlinski, F., *Juristische Methodenlehre und Rechtsbegriff*[2], *Springer, Wien 1991*

Casanovas, P. *et al. (Hrsg.), Proceedings of the 2nd Workshop on Legal Ontologies and Artificial Intelligence Techniques, Stanford University, CA, Juni 2007, elektronische Publikation unter* http://www.ittig.cnr.it/loait/LOAIT07-Proceedings.pdf

Dini, L. *et al., Cross-lingual Legal Information Retrieval Using a WordNet Architecture, in: Proceedings of the 10th ICAIL 2005, ACM Press, New York, 2005, 163-167*

Ebenhoch, P./Gantner, F., *Der Saarbrücker Standard für Gerichtsentscheidungen, JurPC 116/01*

Eckstein, R./Eckstein, S., *XML und Datenmodellierung, dpunkt.verlag, Heidelberg, 2003*

Fellbaum, C. *(Hrsg.), WordNet: An Electronic Lexical Database, MIT Press, MA 1998*

Ferber, R., *Information Retrieval: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web, dpunkt.verlag, Heidelberg, 2003*

Foskett, D., *Thesaurus, in: Sparck Jones, K./Willett, P. (Hrsg.), Readings in Information Retrieval, Morgan Kaufmann, San Francisco, CA, 1997, 111-134*

Francesconi, E./Peruginelli, G., *Searching and Retrieving Legal Literature through Automated Semantic Indexing, in: Proceedings of the 11th ICAIL 2007, ACM Press, New York, 2007, 131-139*

Gaevic, D. *et al., Model Driven Architecture and Ontology Development, Springer, Berlin, 2006*

Gangemi, A. *et al., A Constructive Framework for Legal Ontologies, in: Benjamins, V.R. et al. (Hrsg.), Law and the Semantic Web, 2005 (a.a.O.), 97-124*

Gangemi, A. *et al., Jur-Wordnet, a Source of Metadata for Content Description in Legal Information; in: Proceedings of the ICAIL Workshop on Ontologies, Edinburgh, UK, 2003, elektronische Publikation unter http://www.lri.jur.uva.nl/~winkels/legontICAIL2003.html*

Gangemi, A. et al., *Sweetening Ontologies with DOLCE, in: Proceedings of the 13th EKAW 2002, Lecture Notes in Computer Science Vol. 2473/2002, Springer, London, UK, 166-181*

Gangemi, A. *et al., Sweetening WORDNET with DOLCE, AI Magazine Archive Vol. 24/3 (2003), AAAI, Menlo Park, CA, 13-24*

Gangemi, A., *Design patterns for legal ontology construction, in: Casanovas, P. et al. (Hrsg.), LOAIT 2007 (a.a.O.), 66-85*

Gómez-Pérez, A. *et al., Ontology Engineering, Springer, Berlin, 2004*

Gruber, T.R., *A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition Vol. 5/2 (1993), Academic Press, London et al., 199-220*

Haft, F., *Recht und Sprache, in: Kaufmann, A./Hassemer, W. (Hrsg.), Einführung in die Rechtsphilosophie und Rechtstheorie der Gegenwart6, C.F. Müller, Heidelberg, 1994, 269-291*

Han, J./Kamber, M., *Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, CA, 2001*

Helbig, H., *Die semantische Struktur natürlicher Sprache. Wissensrepräsentation mit MultiNet, Springer, Berlin, 2001*

Hirst, G., *Ontology and the Lexicon, in: Staab, S./Studer, R. (Hrsg.), Handbook on Ontologies, Springer, Berlin/Heidelberg, 2004, 210-229*

Hoekstra, R. *et al., The LKIF Core Ontology of Basic Legal Concepts, in: Casanovas, P. et al. (Hrsg.), LOAIT 2007 (a.a.O.), 43-63*

Hohfeld, W.N., *Fundamental Legal Conceptions as Applied in Judicial Reasoning, The Yale Law Journal Vol. 26/8 (Juni 1917), 710-770*

Kralingen, R.W. van, *Frame-based Conceptual Models of Statute Law, Theses, University of Leiden, The Hague, NL, 1995*

Lancaster, F.W., *Vocabulary Control for Information Retrieval, Information Resources Press, Arlington, VA, 1986*

Levine, D.S. *et al., Toward Connectionist Representation of Legal Knowledge, in: Aparicio, M./Levine, D.S. (Hrsg.), Neural Networks for Knowledge Representation and Inference, Lawrence Erlbaum, Hillsdale, NJ, 1994, 269-282*

Liebwald, D., *Evaluierung juristischer Datenbanken, Verlag Österreich, Wien 2003*

Liebwald, D., *Interfacing between Different Legal Systems: Using the Examples of N-Lex and EUR-Lex, in: Grewendorf/Rathert (Hrsg.), Formal Linguistics and Law, Series for the volume Trends in Linguistics - Studies and Monographs (TiLSM), 2008 (im Druck)*

Liebwald, D., *Semantic Spaces and Multilingualism in the Law: The Challenge of Legal Knowledge Management, in: Casanovas, P. et al. (Hrsg.), LOAIT 2007 (a.a.O.), 131-148*

Matthijssen, L., *Interfacing between Lawyers and Computers, An Architecture for Knowledge-based Interfaces for Lawyers, Law and Electronic Commerce Vol. 8, Kluwer, The Hague, 1999*

McCarty, L.T., *A Language for Legal Discourse: I. Basic Features, in: Proceedings of the 2nd ICAIL 1989, ACM Press, New York, 1989, 180-189*

McCarty, L.T., *Deep Semantic Interpretations of Legal Texts, in: Proceedings of the 11th ICAIL 2007, ACM Press, New York, 2007, 217-224*

McCarty, L.T., *Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning, Harvard Law Review Vol. 90/5 (März 1977), Cambridge, MA, 837-893*

Moens, M.-F. *et al., Abstracting of Legal Cases: The SALOMON Experience, in: Proceedings of the 6th ICAIL 1997, ACM Press, New York, 1997, 114-122*

Moens, M.F., *Automatic Indexing and Abstracting of Documents Texts, The Kluwer International Series on Information Retrieval, Kluwer, Boston et al., 2000*

Moens, M.-F./Dumortier, J., *Automatic Abstracting of Magazine Articles: The Creation of 'Highlight' Abstracts, in: Proceedings of the 21st SIGIR 1998, ACM Press, New York, 1998, 359-360*

Müller, F. *et al., Rechtstext und Textarbeit, Duncker & Humblot, Berlin, 1997*

Müller, F./Burr, I. *(Hrsg.), Rechtssprache Europas. Reflexion der Praxis von Sprache und Mehrsprachigkeit im supranationalen Recht, Duncker & Humblot, Berlin, 2005*

Notholt, J., *Das Semantic Web: Schritte auf dem Weg zum juristischen Einsatz (Teil 1), JurPC 57/2005; Die Standards des Semantic Web (Teil 2), JurPC 65/2005; Die Zukunft des Semantic Web (Teil 3), JurPC 66/2005*

Peters, W. *et al., The Structuring of Legal Knowledge in LOIS, Artificial Intelligence and Law Journal, Vol. 15/2 (Juni 2007), Springer, NL, 117-135*

Rose, D.E., *A Symbolic and Connectionist Approach To Legal Information Retrieval, Lawrence Erlbaum, Hillsdale, NJ, 1994*

Sagri, M.-T./Tiscornia, D., *Semantic Lexicons for Accessing Legal Information, in: Proceedings of the 3rd EGOV 2004, Lecture Notes in Computer Science Vol. 3183/2004, Springer, Berlin/Heidelberg, 72-81*

Salton, G./McGill, M.J., *Information Retrieval - Grundlegendes für Informationswissenschaftler, McGraw-Hill, Hamburg, 1987*

Sartor, G., *The ONE-LEX project and the informational unification of the laws of Europe, in: Tagungsband der Klagenfurter Legistikgespräche 2005, Bildungsprotokolle Vol. 12, Kärntner Verwaltungsakademie, Klagenfurt, 2006, 193-202*

Schieffer, A./Muñoz, A.R., *Vom Change Management zum Change Meaning, Readers Pages, LO (Lernende Organisation) 13/2003*

Schweighofer, E. *et al., Improvement of Vector Representation of Legal Documents with Legal Ontologies, in: Proceedings of the 5th BIS 2002, Poznan University of Economics Press, Poznan, PL, 2002*

Schweighofer, E., *Rechtsinformatik und Wissensrepräsentation, Forschungen aus Staat und Recht 124, Springer, Wien/New York, 1999*

Schweighofer, E./Liebwald, D., *Advanced Lexical Ontologies and Hybrid Knowledge Based Systems: First Steps to a Dynamic Legal Electronic Commentary, AI & Law Journal Vol. 15/2 (2007), 103-115*

Schweighofer, E./Liebwald, D., *LOIS: Juristische Ontologien und Thesauri, in: Schweighofer, E. et al. (Hrsg.), Effizienz von e-Lösungen in Staat und Gesellschaft, Aktuelle Fragen zur Rechtsinformatik 2005, Boorberg, Stuttgart et al., 2005, 79-86*

Schweighofer, E./Winiwarter, W., *Legal Expert System KONTERM - Automatic Representation of Document Structure and Contents, in: Proceedings of the 4th DEXA 1993, Springer, Berlin, 1993, 486-497*

Sharman, R. *et al., Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems; Springer's Integrated Series in Information Systems Vol. 14, 2007*

Smith J.C. *et al., Artificial Intelligence and Legal Discourse: The Flexlaw Legal Text Management System, Artificial Intelligence and Law Vol. 3/1-2 (1995), 55-95, Kluwer, Dordrecht et al., NL*

Sowa, J.F. *(Hrsg.), Principles of Semantic Networks: Explorations in the Representation of Knowledge, Morgan Kaufmann, San Mateo, CA, 1991*

Staab, S./Studer, R. *(Hrsg.), Handbook on Ontologies, Springer, Heidelberg, 2004*

Stamper, R.K., *The Role of Semantics in Legal Expert Systems and Legal Reasoning, Ratio Juris, Vol. 4/2 (1991), Blackwell Publishing, Oxford, 219-244*

Susskind, R.E./Capper, P., *Latent Damage Law: The Expert System, Butterworths, London, 1988*

Turtle, H.F./Croft, W.B., *Inference Networks for Document Retrieval, in: Vidick, J.-L. (Hrsg.), Proceedings of the 13th SIGIR 1990, ACM Press, New York, 1-24*

Uschold, M./Gruninger, M., *Ontologies: Principles, Methods and Applications; Knowledge Engineering Review Vol. 11/2 (1996) 93-136, Cambridge University Press*

Valente, A., *Legal knowledge engineering: A modelling approach, IOS Press, Amsterdam, NL, 1995*

Valente, A., *Types and Roles of Legal Ontologies, in: Benjamins, V.R. et al. (Hrsg.), Law and the Semantic Web, 2005 (a.a.O), 65-76*

Visser, P.R.S., *Knowledge Specification for Multiple Legal Tasks: A Case Study of the Interaction Problem in the Legal Domain, Computer Law Series 17, Kluwer Law International, The Hague, NL, 1995*

Zhang, P./Koppaka, L., *Semantics-Based Legal Citation Network, in: Proceedings of the 11th ICAIL 2007, ACM Press, New York, 2007, 123-130*