

Die Sprache «SemaLogic» als semantische Repräsentation

Eine anforderungsbasierte Sprache zur Modellierung von Prüfungsordnungen und Abbildung von Studienverläufen

Markus von der Heyde¹ , Matthias Goebel²

Abstract: Eine wesentliche Aufgabe im Zuge der Digitalisierung der Verwaltungsprozesse einer Universität ist die Abbildung der Prüfungs- und Studienordnungen (PStO) im Campus-Management-System (CMS). Denn nur wenn die in diesen Ordnungen enthaltenen Regeln logisch auswertbar sind, können sie Abläufe im Student-Life-Cycle-Prozess steuern. Dieser Beitrag schlägt eine drastische Vereinfachung als technische Lösung vor. Auf Basis einer semantischen Repräsentation, die direkt aus den Ordnungen erstellt werden kann, wird die menschlich lesbare Form der Ordnung zum „Programmcode“ und kann direkt Einfluss auf die Umsetzung im CMS nehmen.

Für die Erzeugung einer solchen sprachlichen Übersetzung, welche die notwendige Flexibilität, aber eindeutige Repräsentation erzeugen kann, wurden allgemeine Anforderungen aus der Literatur sowie 20 Studienordnungen analysiert. Die getroffene Auswahl deckt dabei alle Bundesländer, den Großteil der Hochschulformen (Volluniversitäten, Technische Universität, Hochschulen für angewandte Wissenschaften, Musik- und Kunsthochschulen), ein breites Fächerspektrum und vier verbreitete Campus-Management-Systeme ab. Im Weiteren wird aufgezeigt, warum die auf dieser Grundlage entwickelte Sprache SemaLogic die dargelegten Anforderungen erfüllt und dass der Entwicklungsstand der darauf basierenden prototypischen Umsetzung bereits in der Lage ist, die semantische Repräsentation zu erzeugen und logische Prüfungen durchzuführen. Ein Ausblick auf die potentiellen Anwendungsfälle schließt den Beitrag ab.

Keywords: Logische Sprache, Semantik, Studienordnung, Prüfungsordnung, Modellierung.

1 Einleitung

Es gibt eine Reihe von organisatorischen Problemstellungen, wenn Hochschulen und Universitäten neue Campus-Management-Systeme etablieren bzw. diese im Zuge der Digitalisierung ihrer Prozesse intensiver nutzen. Eine systematische Analyse der Erfolgsfaktoren derartiger Entwicklungen wurde von Schreiter et al. [SAA12] vorgenommen. Gleichzeitig ist die Modellierung von geeigneten Datenstrukturen, die in technischen Systemen nachfolgend umgesetzt werden, eine Herausforderung, die seit mehr als 25 Jahren die Hochschulen begleitet [SM95; BCS13]. In einer Vielzahl von Projekten wurde insbesondere die Unsicherheit bei der Abschätzung vom erforderlichen Aufwand für die technische Abbildung

¹ SemaLogic, Weimar, Germany, markus.von.der.heyde@semalogic.de,  0000-0002-6026-082X

² SemaLogic, Weimar, Germany, matthias.goebel@semalogic.de

der Prüfungs- und Studienordnungen (PStO) im Campus-Management-System (CMS) untersucht [Bau11].

Logische Sprachen bzw. alle prozeduralen Sprachen repräsentieren Wissen, Abläufe oder Zusammenhänge in digitaler, durch den Computer ausführbarer Form. Um Regelwerke und Abläufe aus natürlichsprachlichen Texten erzeugen zu können, müssen diese bisher zunächst fachlich interpretiert und in einem zweiten Schritt programmiert werden. Ob eine in aktuellen CMS erzeugte Programmierung mit dem ursprünglichen Text übereinstimmt, kann de facto nicht geprüft werden.

Habtemariam untersuchte bereits 2006 die Anwendbarkeit von logischen Sprachen auf die Modellierung von PStO [Hab06]. Die gewählte Sprache glich aber in der Komplexität einer üblichen Programmierung und war nur für Experten auf dem Gebiet der Studienorganisation sowie der verwendeten Programmiersprache geeignet. Zudem war die Laufzeit der Prüfung der abgebildeten Ordnung auf Konsistenz ungeeignet, um interaktiv Regeln verändern zu können und Optimierungen während der Designphase der Ordnung berücksichtigen zu können. Die semantische Modellierung und Simulation von PStO in der Sprache CLP erfolgte in [Tet08], führte aber nicht zur einer Umsetzung in den verwendeten CMS, da auch hier die Hürde der Programmierung eine gleichzeitige Verwendung von Ordnungstext und semantischer Repräsentation der Ordnung verhinderte. Wie würde sich die Welt aber verändern, wenn der Computer einen natürlichsprachlichen Text „verstehen“ und die darin enthaltenen Regeln und Anweisungen klar und verlässlich abbilden würde? Könnte der Computer dann bereits während der Formulierung textliche Unklarheiten erkennen und korrigieren und so den Menschen unterstützen? Um diese Unterstützung möglich zu machen, ist es notwendig, dass der im Text enthaltene Sinn, also die Bedeutung der Worte, in beiden Interpretationen übereinstimmt. Eine neutrale semantische Repräsentation müsste hierzu aus einem Text verlässlich in ein technisch lesbares und semantisch identisches Abbild überführt werden. Wenn dies gelänge, kann die Interpretation und die anschließende Programmierung – wie sie beispielsweise bei der Abbildung von PStO im CMS stattfinden muss – in wesentlichen Teilen entfallen. Es würden damit potentielle Fehler während der Erzeugung des Textes, der Interpretation bis hin zur Kodierung ausgeschlossen, solange die schriftliche Form eines Textes wie einer PStO tatsächlich die intendierten Regelungen enthält.

2 Ableitung von Anforderungen aus der Literatur

Zwei grundsätzliche Lösungen existieren, damit aus einem Text der PStO eine logische, semantische Struktur z.B. der Modulreihenfolge abgeleitet werden kann: a) Die Disziplin des Natural Language Processing leitet mithilfe von Heuristiken aus der natürlichen Sprache eine wahrscheinlich gemeinte Struktur ab [KM93], als Review siehe [Ara+14]. b) Eine Sprachdefinition formuliert eindeutig logische Operatoren bzw. formale Aussagen (z.B. Constraints) und durch einen Parser werden die Sprachkonstrukte in die semantische Repräsentation überführt [Hab06; Tet08]. Otto und Anton [OA07] kategorisierten diese

grundsätzlichen Ansätze feiner in ein ganzes Spektrum von hybriden Ansätzen, welche sich primär durch die technische Realisation und die Eigenschaften der vorgeschlagenen spezialisierten Sprachkonstrukte unterscheiden. Für alle Fälle ist es zunächst aber essentiell, die in tatsächlichen Texten vorkommenden logischen Operatoren zu klassifizieren und in der Zielrepräsentation die Vollständigkeit sicherzustellen.

Bereits 1957 formulierte Allen in der Dissertation [All57] sechs grundlegende logische Operatoren, die in juristischen Texten maßgeblich präsent sind:

- Conjunction – das logische *UND*, welches alle Bestandteile zusammenbindet
- Exclusive disjunction – das ausschließliche *ODER*, dem genau jeweils eine von mehreren Alternativen gewählt werden kann
- Inclusive disjunction – ein *ODER*, bei dem eine oder mehrere Wahloptionen gewählt werden können
- Negation – das *NICHT*, bei der die Aussage negiert wird
- Implication – als Folgerung, wenn aus einer Aussage eine zweite folgt
- Coimplication – die gegenseitige Folgerung, bei der aus dem jeweiligen Wahrheitswert einer Aussage die jeweils andere abgeleitet werden kann.

Allen formulierte eine neue Schreibweise der Anordnung von Aussagen, in der diese Verbindungen auf natürliche Texte angewendet werden sollten (systematically-pulverized form), damit die von Juristen formulierten Aussagen symbolisch eindeutig würden.

Peek geht in [Pee97] darüber hinaus, in dem die Informationsstruktur, die den Aussagen unterliegt, klassifiziert wird. Den Worten wird dabei innerhalb eines Satzes eine grammatikalische Bedeutung zugewiesen, welche die Funktion der Worte im Satzgefüge eindeutig festlegt. Die Analyse der Sätze erfolgt durch denjenigen, der die Transkription des ursprünglichen juristischen Textes vornimmt. Tetzner und Riedewald schlagen ferner eine Anwendung von symbolischer Simulation zur Prüfung zeitlicher Abhängigkeiten in Regelwerken wie PStO vor [TR05]. Der dort formulierte Ansatz erweitert endliche Automaten um die Auswertung der zeitlichen Regelwerte und ermöglicht, dass komplexe Zusammenhänge problemadäquat modelliert werden können.

Habtemariam wendete in seiner Dissertation [Hab06] die Re-Formulierung der PStO Texte als Constraints an, um mit existierenden Werkzeugen den formalen Studienablauf als Constraint-Erfüllbarkeitsproblem (Constraint Satisfaction Problem, CSP) bzw. Constraint Optimierungsproblem darzustellen und zu lösen. Als Notation der Constraints werden dazu die endlichen Wertemengen der verwendeten symbolische Variablen und deren Zusammenhänge formuliert. Diese Zusammenhänge sind ungerichtete, also bi-direktionale, Beschreibungen, welche potentiell auch nur in partiellen Bedingungen (z.B. im mathematischen Sinne) zwischen zwei Variablen bestehen. Die Auflösung komplexer Aussagen

mit mehreren Variablen in einfache, welche maximal zwei Variablen besitzen, wurde durch grundlegende Arbeiten zur Programmierung mit Constraints (Constraint Logic Programming - CLP) gezeigt.

In ihrer Dissertation erweiterte Tetzner [Tet08] die Constraint Simulation der CLP durch die nutzerfreundlichen Sprachen MODEL-HS und VYSMO, welche die Formulierung von hybriden endlichen Automaten unter der Verwendung von zeitlichen Constraints erlaubte. Die Mächtigkeit der Aussagen blieb aber durch die grundlegenden Aussagen der CLP konstant. Der nutzerfreundliche Text der MODEL-HS gleicht weiterhin eher einer Programmiersprache als einer natürlichen Sprache, erlaubt aber im Gegensatz zur natürlichen Sprache eine eindeutige Formulierung der intendierten Aussagen. Otto und Anton [OA07] formulierten ebenfalls die Anforderung, dass im Fall einer Prüfung der Bedingungen aus den Regelungstexten im Sinne von Audits eine Rückverfolgung von Aussagen auf deren Quelle unerlässlich sei.

Spitta und Mordau [SM95] schlugen eine konsistente Bezeichnung der Strukturelemente von PStO-Begriffen im Sinne einer Ontologie vor, welche durch Brune et al. [BCS13] weiter entwickelt und umgesetzt wurde. Sicherlich ist für die Austauschbarkeit zwischen Hochschulen ein gemeinsames Verständnis der Begriffe essentiell; für die Prüfung der Konsistenz eines Regelwerkes ist diese allgemeine Ebene nicht erforderlich.

Zusammenfassend kann aus der Literatur geschlossen werden, dass zur Formulierung von juristischen Texten, welche eine Struktur formulieren, folgende logische Operatoren zur Gliederung der Anforderungen notwendig sind:

- *UND* zur Verbindung von Aussagen
- *ODER* als aus- und einschließende Variante
- *NICHT* zur Negation von Aussagen
- *BEDINGT* zur Formulierung von Implikationen bzw. Äquivalenz
- *VOR/NACH* zur zeitlichen Abgrenzung
- *FOLGT AUS* zur Angabe der Quelle
- *BASIERT AUF* als Möglichkeit der Versionsangabe

Eine semantische Repräsentation, die konkret dazu geeignet sein soll, PStO strukturell semantisch abzubilden, muss diese bzw. äquivalente Sprachkonstrukte anbieten.

3 Ableitung von Anforderungen aus PStO

Aus existierenden PStO in Deutschland kann stichprobenartig eine Liste von Anforderungen erstellt und die aus der Literatur abgeleitete Liste der logischen Operatoren dagegen geprüft

werden. Dazu ist es wünschenswert, eine möglichst breite Varianz von PStO mit Stichproben aus verschiedenen Hochschulformen, Fächern und Rechtsräumen zu analysieren. Nicht in die Analyse einbezogen sind Anforderungen, die sich außerhalb eines Regelwerkes zur Abbildung logischer Operatoren befinden. Typische Beispiele dafür sind organisatorische Voraussetzungen oder Zusammenhänge, die bereits durch einfache Datenfelder innerhalb vorhandener CMS abgebildet werden. Dies betrifft insbesondere konkrete Prüfungsregularien (z.B. Anwesenheit von einer bestimmte Anzahl von Prüfern), Fristen zur Anmeldung und Prüfungsdurchführung sowie grundsätzliche Zugangsvoraussetzungen.

3.1 Darstellung der Auswahl von PStO

In einer grundlegenden Analyse wurden insgesamt 20 PStO auf Basis der Kriterien Hochschulform, geographische Verteilung und Abdeckung des Fächerspektrums ausgewählt und in Hinblick auf die darin verwendeten logischen Operatoren untersucht.

Die analysierten PStO gliedern sich in die Hochschulformen Universität(8), Fachhochschule(6), Technische Universität(4) und Kunst & Musikhochschule(2). Die geographische Verteilung berücksichtigt Hochschulen aller 16 Bundesländer, wobei Bayern(2), Hessen(2) und Nordrhein-Westfalen(3) mit mehr als einer PStO in die Untersuchung aufgenommen wurden. Bei Zuordnung zum Fächerspektrum lassen sich Geisteswissenschaften(4), Naturwissenschaften(4), technische Fächer(3), Sozialwissenschaften(4), Kunst oder Musik(2), Theologie(1) und Pädagogik(2) unterscheiden.

3.2 Auswertung der Analyse

Im Rahmen der Analyse wurden insgesamt 1746 unterschiedliche Konstrukte untersucht und in Bezug zu struktur-definierenden, logischen Operatoren gesetzt. Von den untersuchten Konstrukten lassen sich rund 62% (1082) den logischen Konstrukten, die aus der Literatur abgeleitet wurden, zuordnen. Die übrigen 38% (664) Konstrukte formulieren Nebenbedingungen und sprachliche Vereinfachungen von logischen Aussagen (s.g. dynamische Gruppen), die für PStO typisch erscheinen.

Die Mehrheit der Konstrukte - annähernd 58% (1018) - basieren auf wenigen Basis-Operatoren aus der Literatur: 26% (451) auf *UND*-Verknüpfungen, 17% (289) auf *ODER*-Operatoren und 16% (278) zeigen eine zeitliche Bedingung oder Zuordnung auf. Diese Standardverknüpfungen bilden damit die zentralen logischen Konstrukte bei der Formulierung der PStO und müssen bei jeder semantischen Repräsentation beachtet werden. Die Nutzung von *NICHT*-Konstrukten 4% (64) stellt dagegen eine Seltenheit in den PStO dar; davon wurden 60 Fälle von einer Universität als explizite Formulierung verwendet. In allen anderen PStO wird der Ausschluss von Möglichkeiten über Positivformulierungen erreicht, die ein *ENTWEDER/ODER* darstellen und damit in der Analyse dem logischen Konstrukt des *ODER* hinzugerechnet wurden.

Der logische Operator *FOLGT AUS* definiert kein strukturgebendes Konstrukt innerhalb der PStO, sondern stellt wie beschrieben den nachvollziehbaren Zusammenhang zwischen einer definierten Konsequenz und der zugrundeliegenden Voraussetzung her. Neben den im Detail analysierten Konstrukten ist der logische Operator *BASIERT AUF* bei allen untersuchten Institutionen aufgrund unterschiedlicher Versionsstände der jeweiligen PStO grundsätzlich gegeben und muss daher zwingend in einer vollständigen semantischen Repräsentation berücksichtigt werden. Darüber hinaus werden Bezüge auf fächerübergreifende oder institutsübergreifende Ordnungen genutzt, die Vorgaben und übergreifende Regeln beinhalten. Dies motiviert die Berücksichtigung eines *INCLUDE* als Funktion der Sprache.

Neben den bereits aus der Literatur bekannten Konstrukten und Operatoren wurden bei der Analyse 38% (664) Formulierungen identifiziert, die für den Studienverlauf strukturgebend wirken: Nebenbedingungen 9% (160), Zuweisung von Werten 2% (41) durch Formeln 15% (264) und Gruppierung ohne logische Verknüpfung 10% (171) sowie Empfehlungen 2% (28). Die Prüfung von Nebenbedingungen erfolgte z.B. durch „*Zur Abschlussarbeit kann nur zugelassen werden, wer mindestens xxx Credit-Points erbracht hat*“. Einfache mathematische Formeln und die Zuweisungen vom Ergebnis wurde z.B. formuliert mit: „*Das Modul gilt als bestanden, wenn mindestens 50% der Punkte erreicht wurden*“. Zum Teil nahmen die Formulierungen gegenseitig Bezug aufeinander: „*Die Gesamtnote der Bachelorprüfung wird als gewichtete Durchschnittsnote berechnet. Die Gewichtung erfolgt nach den Credit-Points der einzelnen benoteten Module*“. Schließlich wurden mehrere Module/Veranstaltungen, um darauf weiter Bezug nehmen zu können, zu Gruppen mit statischen bzw. dynamischen Benennungen zusammengefasst (bspw. durch „*Die Module WP1 bis WP8. . .*“). Eine Sprache, die den semantischen, strukturgebenden Gehalt der PStO abbildet, muss demnach auch diese vier Kategorien enthalten.

Auffällig in der Analyse war, dass im Umfeld der PStO sprachliche Wenn-Dann-Konstrukte verwendet werden, um instanziierte Regelfolgen oder Wertableitungen abzubilden. Eine Formulierung wie „*Wenn der Studierende den Schwerpunkt [. . .] wählt, so müssen aus dem Wahlpflichtbereich [. . .] zwei Module ausgewählt werden*“, entspricht dabei aber eigentlich einer Folgerung aus den bestehenden *UND* und *ODER*-Regeln, die für diesen Studiengang definiert wurden. Andere Formulierungen wie „*Ein Modul ist bestanden, wenn die Modulprüfung mindestens mit ‚ausreichend‘(4,0) bewertet wurde*“ stellen lediglich eine mathematische Ableitung eines Wertes und ebenfalls kein bedingendes logisches Konstrukt dar.

3.3 Schlussfolgerungen

Die Analyse der PStO bestätigt die Notwendigkeit der grundlegenden, in der Literatur genannten logischen Operatoren. Auch wenn die Varianten der bedingten Formulierung von Implikationen und die Negation von Aussagen nur selten bzw. nicht in PStO vorkommen, so sollten sie dennoch in einer semantischen Repräsentation in ihrem Ergebnis zumindest

über entsprechende Formulierungsmöglichkeiten in Form von *UND* oder speziellen *ODER*-Operatoren verfügbar sein.

Ein weiteres wesentliches Ergebnis der Analyse ist, dass sich Lösungen zur Abbildung von PStO nicht nur auf die Formulierung von logischen Operatoren beschränken dürfen. Es ist essentiell, dass die semantischen Regeln gleichzeitig dafür genutzt werden können, um Nebenbedingungen und mathematische Funktionen auf dieser Basis beschreiben zu können. Um einfache Bezüge ohne eine logische Einordnung von Modulen vornehmen zu können, müssen überdies Gruppen gebildet werden können.

4 Anforderungen aus bisherigen Implementationen

Die überwiegende Menge der Abbildungen von PStO in Deutschland erfolgte in den Systemen der Hersteller HIS (HISinOne, HIS GX), Datenlotsen (CampusNet), SAP (SLCM) und durch das System der TU Graz (CAMPUSonline). Um für den deutschen „Markt“ eine relevante Lösung zu präsentieren, sollten die Fähigkeiten dieser Systeme inkludiert sein. Die öffentlich zugänglichen Dokumente sowie Dokumente aus der Zusammenarbeit im ZKI (AK-Campus-Management) wurden dementsprechend analysiert.

Die Beschreibungen der Fähigkeiten der Systeme auf den Webseiten sind sehr allgemein gehalten. Zum Teil wird die Anwendbarkeit für ein typisches Fächerspektrum genannt. Allgemeine Aussagen zur Versionierung und Wiederverwendbarkeit von Modul-Definitionen lassen auf die strukturierte Speicherung der Information in Datenbanken schließen. Eine Verbindung zwischen dem PStO Texten und der im CM-System hinterlegten Abbildung wurde nicht gefunden.

5 Entwicklung der Sprache

Unter Beachtung der Anforderungen aus der Literatur und den Ergebnissen der Analyse von bestehenden PStO wurde eine neue semantische Repräsentation als Kodierung von Regeln entwickelt. Dazu passend wurde eine neue technologische Lösung erstellt, welche die Wandlung von natürlichem Text in diese eindeutige strukturelle Repräsentation stark vereinfacht. Texte können damit natürlichsprachlich aussehen und sowohl vom Menschen als auch vom Computer gleichermaßen gedeutet werden. Somit könnten PStO von den Gremien formuliert und direkt aus der vorhandenen Formulierung vom Computer verlässlich interpretiert werden.

Durch die Vereinbarung einer klaren Satzstruktur entfallen Mehrdeutigkeiten, die in natürlicher Sprache vorkommen. Auf eine Klassifikation der Satzbausteine (im Sinne [Pee97]) in der natürlicher Sprache wird verzichtet und stattdessen eine der menschlichen Sprache nahe und vereinfachte Grammatik vereinbart, die eine größere Flexibilität als klassische Programmiersprachen ermöglicht.

5.1 Beschreibung der Sprache

Die semantische Repräsentation enthält u.a. die folgenden Konstrukte (und passende Beispiele bzw. Anwendungsszenarien jeweils in Klammern):

- Unterscheidung von Definitionen, Nebenbedingungen und Empfehlungen
- Intervalle von Werten aus ganzen oder reellen Zahlen bzw. Symbolen
- Exklusiver und inklusiver Oder-Operator mit Anzahl der gültigen Optionen (mindestens 3, höchstens 5 von insgesamt 8 Alternativen)
- Verbindung von notwendigen Aussagen durch *UND*-Operator
- bedingte mathematische und boolesche Funktionen zur Ableitung und Festlegung von Werten (Summe „Workload-Punkte“ für bestandene Module) auf Basis der über Operatoren verbundenen Symbole
- Statische und dynamische Gruppen (alle Kurse Biologie Grundlagen 1 bis Biologie Grundlagen 9)
- Abbildung der zeitlichen Abfolge (Modul A liegt vor Modul B)
- Versionierung (für alle Definitionen, Nebenbedingungen und Empfehlungen)
- angekündigte Werte (für erwartete Benotungen)
- Inkludierung von Regelwerken (Nutzung und Definition von Rahmenregelwerken)
- Kaskadierung von begrifflichen Ebenen (Biologie.Grundstudium)
- Festlegung von Wertebereichen (z.B. für Notenwerte oder Semesterlaufzeiten)
- Instanziierung (Belegung von Werten mit individuellen Studienergebnissen)

Folgende Konstrukte werden nicht als Sprachelemente explizit angeboten, da diese durch andere Konstrukte substituierbar sind: Die Negation als logische Basis wird über den Ausschluss innerhalb der *ODER*-Mengen realisiert, da dort auch null von null Optionen einer Menge gewählt werden können. Die Implikation (*Wenn-Dann*) wird unterschieden nach logischen und zeitlichen Bedingungen, entweder durch eine Gleichheit (*Äquivalenz*) oder durch eine Zusammenfassung von Aussagen.

5.2 Konsistenzprüfung

Im Rahmen der Analyse wurden semantische Problemstellungen, die in aktuell beschlossenen PStO vorliegen, erfasst und beispielhaft fragwürdige Stellen gesammelt. Eine bereits in [All57] ausgiebig analysierte sprachliche Unschärfe ergibt sich, da im Fließtext Aussagen

nicht klar gruppiert werden können. So wird z.B. das Wort „und“ sowohl als Verbindung einer Aufzählung (auch mit *ODER*-Charakter) als auch als Verbindung von gleichzeitig notwendigen Voraussetzungen (boolesches *UND*) verwendet. Aussagen, die mit *UND* bzw. *ODER* verbunden werden, stellen somit keinen eindeutigen sprachlich logischen Bezug her.

Zur Illustration einige Beispiele:

- „*Wer das Wahlpflichtmodul WP 76 wählt, darf nicht die Wahlpflichtmodule WP 30 und WP 40 wählen.*“ Ist hier gemeint, dass man, wenn WP 76 gewählt wurde, a) weder WP 30 noch WP 40 wählen darf, oder dass man b) WP 30 und WP 40 nicht gemeinsam wählen darf? Vermutlich ist Interpretation a) gemeint, aber sprachlich sauberer wäre dann: Wer X wählt, darf weder Y noch Z wählen.
- „*[Dazu sind] für den Wahlpflichtbereich xyz die Wahlpflichtmodule (WP 14 oder WP 15), WP 40 und WP 55 zu wählen.*“ Hier liegt die Interpretation, dass WP 14 und WP 15 sich gegenseitig exklusiv ausschließen, nahe. Sprachlich eindeutig sind so formulierte *ODER*-Bedingungen nicht, so dass eine Wahl von WP 14 und WP 15 nicht eindeutig vermieden wird.

Derartige sprachliche Probleme betreffen zunächst nur den Übersetzungsschritt aus einem natürlichsprachlichen Text in eine eindeutige Semantik. Die neu entwickelte Sprache bietet dazu eine flexible, jedoch die Eindeutigkeit erhaltende Basis.

Wenn die aus dem Text entstehende mathematische Bedeutung der Repräsentation eindeutig ist, kann die formulierte Ordnung durch die neue Technologie zusätzlich auf Konsistenz geprüft werden: Logische Bezüge zwischen Begriffen und zeitliche Abfolgen werden auf Zirkelbezüge geprüft; Wahloptionen, die nur scheinbar existieren, aber durch Nebenbedingungen letztlich ausgeschlossen werden, werden aufgedeckt; Lücken im Übergang zwischen Versionen werden erkannt. Eine erste Liste von mehr als 30 Konstellationen wurde im Rahmen der Implementierung des Prototyps erstellt. Diese können zwar durch die Flexibilität der Sprache in einer semantischen Repräsentation abgebildet werden, sind in sich aber nicht notwendigerweise konsistent. Ein umfangreicher Nachweis, dass derartige Fehler tatsächlich in PStO vorliegen, kann allerdings erst erbracht werden, wenn die Technologie zur Abbildung von vielen Ordnungen verwendet wird und dabei diese Fehler identifiziert werden. Eine Anwendung der Technologie in der Breite steht aber noch aus, so dass zu diesem Zeitpunkt keine Übersicht erstellt werden kann.

5.3 Allgemeingültigkeit

Die durch die Anwendung der strukturgebenden semantischen Sprache erzeugte Repräsentation ist allgemeingültig und unabhängig von spezifischen Eigenschaften der Umsetzung eines konkreten CMS.

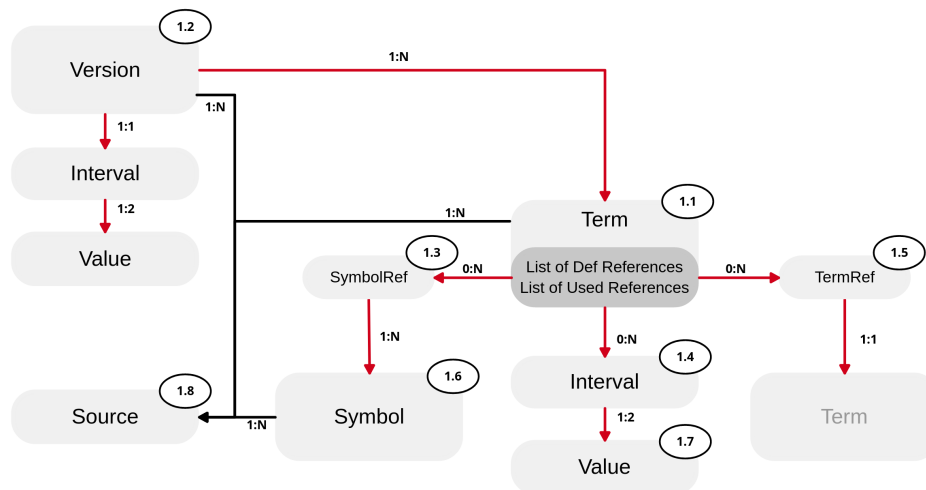


Abb. 1: Allgemeine Objektstruktur der semantischen Repräsentation

Abbildung 1 zeigt die gewählte konkrete Objektstruktur der semantischen Repräsentation. Grundsätzlich werden alle Sprachkonstrukte (Definitionen, Nebenbedingungen und Empfehlungen) als Term **1.1** aufgefasst, der einer Versionierung **1.2** unterliegt. Terme referenzieren potentiell Symbole **1.3**, Intervalle **1.4** und weitere Terme **1.5**. Sie dienen damit sowohl der Abbildung logischer Operatoren als auch der Abbildung von Funktionen oder mathematischen Rechenoperationen. Symbole **1.6** dienen der Kaskadierung und Abstraktion von Namensräumen. Die Intervalle werden verwendet, um Versionen oder die Menge der ODER-Auswahl zu repräsentieren, deren Unter- und Obergrenze durch Werte **1.7** dargestellt werden. Ein Intervall mit einem Wert entspricht der normalen Speicherung von Ganzzahlen, Fließkommazahlen, Strings oder symbolischen Werten wie ∞ . Schließlich werden alle Sprachkonstrukte mit Bezug zum Quelltext (Source) **1.8** verwaltet. Die rekursive Schachtelung aller Terme erlaubt beliebige Kaskadierung der abzubildenden semantischen Struktur. Die Unterscheidung verschiedener Statements ist durch die Typisierung der Terme jederzeit erweiterbar.

Diese Allgemeingültigkeit der logisch semantischen Regeln ist entscheidend, damit grundsätzlich die Möglichkeit gegeben ist, diese Repräsentation sowohl in einen Text zurückzuführen als auch für technische Systeme wie die CMS bereitzustellen.

Somit wäre eine technische Lösung potentiell sogar in der Lage, automatisch auf Eigenschaften oder Regelformulierungen der PStO aufmerksam zu machen, deren Abbildung in einem konkreten System problematisch sein könnten. Gleiches gilt für PStO, die von einem CMS in ein anderes übertragen werden sollen. Hier kann die Zwischenebene der neutralen Repräsentation genutzt werden, um Vereinfachungen der Regelformulierungen so vorzunehmen, dass ein konkretes Zielsystem optimal unterstützt wird. Voraussetzung dafür

ist die generelle Unterstützung einer einheitlichen semantischen Repräsentation innerhalb der Systeme nach einem gemeinsamen Standard oder zumindest die Erstellung geeigneter Import- & Export-Konfigurationen.

6 Aktueller Stand der Umsetzung

Das in diesem Beitrag vorgestellte Projekt befindet sich aktuell in der Erstellung eines Prototypen zur Verifikation der getätigten Aussagen.

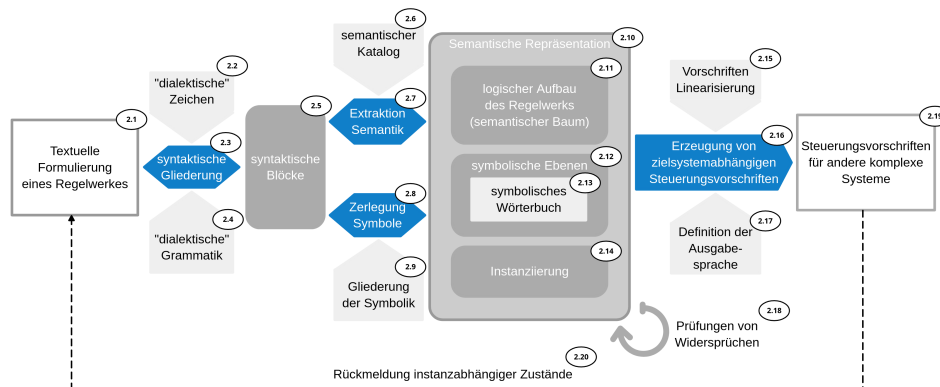


Abb. 2: Funktionaler Aufbau des Prototyps

Abbildung 2 stellt den funktionalen Aufbau des Prototyps dar, der auf der oben dargestellten Form der semantischen Repräsentation basiert. Die von uns implementierte Standard-Sprache erlaubt die Erweiterung um beliebige Symbole: Alle Standard-Symbole sind flexibel für s.g. Dialekte neu oder zusätzlich zu definieren (2.2); dies gilt selbst für Kommentar-Start- und -Ende-Zeichen, Leerzeichen und Zeilenenden. Die zum Einlesen von Texten (2.1) verwendete Syntax ist auf Basis der Übergänge einer State Machine konfigurierbar (2.4) und ohne Programmänderung anpassbar. Die Möglichkeit, die Ausgabe auf Basis einer Syntax-Konfiguration gestalten zu können, bereitet die schnelle Entwicklung von Input/Export-Schnittstellen für CMS-Systeme vor (2.15, 2.17).

Aktuell ist der Prototyp bereits in der Lage, sämtliche Sprachkonstrukte korrekt zu identifizieren (2.3), einzulesen und die semantische Repräsentation (2.10) zu erzeugen (2.7, 2.8). Diese kann vollständig in einem standardisierten Format ausgegeben werden und dem Menschen damit den Einblick in das repräsentierte Regelwerk bzw. eine repräsentierte Instanz (z.B. einen Studienverlauf) geben. Für die semantische Prüfung ist die Erkennung von logischen Fehlern (2.18) weitgehend umgesetzt und beinhaltet: Schleifen in jeglichen semantischen Konstrukten (mathematische Formeln, Zeit, logische Ausdrücke); Widersprüche in komplexen ODER- / UND-Kombinationen; Widersprüche zwischen Versionen des Regelwerks.

6.1 Beispiel

Der Abdruck einer kompletten PStO wäre im Rahmen des Artikels nicht sinnvoll, da die Sprache primär die strukturgebenden Aspekte abbildet. Es wurde daher eine Passage aus einer analysierten Studienordnung ausgewählt, die auf kleinem Raum viele verschiedene Sprachkonstrukte benutzt. Zur Übersichtlichkeit werden alle enthaltenen Aussagen nummeriert. Zunächst wird der Abschnitt im Original zitiert. Anschließend wird der semantische Gehalt in SemaLogic dargestellt. Abschließend wurde eine erweiterte Sprach-Symbolik gewählt, die den Text möglichst nahe am originalen Text bzw. der deutschen Sprache belässt, aber vollständig und eindeutig in die semantische Repräsentation überführbar ist.

- (1) Aus den Wahlpflichtmodulen WP 18 bis WP 26 und WP 68 bis WP 77 sind insgesamt zwei Wahlpflichtmodule zu wählen.
- (2) Dabei soll im 3. und 6. Fachsemester jeweils ein Wahlpflichtmodul gewählt werden.
- (3) Wer das Wahlpflichtmodul WP 19 wählt, darf nicht die Wahlpflichtmodule WP 2 und WP 3 wählen.
- (4) Wer das Wahlpflichtmodul WP 76 wählt, darf nicht die Wahlpflichtmodule WP 30 und WP 40 wählen.

List. 1: Text der PStO

Als Standard-Sprache formuliertes Beispiel:

- (1) Wahlpflichtmodule 2|2 {Wahlpflicht ~ WP 18 | WP 26, WP 68 | WP 77 ~}
- (2) ! 3te Fachsemester 1|1 {Wahlpflicht}
! 6te Fachsemester 1|1 {Wahlpflicht}
- (3) [WP 19, 0|0 {WP 2, WP 3}]
- (4) [WP 76, 0|0 {WP 30, WP 40}]

List. 2: SemaLogic als Standard-Sprache

Als Sprachanpassung formuliertes und semantisch identisch überführbares Beispiel:

- (1) Wahlpflicht umfasst WP 18 bis WP 26, WP 68 bis WP 77 als Gruppe.
Wahlpflichtmodule enthält 2 bis 2 Alternativen aus Wahlpflicht.
- (2) Empfohlen wird: 3te Fachsemester enthält 1 bis 1 Alternativen aus Wahlpflicht.
Empfohlen wird: 6te Fachsemester enthält 1 bis 1 Alternativen aus Wahlpflicht.
- (3) Nebenbedingung 1 besteht aus WP 19 sowie 0 bis 0 Alternativen aus WP 2, WP 3..
- (4) Nebenbedingung 2 besteht aus WP 76 sowie 0 bis 0 Alternativen aus WP 30, WP 40..

List. 3: GI-Paper Dialekt

Für den letzten Abschnitt wurden folgende Definitionen vorgenommen, so dass der Prototyp die Sprachelemente zusätzlich auf die Standard-Sprache mappen kann:

- Start einer Gruppe: „ umfasst “ \Rightarrow „~“
- Ende der Gruppe: „ als Gruppe“ \Rightarrow „~“
- Trennzeichen für Intervalle: „ bis “ \Rightarrow „|“
- Beginn einer Empfehlung: „Empfohlen wird:“ \Rightarrow „!“
- Überlesen als Leerzeichen: „ enthält “ \Rightarrow „ “
- Start einer *ODER*-Regel: „ Alternativen aus “ \Rightarrow „{“
- Start einer *UND*-Regel: „ besteht aus “ \Rightarrow „[“
- Trenner zwischen Symbolen: „ sowie “ \Rightarrow „,“
- Ende von *UND*- und *ODER*-Mengen: „,“ \Rightarrow „}“ und „]“

Zugegebenermaßen leidet der flüssige deutsche Satzbau etwas unter den Restriktionen der statischen Grammatik der bisher definierten Sprache. Dennoch sind die im letzten Abschnitt formulierten Sätze bereits näher an der natürlichen Sprache als an der hier definierten Standard-Sprache oder einer anderen formalen Programmiersprache wie Prolog. Dem Nutzer steht es frei, durch weitere Definitionen die Formulierung noch weiter an den üblichen Sprachgebrauch anzupassen.

7 Ausblick

Die aktuelle standardisierte Sprache und die damit verbundene semantische Repräsentation ist vollständig auf wenigen Seiten beschreibbar und wird in Kürze veröffentlicht. Die technische Umsetzung zur Steuerung oder Konfiguration komplexer Plattformen ist zusammen mit der Sprachdefinition zum Patent angemeldet. Es wird ferner angestrebt, die aktuellen Arbeiten in die Bemühungen um den XHochschul-Standard einfließen zu lassen.

7.1 Semantik als Basis für weitere Anwendungsfälle

Die von uns definierte semantische Repräsentation eignet sich sowohl zur Darstellung eines Regelwerks einer PStO als auch zur Speicherung von Studienverläufen. Weiterhin kann auf Basis der semantischen Repräsentation die Zahl der tatsächlich unterschiedlichen Varianten eines Studiums berechnet werden und vieles mehr. Da die Repräsentation auch mit Werten einer individuellen Instanz - also einem konkreten Studienverlauf - befüllt werden kann, ist eine Prüfung von hypothetischen Studienverläufen möglich. Wenn

zukünftig mehrere dieser Hypothesen gleichzeitig getestet werden, entstehen dadurch für die modellierte PStO statistische Vorhersagen des Kapazitätsbedarfs bei gegebenen Annahmen von Studierendenzahlen.

Gemeinsam mit der zugrunde liegenden semantischen Beschreibung der PStO bietet die semantische Repräsentation eine ideale Plattform für den Austausch bzw. die Anerkennung von Studienleistungen auf nationaler und europäischer Ebene. Die Besonderheit besteht dabei in der Neutralität des gespeicherten Wissens und der Möglichkeit, bei der Ausgabe eine andere syntaktische Grammatik und Sprache zu wählen als bei der Kodierung. Natürlich müssten verwendete Begriffe wie „Biologie“ extern als Matching bereitgestellt werden, da unsere Technologie die Übersetzung einzelner, einfacher Begriffe voraussetzt und sich auf die semantische Kodierung der Zusammenhänge fokussiert.

7.2 Einschränkungen

Die vorgeschlagene semantische Repräsentation, welche strukturierte, logische Zusammenhänge, die durch Regelwerke formuliert werden, abbilden kann, ist nicht für alle denkbaren Szenarien von juristischen Texten geeignet. So werden z.B. keine dehnbaren Begriffe („in geeigneter Weise“) oder solche mit klaren juristischen Bedeutungen („unverzüglich“ entspricht „ohne schuldhaftes Zögern“) abgebildet. Die vorgeschlagene Sprache fokussiert auf strukturgebenden, logischen Regeln, welche die Voraussetzungen zum Erreichen eines angestrebten Zustands (z.B. dem Abschluss eines Studiums) definieren.

Die o.g. Liste der Konstellationen, bei denen potentielle Widersprüche und andere Mängel der Konsistenz auftreten, ist noch nicht abgeschlossen. Ferner gehen wir aktuell davon aus, dass eine solche Liste endlich sein kann, damit letztlich alle Mängel durch ein automatisches Verfahren gefunden werden können. Ein Beweis dieser Annahme steht aus.

Schließlich wird die Anerkennung von Leistungen potentiell von der semantischen Struktur profitieren können. Eine Zuordnung der Inhalte ist Bestandteil der Modulbeschreibungen und nicht Teil der Aufbaustruktur der in den PStO definierten Studiengänge. In Konsequenz wird diese Ebene nicht in der hier vorgestellten Sprache abgebildet.

7.3 Schlussfolgerungen

Um die Wirksamkeit der semantischen Prüfung zu erreichen, wurde die in diesem Beitrag vorgeschlagene Sprache in Relation zur natürlichen Sprache stark vereinfacht. Die große Flexibilität in der Nutzung von vereinbarten Sprachsymbolen und einer ggf. durch Konfiguration adaptionsfähigen Grammatik erlaubt Satzstrukturen, die dem korrekten (z.B. deutschen) Satz sehr ähnlich sind und intuitiv verstanden werden. Eine Neuformulierung von Ordnungen in den Passagen, die den studienorganisatorischen Aufbau betreffen, muss dennoch in angemessener Weise erfolgen, um das Konzept anwenden zu können. Für alle

Ordnungen, die in solcher Weise umformuliert werden, kann ein Kompromiss gefunden werden, bei dem der Text sowohl vom menschlichen Leser als auch vom Computer identisch verstanden wird. Die zusätzliche Umsetzung in Campus-Management-Systemen zur Strukturierung des Studiums in Modulen und deren gegenseitige Abhängigkeiten kann dann entfallen.

8 Danksagung

Die Autoren möchten sich beim Team der LMU für den offenen Austausch zu den Schwierigkeiten bei der Abbildung von PStO bedanken. Weiterhin bedanken sich die Autoren bei den Gutachtern für die nützlichen Hinweise zum Manuskript.

Literatur

- [All57] Layman E. Allen. „Symbolic Logic: A Razor-Edged Tool for Drafting and Interpreting Legal Documents“. Diss. Yale Law School, 1957.
- [KM93] J.-T. Kim und D.I. Moldovan. „Acquisition of semantic patterns for information extraction from corpora“. In: *Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications*. Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications. März 1993, S. 171–176. DOI: 10.1109/CAIA.1993.366645.
- [SM95] Thorsten Spitta und Jörn Mordau. *Entwicklung und Ergebnisse eines allgemeingültigen Fachkonzeptes für die Prüfungsverwaltung an Hochschulen*. Universität Bielefeld, Okt. 1995, S. 127–147. URL: <https://pub.uni-bielefeld.de/download/2669014/2679395/PrufungsVerwaltung-Siegen95.pdf>.
- [Pee97] Niels Peek. „Representing Law in Partial Information Structures“. In: *Artificial Intelligence and Law 5.4* (1. Dez. 1997), S. 263–290. ISSN: 1572-8382. DOI: 10.1023/A:1008238332032. URL: <https://doi.org/10.1023/A:1008238332032> (besucht am 18.07.2020).
- [TR05] Elke Tetzner und Günter Riedewald. „Spezifikation und Verifikation in regelbasierten Beratungssystemen auf der Grundlage hybrider Automaten“. In: *Programmiersprachen und Rechenkonzepte*. 22. Workshops der GI-Fachgruppe 2.1.4. Bd. Bericht Nr. 0513. Bad Honnef, 2005, S. 77. URL: <http://www-ps.informatik.uni-kiel.de/fg214/Honnef2005/TechnischerBericht0513.pdf#page=85>.
- [Hab06] Dereje Tewelde Habtemariam. „Simulation von Prüfungsordnungen und Studiengängen mit Hilfe von Constraint-logischer Programmierung“. Diss. Frankfurt am Main, 2006. URL: <https://d-nb.info/1046848909/34>.

- [OA07] Paul N. Otto und Annie I. Anton. „Addressing Legal Requirements in Requirements Engineering“. In: *15th IEEE International Requirements Engineering Conference (RE 2007)*. 15th IEEE International Requirements Engineering Conference (RE 2007). ISSN: 2332-6441. Okt. 2007, S. 5–14. DOI: 10.1109/RE.2007.65.
- [Tet08] Elke Tetzner. „Nutzerfreundliche Modellierung mit hybriden Systemen zur symbolischen Simulation in CLP“. Diss. Rostock: Universität Rostock, 5. Sep. 2008. 282 S. URL: <https://d-nb.info/993978975/34>.
- [Bau11] Nils-Joachim Bauer. „Sieben Jahre integriertes Campus Management an deutschen Hochschulen“. In: *Allgemeine Aspekte zur prozessorientierten Hochschule*. Bock und Herchen Verlag, Aug. 2011, S. 39–50. URL: http://www.dini.de/fileadmin/docs/Prozessorientierte_Hochschule_2011.pdf#page=39 (besucht am 13.07.2020).
- [SAA12] Jan Schreiter, Rainer Alt und Gunnar Auth. „Business Engineering bei der Einführung von Campus-Management-Systemen–Herausforderungen und Potenziale“. In: *INFORMATIK 2012* (2012). Publisher: Gesellschaft für Informatik eV, S. 642–656. URL: <https://dl.gi.de/handle/20.500.12116/17866>.
- [BCS13] Henning Brune, Marco Carolla und Thorsten Spitta. „Studiengangmodellierung - Ein implementierter Diskussionsansatz -“. In: (2013). Accepted: 2017-12-06T09:52:38Z Publisher: Gesellschaft für Informatik e.V. ISSN: 1610-5753. DOI: 10.1007/BF03344342. URL: <http://dl.gi.de/handle/20.500.12116/8849> (besucht am 13.07.2020).
- [Ara+14] Denis Andrei de Araujo u. a. „Information extraction for legal knowledge representation – a review of approaches and trends“. In: *Revista Brasileira de Computação Aplicada* 6.2 (29. Mai 2014), S. 2–19. ISSN: 2176-6649. DOI: 10.5335/rbca.2014.3542. URL: <http://seer.upf.br/index.php/rbca/article/view/3542> (besucht am 18.07.2020).