

Distributed Discussions in Online Social Networks

Masterarbeit

Florian Müller

Betreuer: Prof. Dr. Max Mühlhäuser

Verantwortlicher Mitarbeiter: Dipl.-Inform. Kai Höver

Darmstadt, September 2013



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Informatik
Telekooperation
Prof. Dr. Max Mühlhäuser

Ehrenwörtliche Erklärung

Hiermit versichere ich, die vorliegende Masterarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in dieser oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, September 2013

(Florian Müller)



Zusammenfassung

Inhalt der Zusammenfassung

Abstract

Content of the abstract



Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen	3
2.1. Zugriff auf Webanwendungen	3
2.1.1. Representational State Transfer (REST)	3
2.1.2. Simple Object Access Protocol (SOAP)	4
2.2. Datenintegration	5
2.2.1. Semantic Web	5
2.2.2. Ontologien	9
2.3. Datenverteilung	12
2.3.1. Java Messaging Service	12
2.3.2. Enterprise Integration Pattern (EIP)	13
2.4. Lernplattformen und soziale Online-Netzwerke	15
2.4.1. Moodle	15
2.4.2. Canvas	16
2.4.3. Youtube	17
2.4.4. Facebook	18
2.4.5. Google+	18
2.5. Verwandte Arbeiten und Projekte	18
2.5.1. What happens when Facebook is gone?	18
2.5.2. Reclaim Social	19
3. Analyse	21
3.1. Ein Beispiel	21
3.1.1. Beiträge lesen und konvertieren	21
3.1.2. Beiträge in eine andere Webseite schreiben	23
3.2. Identifizierung der Komponenten	23
3.3. Wahl vorhandener Techniken	24
3.3.1. Welches Zwischenformat?	24
3.3.2. Welches Nachrichtensystem?	25
4. Eigener Ansatz: Social Online Community Connectors (SOCC)	27
4.1. Datenformat	28
4.2. Konfiguration	29
4.2.1. SOCC Connector Config Ontologie	29
4.2.2. Services	30
4.2.3. Benutzerdaten	31
4.2.4. Authentifizierung	32
4.2.5. Autorisierung	34

4.3. Design eines Connectors	36
4.3.1. SOCC Context	36
4.3.2. AccessControl	36
4.3.3. ClientManager	37
4.3.4. StructureReader	38
4.3.5. PostReader	38
4.3.6. PostWriter	39
4.4. SOCC-Camel	40
4.4.1. SoccPostPollingConsumer	40
4.4.2. SoccPostProducer	41
5. Implementierung und Evaluation	45
5.1. Verwendete Programme und Bibliotheken	45
5.1.1. Apache Maven	45
5.1.2. RDF2Go	45
5.2. Implementierung der Connectoren	45
5.2.1. Allgemeine Informationen zum Mapping nach SIOC	45
5.2.2. Allgemeine Probleme und Lösungen bei der Implementierung	47
5.2.3. Moodle Connector	47
5.2.4. Facebook Connector	51
5.2.5. Google+ Connector	55
5.2.6. Youtube Connector	58
5.2.7. Canvas Connector	61
5.3. Evaluation	64
6. Zusammenfassung und Ausblick	67
6.1. Ausblick	68
A. Anhang	69
A.1. SOCC Connector Config Ontologie	69
A.2. SIOC Services Authentication Module	71
Literaturverzeichnis	76

Abbildungsverzeichnis

2.1. Einfacher RDF-Graph	7
2.2. Aufbau von SIOC (modifiziert) - Originalquelle: [13]	11
2.3. JMS Programmiermodell - Original Bild: http://docs.oracle.com/javaee/1.3/jms/tutorial/1_3_1-fcs/doc/images/Fig3.1.gif (Zugriff: 2013-09-15) . .	12
2.4. Moodleinstanz der TU Darmstadt	16
2.5. Instructure Canvas	17
3.1. Benutzer erstellt einen Beitrag im sozialen Netzwerk A.	21
3.2. Komplexität ohne und mit dem Einsatz eines Zwischenformats - Originalbild: [36]	22
3.3. Lesen des erstellten Beitrags und konvertieren in das Zwischenformat.	23
3.4. Konvertierten des Beitrags in das Format B und schreiben n das soziale Netzwerk B	24
4.1. Übersicht der Komponenten der SOCC	28
4.2. Schema der SOCC Connector Config Ontology	29
4.3. SIOC Services Module	31
4.4. Zusammenhang von Person, UserAccount und Service. Die inversen Eigenschaften sioc:account_of und siocs:service_of wurden zu einer besseren Übersicht weggelassen	32
4.5. SIOC Services Authentication Ontology	34
4.6. Basic Access Control Ontologie	35
4.8. SOCC Context	36
4.9. AccessControl	36
4.10. ClientManager	37
4.11. StructurReader	38
4.12. PostReader	38
4.7. UML Klassendiagramm eines Connectors	42
4.13. UML Sequenzdiagramm eines PostWriters	43
4.14. Übersicht des Socc-Camel Moduls in EIP-Notation	43
5.1. Erhalten der Kommentarstruktur beim Schreiben von Beiträgen	48
5.2. UML Klassendiagramm von CanvasLMS4J	65



Tabellenverzeichnis

2.1. Wichtigsten HTTP Operationen mit REST	4
2.2. Einige Beispiel von EIP	14
3.1. Anzahl Konverter bei drei Webseiten	22
4.1. Variablennamen und Ersetzung innerhalb von <code>unknownMessageTemplate</code>	30
5.1. Allgemeine Platzhalter und deren Beschreibung für das Mapping nach SIOC	46
5.2. URI Format für SIOC-Objekte aus Moodle	50
5.3. Für SOCC wichtige Facebook Permissions	53
5.4. Format der URIs für Canvas	64



1 Einleitung

Durch die Omnipräsenz des Internets im heutigen Alltag haben sich viele Bereiche unseres Lebens sehr verändert. Unter anderem die Art wie wir uns weiterbilden und neue Dinge lernen verlagert sich immer mehr dort hin. Prägend für diesen Umbruch ist der Begriff des „E-Learnings“. Besonders neue Technologien im Zuge des so genannten Web 2.0 wie Blogs, Wikis Diskussionsseiten und sozialer Netzwerke machen es immer leichter neues Wissen zu erwerben und es mit anderen zu teilen. Gerade beim Lernen spielt die Diskussion mit Gleichgesinnten eine wichtige Rolle [14]. Es wurden Studien durchgeführt, die zeigen dass Studenten welche sich an online Diskussionen teilnehmen dazu tendieren bessere Noten zu bekommen, als solche die nicht teilnahmen [12, 30]. Auch für zurückhaltende Studenten ist E-Learning eine Verbesserung, da sie sich so eher an Diskussion beteiligen als zum Beispiel in der Vorlesung oder der Lerngruppe [25].

Qiyun Wang et. al. [37] zeigen in ihrer Studie, dass sich Gruppen im sozialen Netzwerk Facebook für den E-Learning Einsatz als Learning Management System (LMS) gut nutzen lassen. Teilnehmer konnte auf der Gruppenseite durch Kommentare und Chats mit einander diskutieren. Gerade die Organisation der Lerngruppe als auch die Benachrichtigung über Ereignisse funktionierte reibungslos. Jedoch uneingeschränkt konnte Facebook als LMS nicht empfohlen werden. Bemängelt wurde unter anderem die aufwändige Integration von Lernmaterialien „tutor noticed that it was quite troublesome to add teaching materials“ [37, S. 435] und dass es nicht möglich war Diskussionen in einzelne Themen zu unterteilen. Alle Kommentare wurden nur als eine chronologische Liste dargestellt. Seit 2013 ist es aber auch auf Facebook möglich auf Kommentare direkt zu antworten¹ und so sind auch forumsähnliche Diskussionen realisierbar. Jedoch ist diese Erweiterung auf „Pages“ beschränkt. Über eine Ausweitung auf Gruppenseiten ist nichts bekannt.

Aber nicht nur soziale Netzwerke sind für Diskussionen innerhalb von E-Learning geeignet, Foren oder Blogs sind ebenfalls sehr beliebte Plattformen. Jedoch ein Problem bei der Nutzung des Internets zum Lernen liegt darin, dass es in der Regel nicht nur eine Plattform genutzt wird, sondern häufig mehrere simultan. Zum Beispiel könnte für einen Kurs ein eigenes Forum im LMS des Veranstalters und nebenbei noch eine Gruppe in Facebook existieren. Teilnehmer, die vorzugsweise nur eine der Plattformen nutzen, erhalten vielleicht von wichtigen Diskussion auf der anderen Plattform keine Kenntnis. Durch diese Inselbildung werden Themen mehrfach behandelt, da Suchfunktionen nur innerhalb der eigenen Plattform suchen und von der Existenz in der Anderen nichts wissen. Eine Integration von zusätzlichen Wissensquellen ist nur schwer möglich und erfolgt immer wieder nur in verbaler Form wie „Schau dir auf der Seite x den Artikel y an“.

Aus diesen Gründen soll in dieser Arbeit ein Ansatz entwickelt werden der es ermöglicht verteilte Diskussionen zusammen zu führen und wiederverwenden zu können. Ein solcher Ansatz muss dazu mehrere Anforderungen erfüllen. Da es sich bei online Plattformen in der Regel um

¹ <https://www.facebook.com/notes/facebook-journalists/improving-conversations-on-facebook-with-replies/578890718789613>

abgeschottete „Datensilos“[4] handelt auf die nur über zum Großteil heterogene Schnittstellen zugegriffen werden kann, ist es hier wichtig eine einheitlich Schnittstelle für den Zugriff auf die gespeicherten Diskussionsdaten zu schaffen. Nicht nur in der Art des Zugriffs unterscheiden sich die einzelnen Plattformen, auch das Format der Daten ist davon Betroffen. Um Diskussionen zwischen den Plattformen überhaupt austauschen zu können ist demzufolge eine Umwandlung in ein gemeinsames Datenformat notwendig, welches erst eine Interoperabilität möglich macht. Als letztes muss noch ein System zur automatischen Synchronisation entwickelt werden wodurch verteilte Diskussionen aktuell gehalten werden können.

Diese Arbeit gliedert sich dazu in folgende Kapitel:

Kapitelbeschreibung

2 Grundlagen

mehr QUELLEN!!!

Grundlagen Einleitung schreiben

2.1 Zugriff auf Webanwendungen

Der Zugriff auf Daten von einer Webanwendung ist in den seltensten Fällen durch eine direkte Anbindung an die dahinter liegende Datenbank möglich beziehungsweise gewünscht. Gerade wenn das eigene Geschäft von diesen Daten abhängt, will man nur ungern alles mit allen teilen. Um trotzdem Dritten die Nutzung zu ermöglichen, wird dazu eine eine der Zugriff über eine vordefinierte Programmierschnittstelle (API) gestattet. Für Anwendungen und Dienste im Web sind die folgenden zwei Ansätze für die Architektur einer solchen API besonders verbreitet.

2.1.1 Representational State Transfer (REST)

Eine sehr beliebte Architektur für den öffentlichen Zugriff auf Webanwendungen ist *Representational State Transfer* (REST) [18, S. 76]. REST baut auf HTTP auf und definiert einige Beschränkungen die eine REST basierter Dienst erfüllen muss.

Die Grundidee besteht darin, dass hinter einer URL eine bestimmte Ressource sich verbirgt auf die man von außen zugreifen möchte. REST schreibt aber nicht vor in welchen Datenformat diese Ressource übermittelt werden soll, sondern dass das zurückgelieferte Format der Ressource änderbar ist.

„REST components communicate by transferring a representation of a resource in a format matching one of an evolving set of standard data types, selected dynamically based on the capabilities or desires of the recipient and the nature of the resource.“[18, S. 87]

Dadurch soll einer einfachere Verwendbarkeit in unterschiedliche Systemen ermöglicht werden. So kann für eine Webanwendung beim aufrufen im Browser eine HTML-Datei zurück geliefert werden, die sofort betrachtet werden kann und falls ein Programm darauf zugreift wird ein maschinenlesbares Format verwendet. Neben HTML sind auch noch XML und JSON sehr verbreitete Formate. Die Kommunikation wird dabei komplett Zustandslos abgehalten und alle Zusatzinformationen müssen immer mitgeliefert werden. Durch die Zustandslosigkeit skaliert das System viel besser, da Ressourcen sofort wieder frei gegeben werden können und nicht für spätere Anfragen gespeichert werden müssen.

Wie schon beschrieben, nutzen REST basierte Dienste HTTP als Grundlage zur Kommunikation. Die dort definierten Operationen werden mit REST zur Auslieferung und Manipulation der

Tabelle 2.1.: Wichtigsten HTTP Operationen mit REST

Operation	Beschreibung
GET	Liefert die hinter einer URL liegende Ressource an den Aufrufer zurück.
POST	Dient zum Anlegen einer neuen Ressource. Die URI der neuen Ressource ist beim Aufruf noch unbekannt und wird von Service bestimmt.
PUT	Wird zum Ändern eine bestehenden Ressource genutzt.
DELETE	löscht, wie der Name schon sagt, eine Ressource dauerhaft.

Ressourcen verwendet. Zur Grundausstattung gehören dabei GET, POST, PUT und DELETE (siehe Tabelle 2.1). Die anderen Operationen HEAD, TRACE, OPTIONS und CONNECT sind eher selten anzutreffen.

2.1.2 Simple Object Access Protocol (SOAP)

Das *Simple Object Access Protocol*[29] (SOAP) ist ein vom W3C standardisiertes Netzwerkprotokoll für den Austausch von Daten zwischen heterogenen Systemen. SOAP schreibt einen bestimmten Aufbau von Nachrichten vor, innerhalb von denen die Daten transportiert werden. Als Repräsentation für diese Nachrichten wird auf XML gesetzt. Bei der Wahl des Transportprotokolls werden dahingegen keine Vorgaben gemacht und es ist frei wählbar. Häufig wird es aber in Verbindung mit HTTP und TCP verwendet.

Listing 2.1: SOAP Nachricht

```
1 <Envelope xmlns="http://www.w3.org/2003/05/soap-envelope">
2   <Header>
3     <!-- header information -->
4   <Header>
5   <Body>
6     <!--body content-->
7   </Body>
8 </Envelope>
```

Eine Nachricht besteht im Grunde aus drei Elementen: den *Envelope*, einen optionalen *Header* und einem *Body* (siehe Listing 2.1). Der Envelope fungiert, wie die Übersetzung schon sagt, als Briefumschlag für die zu transportierenden Daten. Innerhalb jedes Envelopes können zusätzliche Meta-Informationen im Header Element gespeichert werden. Die eigentlichen Daten befinden sich im Body Element des Envelopes. Wie der Inhalt von Header und Body auszusehen haben wird von SOAP nicht vorgeschrieben. Dies können weitere XML Elemente oder einfache Zeichenketten sein.

Gebräuchlich ist der Einsatz von SOAP bei sogenannten *Remote Procedure Calls* (RPC). Unter RPC versteht man den Aufruf einer Funktion von einem entfernten Dienst und das Zurückerliefern einer eventuell vorhandenen Antwort. Welche Funktionen von einem Dienst zur Verfügung stehen wird in einer *Web Services Description Language* [11] (WSDL) Datei beschrieben. Diese WSDL Datei wird in XML Format geschrieben und enthält alle wichtigen Informationen für RPC Aufrufe, die von einem Dienst zur Verfügung gestellt werden:

types enthält Definition von Datentypen die in einer Message eingesetzt werden können. Zur Definition der Datentypen wird das Vokabular von XML Schema¹ eingesetzt.

message Elemente beschreiben die Datentypen aus denen eine Nachricht aufgebaut ist.

portType definiert eine Menge an zur Verfügung stehenden Operationen. Inklusive Eingabe- und Ausgabeparameter. In der WSDL Version 2.0 wurde portType in *interface* umbenannt.

binding beschreibt das Format und den Protokollablauf mehrerer Operationen. Zum Beispiel wie Eingabe- und Ausgabeparameter kodiert werden sollen.

port Definiert eine Adresse hinter der sich ein Binding befindet. Üblicherweise in Form ein URI. Seit WSDL 2.0 wird statt port der Begriff *endpoint* verwendet.

service dient zum Zusammenfassen mehrerer Ports zu einen einzigen Dienst.

Wird eine solche WSDL Datei öffentlich zugänglich gemacht, kann festgestellt werden welche Funktionen ein Dienst anbietet und automatisch APIs für unterschiedliche Systeme generiert werden. Der weitere Datenaustausch erfolgt dann über SOAP Nachrichten.

2.2 Datenintegration

Datenintegration-Einleitung schreiben

2.2.1 Semantic Web

Seit den Anfängen des *World Wide Webs* (kurz als WWW oder Web bezeichnet) hat die Masse an abrufbaren Information immer mehr zugenommen. Die Vorteile des Webs liegen eindeutig in der guten Aktualität und Erreichbarkeit von überall auf Erde. Die Menge an Informationen sind aber auch ein Problem im Web. Da diese überall verteilt sind, ist es schwer für einen einzelnen alleine alles zu einem Thema zu finden. Suchmaschinen wie Google², Yahoo³ oder Microsoft Bing⁴ leisten hier gute Dienste. Doch für Maschinen ist es noch immer nicht einfach die Inhalte von Webseiten zu verstehen, da diese für Menschen gemacht wurden [22]. Auch der Erlernen

¹ <http://www.w3.org/XML/Schema>

² <https://www.google.com>

³ yahoo.com

⁴ <http://www.bing.com/>

von neuen Wissen anhand vorhandener Informationen ist in der aktuellen Form des Webs nur schwer realisierbar.

2001 machte Tim Berners-Lee (der Erfinder des Webs) den Vorschlag [5] das Web mit maschinenlesbaren Informationen zu erweitern und so die Verarbeitung mit Computerprogrammen zu vereinfachen. Die Idee des *Semantic Webs* wurde geboren. Der Inhalt des Webs wird mit semantischen Information so erweitert, dass Programme das sich zwei Texte an unterschiedlichen Stellen des Webs um das selbe Thema handeln. Aber auch die Anzeige von impliziten Wissen, wenn jemand zum Beispiel eine Telefonnummer in Los Angeles, USA sucht und aus Deutschland anrufen will, wird im mitgeteilt dass er wegen der Zeitdifferenz von neun Stunden lieber erst Nachmittags anrufen sollte, wäre so einfacher möglich.

Resource Description Framework

Eine Baustein des Semantic Webs ist das *Resource Description Framework* (RDF). Wie der Name schon suggeriert, dient RDF zur Beschreibung von einzelnen Ressourcen innerhalb des Webs. Nach [26, 27] bestand die Motivation bei der Entwicklung von RDF Information über Ressource in einen offenen Datenmodell zu speichern, so dass diese Daten von Maschinen automatisch verarbeitet, manipulieren und untereinander ausgetauscht werden können. Gleichzeitig sollte es auch einfach von jedem erweitert werden können „RDF is designed to represent information in a minimally constraining, flexible way“[26].

Das Datenmodell von RDF ist zur effizienten Verarbeitung sehr einfach aufgebaut. Die Grundlage bilden Tripel aus Subjekt, Prädikat und Objekt. Einer oder mehrere solcher Triple zusammen werden als gerichteter RDF-Graph bezeichnet. Subjekt und Objekt stehen über das Prädikat mit einander in Beziehung, wobei die Beziehung immer vom Subjekt zum Objekt geht. Das Prädikat wird auch als Eigenschaft (engl. Property) bezeichnet. Gemeinsam beschreibt das Triple immer eine Aussage über eine oder zwei Ressourcen. Zum Beispiel „Die Dose enthält Kekse“ wäre eine Aussage, dass in einer Dose sich Kekse befinden. Die Dose ist dabei das Subjekt, enthält das Prädikat und Kekse das Objekt. Ein Triple ist quasi ein einfacher Satz in der natürlichen Sprache [21]. Für Subjekt, Prädikat und Objekt werden in RDF *Uniform Resource Identifier* (URI), *Literale* oder *leere Knoten* (im englischen *Blank Nodes* genannt) verwendet.

URIs sind eindeutige Bezeichner die eine beliebige reale oder abstrakte Ressource darstellen und werden wie in RFC 2396⁵ beschrieben formatiert. Relative URIs sollten aber nach [26] aber nach Möglichkeit vermieden werden. URIs bilden eine Verallgemeinerung der im Web gebräuchlichen Uniform Resource Locator (URL).

Literale bestehen aus einfachen Zeichenketten die zum Speichern der Informationen dienen. Zusätzlich können Literale mit der Angabe der verwendeten Sprache `Objekt"@de` oder des Datentyps `"42"8sd:integer` erweitert werden. Bei Literalen ist darauf zu achten, dass die Literale `Objekt"` und `Objekt"@de` auf den ersten Blick zwar den selben Wert beschreiben, aber aus Sicht von RDF nicht die selben sind. Sowohl die angegebene Spracht als auch der Datentyp müssen übereinstimmen.

⁵ <http://www.isi.edu/in-notes/rfc2396.txt>

Leere Knoten werden als alle Knoten im RDF Graphen beschrieben, welche weder eine URI noch ein Literal sind. Sie dienen häufig dazu, um Subjekte zu beschreiben für die nicht unbedingt eine eigene URI nötig ist und sind nur innerhalb eines Graphen eindeutig. Für die Referenzierung außerhalb des RDF-Graphen sind leere Knoten ungeeignet.

Doch nicht jeder davon ist in jeden Teil des Tripels erlaubt. Das Subjekt ist entweder eine URI oder ein leerer Knoten, wobei das Prädikat nur eine URI sein kann. Dahingegen ist es beim Objekt möglich eine URI, einen leeren Knoten oder ein Literal zu verwenden.

Repräsentation von RDF-Graphen

In Laufe der Zeit von RDF wurde verschiedene Möglichkeiten erfunden einen RDF-Graphen darzustellen. In diesem Abschnitt werden drei Formen vorgestellt die auch in dieser Arbeit zur Visualisierung benutzt werden.

Graphische Darstellung

Graphisch lässt sich RDF als gerichteter Graph mit Knoten und Kanten darstellen. Ressourcen werden dabei als elliptische Knoten, Literale als Rechtecke und die Prädikate als gerichtete Kante gezeichnet. Ein Beispiel ist in Abbildung 2.1 zu sehen.

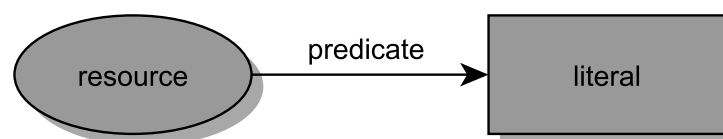


Abbildung 2.1.: Einfacher RDF-Graph

RDF/XML

RDF/XML ist eine verbreitete Form RDF-Dokumente zu beschreiben. Die Basis bildet hierbei die Verwendung der Extensible Markup Language (XML). In Listing 2.2 ist ein Beispieldokument in RDF/XML zu sehen. Das in Zeile 2 zu sehende `rdf:RDF` Element zeigt, dass sich innerhalb von ihm die RDF-Beschreibung des Dokuments befindet. In diesem Element werden mit `xmlns:` einige Präfixe für Namensräume definiert, um das Dokument übersichtlicher zu halten. Alle Präfixe werden danach mit den angegebenen Namensraum ersetzt. Das `Description` in Zeile 5 stellt die Beschreibung einer Ressource im RDF-Graphen. Die URI der Ressource wird mit dem Attribut `rdf:about` definiert. Innerhalb des `Description` Elements befinden sich die Prädikate. In Zeile 6 steht also, dass die Ressource die Eigenschaft `externs:enthaelt` besitzt und diese das Literal `Kekse`. Wäre das Objekt nicht wie hier ein Literal sondern eine weitere Ressource, könnte man über das Attribut `rdf:resource` für das `externs:enthaelt` auf diese Ressource verweisen.

Listing 2.2: RDF/XML Beispiel

```
1 <?xml version="1.0"?>
```

```

2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:externs="http://www.example.org/terms#">
4
5   <rdf:Description rdf:about="http://www.example.org/dose">
6     <externs:enthaelt>Kekse</externs:enthaelt>
7   </rdf:Description>
8 </rdf:RDF>

```

Turtle

Turtle⁶ (Ausgeschrieben: *Terse RDF Triple Language*) ist eine weitere Möglichkeit RDF-Graphen darzustellen und ist eine Gering aus der Sprache N3 (Kurzform für Notation 3) hervor. In Turtle wird das Triple aus Subjekt, Prädikat und Objekt hintereinander geschrieben und zwischen jeden mindestens ein Leerzeichen gelassen. Als Abschluss folgt nach jedem Triple noch ein Punkt. Der Punkt verdeutlicht noch einmal die Ähnlichkeit mit gesprochenen Sätzen. Listing 2.3 zeigt das Beispiel mit der Keksdose noch einmal in Turtle Notation.

Listing 2.3: Turtle Beispiel

```

1 <http://www.example.org/dose> <http://www.example.org/terms#enthaelt>
  "Kekse" .

```

In Turtle ist darauf zu achten, dass alle URIs immer zwischen Spitzengklammern stehen müssen. Literale werden in Anführungszeichen geschrieben. Da nun einzelne Prädikate beziehungsweise allgemein URIs recht häufig innerhalb eines Graphen auftauchen können, kann es einfacher sein diese abzukürzen. Wie schon in RDF/XML können auch in Turtle Präfixe definiert werden um Schreibarbeit zu sparen.

Listing 2.4: Turtle Präfixe

```

1 @prefix externs: <http://www.example.org/terms#> .
2 <http://www.example.org/dose> externs:enthaelt "Kekse" .

```

In der ersten Zeile von Listing 2.4 wird durch Einleiten mit dem Schlüsselwort @prefix ein neuer Präfix externs: für den Namensraum http://www.example.org/terms# festgelegt. Dieser Präfix kann nun überall innerhalb des Dokumentes verwendet werden, wobei die Spitzengklammern dann weggelassen werden können.

Listing 2.5: Turtle abkürzende Schreibweise

```

1 @prefix externs: <http://www.example.org/terms#> .
2 <http://www.example.org/dose> externs:farbe "blau";
3   externs:enthaelt "Kekse", "Geld" .

```

Listing 2.5 zeigt nochmal ein drittes Beispiel, wie redundante Angaben eingespart werden. Wie man in der zweiten Zeile sehen kann, wird keine Frage für die Dose angegeben das Triple aber mit einem Semikolon abschlossen und nicht mit einem Punkt. Durch das Semikolon ist es möglich das

⁶ Turtle Spezifikation: <http://www.w3.org/TeamSubmission/turtle/>

Subjekt mehrfach wieder zu verwenden, wenn sich nur Prädikat und Objekt ändern. So können sich mehrere Eigenschaften einer Ressource platzsparend schreiben ohne das Subjekt immer wieder anzugeben. Ändert sich dahingegen nur das Objekt können mehrere durch Kommata getrennt hintereinander geschrieben werden. Die dritte Zeile beschreibt zum Beispiel, dass in der Dose nicht nur Kekse sondern auch Geld steckt. Leere Knoten können dann noch in Turtle durch angeben einer geöffneten eckigen Klammer gefolgt von einer sich Schließenden dargestellt „[]“. Soll ein leerer Knoten innerhalb eines Graphen referenziert werden, kann er auch als `_:LABEL`, wobei LABEL ein beliebiger Beizeichner ist, geschrieben werden.

2.2.2 Ontologien

Sollen Daten aus verschiedenen Quellen zusammengefügt werden, stellt sich häufig das Problem dass Teile dieser Daten zwar den gleichen Sinn haben, aber aufgrund der Sichtweise des jeweiligen System eine andere Bezeichnung besitzen. Das kann zum Beispiel zu Missverständnissen bei der Verarbeitung führen oder dass Teile eines anderen Systems nicht wiederverwendet werden können [36]. Einen Ausweg aus diesem Dilemma kann die Verwendung von Ontologien zeigen. Ontologien können allgemein als Wissensbasis [36, 22] bezeichnet werden und liefern eine formale Spezifikation über eine bestimmte Interessensdomäne. Sie beschreibt nicht nur wie das verwendete Vokabular aussieht, sondern legt auch fest welche einheitliche Bedeutung jede Vokabel hat.

Im Bereich des Semantic Webs sind heutzutage zwei Sprachen für die Erstellung von Ontologien weit verbreitet. Diese sind *RDF Schema* (RDFS)[9] und die darauf aufbauende *Web Ontology Language* (OWL)[31]. Beide Sprachen basieren auf RDF, so können sie zusammen mit jedem System verwendet werden, das RDF versteht. Mit ihnen ist man im Stande Klassen von abstrakten Objekten und deren Eigenschaften zu definieren und diese in eine Vererbungshierarchie einzugliedern. Im Gegensatz zu RDFS können in OWL diverse Einschränkungen definiert werden, wie zum Beispiel dass eine Eigenschaft nur einmal pro Objekt einer Klasse vorhanden sein darf. Die Anhänge A.1 und A.2 zeigen zwei Ontologien in der Sprache OWL, welche innerhalb dieser Arbeit entwickelt wurden und in Abschnitt 4.2.1 und 4.2.4 beschreiben werden.

überleitung zu FOAF und SIOC

Friend of a Friend (FOAF)

*Friend of a Friend*⁷ (FOAF) ist ein 2000 gestartetes Projekt und versucht Personen innerhalb des Webs, inklusive der Verbindungen zwischen ihnen und anderen, sowie dem was sie machen, in maschinenlesbarer Form abzubilden. FOAF stellt hierzu ein Vokabular [10] auf der Basis von RDF für solche sozialen Netzwerke zur Verfügung. Das Vokabular von FOAF gliedert sich dazu in einen „FOAF Core“ und einen „Social Web“ Bereich. Der Core-Bereich die Klasse *Agent* für alle Dinge die eine Handlung ausführen können, also sowohl natürliche Personen, Gruppen oder Organisationen als auch Computerprogramme oder Maschinen. Für diese gibt es jeweils noch einzelne Unterklassen die von der Klasse *Agent* erben. Objekten dieser Klassen können eigene Eigenschaften wie einen Namen, ein Alter oder wen sie kennen gegeben werden. Der Social Web

⁷ <http://www.foaf-project.org>

Bereich enthält alle Teile die für das Web interessant wären. Das wären zum Beispiel welche E-Mail-Adresse eine Person besitzt, welche Benutzerkonten im bei welcher Webseite gehören oder wie die URL seiner Homepage lautet. Das FOAF Projekt sieht sich aber selber nicht als Konkurrenz gegenüber den etablierten sozialen Online-Netzwerken, sondern eher ein Ansatz für einen besser Austausch zwischen den einzelnen Seiten [10, Abstract].

Listing 2.6 zeigt ein Beispiel FOAF-Dokument in RDF/XML. Es beschreibt die fiktive Person „Max Mustermann“, mit Vor- und Nachname sowie der Hashwert seiner E-Mail-Adresse (foaf:mbox_sha1sum). Diese Person kennt (foaf:knows) kennt eine Person mit dem Namen „John Doe“ der ebenso eine E-Mail-Adresse mit den angegebenen Hashwert besitzt. Statt die Eigenschaften von John Doe hier noch einmal vollständig anzugeben, wird über die Eigenschaft rdfs:seeAlso auf ein weiteres FOAF-Dokument verwiesen, dass die fehlenden Daten enthält.

Listing 2.6: FOAF Beispiel

```
1 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
2     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
3     xmlns:foaf="http://xmlns.com/foaf/0.1/">
4   <foaf:Person>
5     <foaf:name>Max Mustermann</foaf:name>
6     <foaf:firstName>Max</foaf:firstName>
7     <foaf:surname>Mustermann</foaf:surname>
8     <foaf:mbox_sha1sum>dce4fc922158f8b26fbf0a65ea32bfab58488bd2</
foaf:mbox_sha1sum>
9     <foaf:knows>
10      <foaf:Person>
11        <foaf:name>John Doe</foaf:name>
12        <foaf:mbox_sha1sum>479ea35d3522662b70dc7afd721853485c95db57</
foaf:mbox_sha1sum>
13        <rdfs:seeAlso rdf:resource="http://www.example.org/people/jd/
foaf.rdf"/>
14      </foaf:Person>
15    </foaf:knows>
16  </foaf:Person>
17 </rdf:RDF>
```

Semantically-Interlinked Online Communities (SIOC)

Semantically-Interlinked Online Communities (SIOC, ausgesprochen „schock“) ist ein Projekt, welches von Uldis Bojārs und John Breslin begonnen wurde um unterschiedliche, webbasierte Diskussionsplattformen (Blog, Forum, Mailinglist, ...) untereinander verbinden zu können [13, 8, 7]. Der Kern von SIOC besteht aus einer Ontologie, welche den Inhalt und die Struktur dieser Plattformen in ein maschinenlesbares Format bringt und es erlaubt diese auf semantischer Ebene zu verbinden. Auch soll es so möglich sein Daten von einer Plattform zu einer Anderen zu transferieren und so einfacher Inhalte austauschen zu können. Als Basis für SIOC dient RDF, die Ontologie selber wurde in RDFS und OWL designet. Um nicht das Rad neu erfinden zu müssen

greift SIOC auf schon bestehende und bewährte Ontologien zurück. Für die Abbildung von Beziehungen zwischen einzelnen Personen wird FOAF und für einige Inhaltliche- und Metadaten (Titel, Inhalt, Erstelldatum, ...) Dublin Core Terms⁸ eingesetzt.

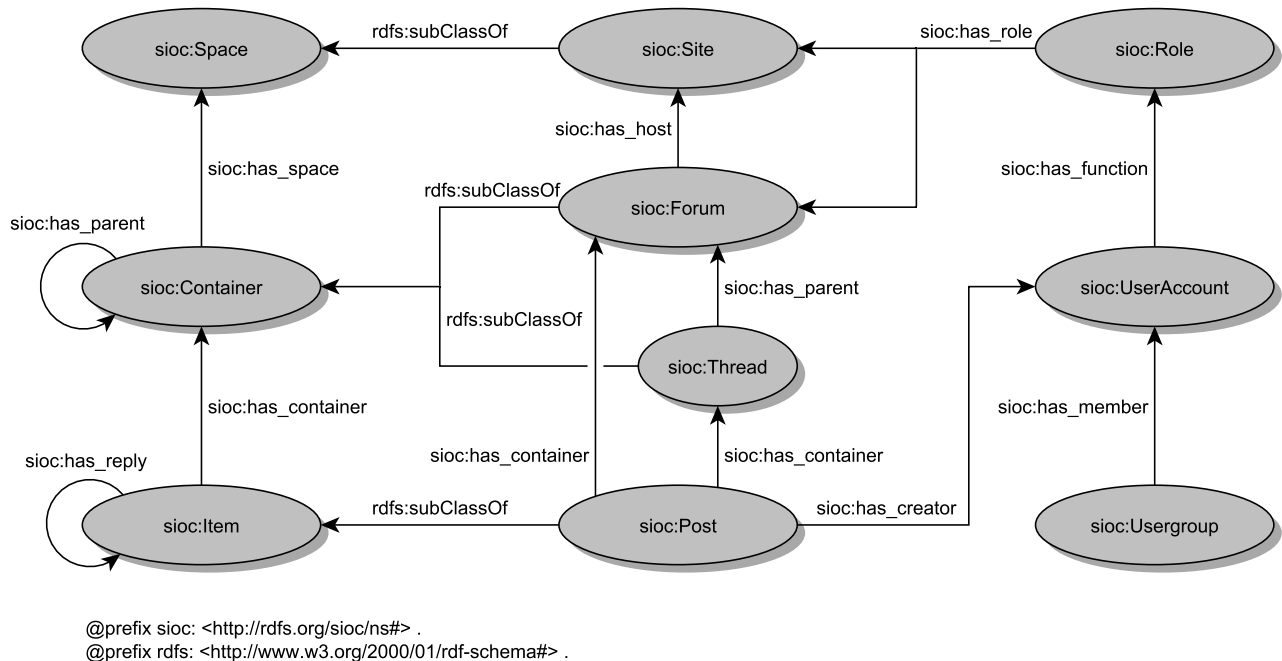


Abbildung 2.2.: Aufbau von SIOC (modifiziert) - Originalquelle: [13]

Die wichtigsten Klassen von SIOC sind in Abbildung 2.2 in der mittleren Spalte zu sehen. Die Klasse *Site* ist für die Beschreibung von allgemeinen Webseiten in denen Beiträge innerhalb von Containern verfasst werden. Ein solcher Container ist die Klasse *Forum* und steht für einen Ort an dem Diskussionen geführt werden. Enthält ein Forum unterschiedliche Diskussionen zu unterschiedlichen Themen, kann es noch einmal in unterschiedliche *Thread* unterteilt werden, welche immer ein Forum als Elternteil haben (*has_parent*). Beide Klassen leiten sich von der Klasse *Container* als einen allgemeinen Ort für Beiträge ab. Die einzelnen Beiträge werden durch die Klasse *Post* beziehungsweise von der übergeordneten Klasse *Item* modelliert. Beiträge gehören in der Regel immer zu einem bestimmten Container oder mindestens zu einer Webseite. Es ist auch möglich Beiträge als Kommentar zu anderen Beiträgen über die Eigenschaft *has_reply* abzubilden. Jeder Beitrag besitzt mindestens einen Autor der ein Benutzerkonto auf der betreffenden Seite besitzt. Für die Beschreibung eines solchen Benutzerkontos wird die Klasse *UserAccount* verwendet. Dieses Benutzerkonto kann nun zu einer Gruppe von anderen Konten gehören, zum Beispiel einer Lerngruppe. Über die Klasse *Role* kann ein Benutzerkonto eine bestimmte Rolle innerhalb einer Seite, eines Forums und so weiter zugeteilt werden. Ein Beispiel dafür wäre die Rolle eines Moderators, der überwacht ob die Regeln der Seite in den einzelnen Foren eingehalten werden.

⁸ <http://dublincore.org/documents/dcmi-terms>

2.3 Datenverteilung

Datenverteilung-Einleitung schreiben

2.3.1 Java Messaging Service

Das *Java Messaging Service* (JMS) ist eine Sammlung von Schnittstellen für das Erstellen, Senden und Empfangen von Nachrichten zwischen Clients [34]. JMS erlaubt eine Entwicklung von verteilten Anwendungen, die nicht nur lose gekoppelt, sondern auch asynchron und zuverlässig arbeiten.

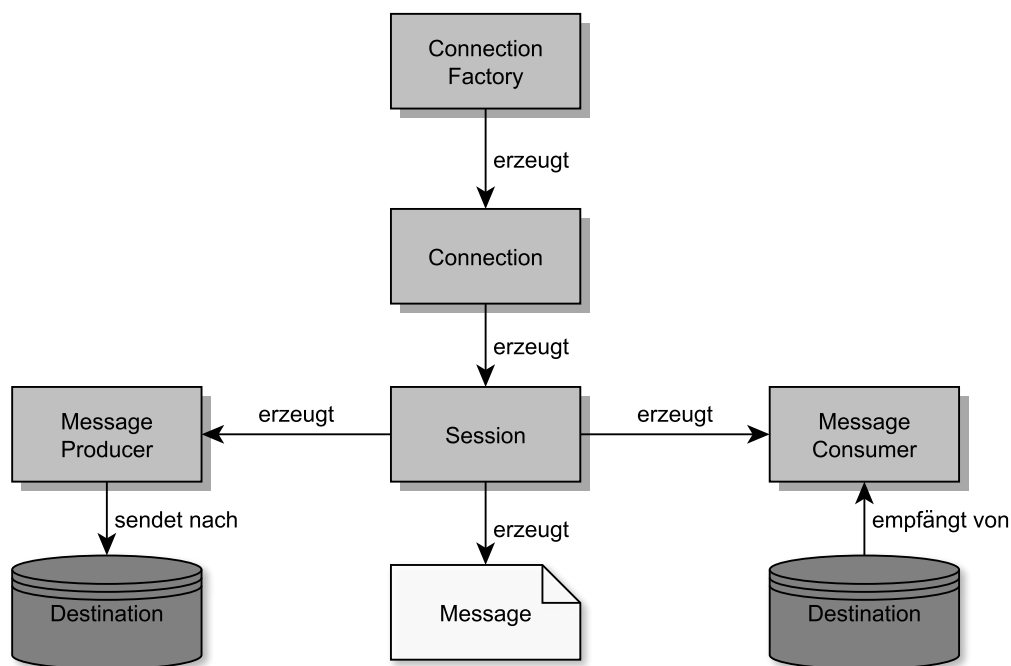


Abbildung 2.3.: JMS Programmiermodell - Original Bild: http://docs.oracle.com/javaee/1.3/jms/tutorial/1_3_1-fcs/doc/images/Fig3.1.gif (Zugriff: 2013-09-15)

Die Grundstruktur einer JMS Anwendung besteht aus einem *JMS Provider*, der die Schnittstellen von JMS implementiert. Dieser JMS Provider wird auch als *Message Oriented Middleware* (MOM) bezeichnet und kümmert sich auch darum, dass Nachrichten zuverlässig verschickt werden. Einen JMS Client von dem Nachrichten verschickt und empfangen werden. Den Nachrichten selber und den sogenannten *Administered Objects*. Diese bestehen aus vorkonfigurierten *ConnectionFactory*s für das Erstellen von Verbindungen (*Connections*) zwischen Client und MOM und *Destinations* als Sende- und Empfangspunkte von Nachrichten. Die Administered Objects können im Client über das Java Naming and Directory Interface (JNDI)⁹ API abgefragt werden. Alle Clients, die nicht die JMS API sondern die implementierte API der MOM direkt verwenden, werden *Native Client* genannt.

⁹ JNDI API: <http://www.oracle.com/technetwork/java/index-jsp-137536.html>

JMS unterstützt zwei Verbindungsarten zum Übertragen von Nachrichten: Queue-basiert und Topic-basiert. Als Queue-basiert wird eine Punkt-zu-Punkt Verbindung bezeichnet. Hier werden Nachrichten nur zwischen zwei Clients übertragen und gegebenenfalls in einer Warteschlange zwischengespeichert. Hinter Topic-basiert verbirgt sich ein Publish-Subscribe Mechanismus bei dem ein Client Nachrichten an eine bestimmtes Topic-Destination schickt und andere Clients sich auf dieses Topic anmelden können, die dann die Nachrichten des ersten Clients zugeschickt bekommen. Ob Queue oder Topic-basiert, wird über das verwendete Destination Objekt ausgewählt.

Sollen nun Nachrichten von einem Client verschickt beziehungsweise empfangen werden, muss mit einer Connection Factory eine neue Connection zu einer MOM aufgebaut werden. Mit dieser Connection wird danach ein *Session*-Objekt erstellt, das als Kontext zum Senden und Empfangen verwendet wird. Sollen Nachrichten gesendet werden, muss mit einem Destination-Objekt ein *MessageProducer* und danach eine Nachricht mit dem Session-Objekt erstellt werden. Die Nachricht wird dann über den MessageProducer versendet. Für das Empfangen ist ein *MessageConsumer* verantwortlich. Abbildung 2.3 zeigt noch einmal den Zusammenhang aller JMS Komponenten.

Listing 2.7: JMS Beispiel

```
1 Context ctx = new InitialContext();
2 ConnectionFactory connectionFactory = (ConnectionFactory) ctx.lookup(
   "ConnectionFactory");
3
4 Connection connection = connectionFactory.createConnection();
5 connection.start();
6
7 Session session = connection.createSession();
8 Destination destination = session.createTopic("topic-test");
9
10 MessageProducer msgProducer = session.createProducer(destination);
11 Message msg = session.createTextMessage("Hallo World!");
12 msgProducer.send(msg);
```



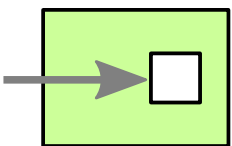
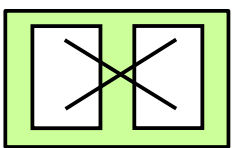
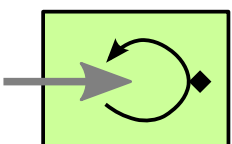
Listing 2.7 zeigt ein kleines Beispiel zum Senden einer Textnachricht mit JMS. Die erste und zweite Zeile zeigt wie ein JNDI Kontext erstellt und nach einer vordefinierten Connection Factory mit dem Namen „ConnectionFactory“ gesucht wird. Mit dieser Connection Factory wird dann eine neue Connection erstellt und gestartet. Danach folgt in Zeile 7 und 8 das Erstellen einer neuen Session und die Definition eines Topics mit dem Namen „topic-test“ als Destination. In der zehnten Zeile wird dann der MessageProducer zum Senden von Nachrichten und in der Folgezeile die zu sendende Textnachricht erzeugt. Diese wird dann in der letzten Zeile an das Topic verschickt.

2.3.2 Enterprise Integration Pattern (EIP)

Bezeichnungen wie „Iterator“, „Factory Method“, „Observer“ oder „Proxy“ hat bestimmt schon jeder Programmierer mindestens einmal gehört. Hierbei handelt es sich um sogenannte Entwurfsmuster für Softwareprogramme. Sie sind Schablonen für Lösungen von Problemen, die in

der Entwicklung von Software immer wieder auftreten und sich als hilfreich erwiesen haben. Auch in der Integration von verschiedenen Geschäftsanwendungen in ein System treten solche Muster immer wieder auf. Gregor Hohpe und Bobby Woolf beschreiben in ihrem Buch „Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions“[23] eine Vielzahl solcher Enterprise Integration Patterns (EIP) für die Integration mit MOM. Alle hier aufzuzählen würde den Rahmen dieser Arbeit sprengen. Aus diesem Grund werden in Tabelle 2.2 fünf Muster inklusive der verwendeten Symbole vorgestellt die später noch vorkommen werden. Die restlichen Muster sind auf der Webseite von EIP¹⁰ zu finden.

Tabelle 2.2.: Einige Beispiel von EIP

Icon	Name	Beschreibung
	<i>Message</i>	Über Nachrichten werden Daten zwischen zwei oder mehr Systemen ausgetauscht.
	<i>Channel</i>	Ein Channel beschreibt einen Nachrichtenkanal über dem Nachrichten von einem Systemen in ein anderes verschickt werden können.
	<i>Endpoint</i>	Endpoints sind Schnittstellen in einem System, von dem Nachrichten in einen Kanal gesendet oder von da Empfangen werden.
	<i>Message Translator</i>	Nicht immer liegt eine Nachricht im richtigen Format für ein System vor. Durch Message Translators können diese in das gewünschte Format übersetzt werden.
	<i>Polling Consumer</i>	

Apache Camel

Apache Camel[2] (kurz: Camel) ist ein Projekt der Apache Software Foundation (kurz: Apache) für das Routen und Verteilen von Nachrichten zur Integration von System auf der Basis von definierten Regeln. Camel stellt dazu eine Java basierte API für den Einsatz der EIP bereit. Die Regeln für die Routen, die Nachrichten nehmen können, können direkt in Java aber auch durch das Spring Framework¹¹ in XML definiert werden.

Listing 2.8: Apache Camel Beispiel

```

1 RouteDefinition rd = new RouteDefinition()
2   .from('timer://helloTimer?period=3000')
```

¹⁰ <http://www.enterpriseintegrationpatterns.com/toc.html>

¹¹ <http://www.springsource.org/>

```
3      .to('log:helloLog');
4
5 CamelContext camelContext = new DefaultCamelContext();
6 camelContext.addRouteDefinition( rd );
7 camelContext.start();
```

Listing 2.8 zeigt ein kleines Beispiel, wie eine Nachrichtenroute in Java definiert werden kann. In der ersten Zeile wird über die Klasse `RouteDefinition` eine neue Route erstellt. Über die Methode `from` wird der Endpunkt, von dem die Route ausgeht, festgelegt. Welcher Endpunkt das genau sein soll, kann auf zwei Arten definiert werden. Man übergibt der Methode direkt ein Objekt einer Klasse, die das Interface `Endpoint` implementiert, oder man macht es, wie hier im Beispiel, über eine URI. Die URI baut sich auf folgende Weise zusammen. Der Teil bis zum Doppelpunkt, das sogenannte „Schema“, legt die Komponente (engl. *Component*) fest, von der ein Endpunkt erzeugt werden soll. In diesem Fall ist es eine Timer-Komponente, deren Endpunkt im periodischen Abstand Event-Nachrichten verschickt. Der Rest der URI wird zur Konfiguration an den Endpunkt übergeben. „helloTimer“ steht hier für einen Namen für den Timer und der Parameter „period“ gibt den zeitlichen Abstand zwischen zwei Event-Nachrichten an. Das Ziel der Route wird mit der Methode `to` in Zeile Drei festgelegt. Für das Ziel wird hier eine Log-Komponente mit dem Namen „helloLog“ festgelegt, die alle reinkommenden Nachrichten protokolliert. In der fünften Zeile wird ein `CamelContext` Objekt, das für die Verwaltung und Ausführung der Routen verantwortlich ist. Die eben erstellte Route wird dann dem `CamelContext` hinzugefügt und die Ausführung in der letzten Zeile gestartet. Nun wird der Timer alle 3000 Millisekunden eine neue Event-Nachricht erzeugen und an den `CamelContext` schicken. Dieser leitet die Nachricht dann an das durch die Route definierte Ziel und wird dort von der Log-Komponente

Für Camel existieren schon eine Menge vorgefertigter Komponenten¹², die ein großes Spektrum an Anwendungsmöglichkeiten abdeckt. Es gibt Komponenten für die Anbindung an HTTP-Webserver, JMS-Provider, E-Mail-Server, RSS/Atom-Feeds und viele mehr.

2.4 Lernplattformen und soziale Online-Netzwerke

An dieser Stelle sollen noch kurz einige Lernplattformen und soziale Online-Netzwerke vorgestellt, die im späteren Verlauf dieser Arbeit für die Implementierung verwendet wurden. Im Einzelnen waren dies *Moodle*, *Canvas*, *Youtube*, *Facebook* und *Google+*, da sie einen guten Schnitt von den Plattformen bilden, die heutzutage sowohl im Bereich E-Learning als auch von der breiten Masse verwendet werden.

2.4.1 Moodle

Moodle¹³ ist ein weit verbreitetes Open Source Online LMS. Die Hauptaufgabe liegt im Verwalten von online Kursen im Bereich E-Learning. Hierzu bietet Moodle von Haus aus eine große Menge an Funktionen für die Verwaltung des Kurses und die Kommunikation zwischen Lehrenden und

¹² <http://camel.apache.org/components.html>

¹³ <https://moodle.org/>

Lernenden. Es bietet die Möglichkeit Aufgaben die Teilnehmern zu verteilen, Fragebögen zu erstellen, zusätzlichen Kursmaterialien bereitzustellen und den Lernerfolg durch Benotung und Feedback zu kontrollieren. Funktionen für die Unterstützung des kollaborativen Lernens sind ebenfalls vorhanden. Teilnehmer können Lerngruppen bilden, sich über persönliche Nachrichten austauschen, gemeinsam an Wikis arbeiten oder in Foren diskutieren.

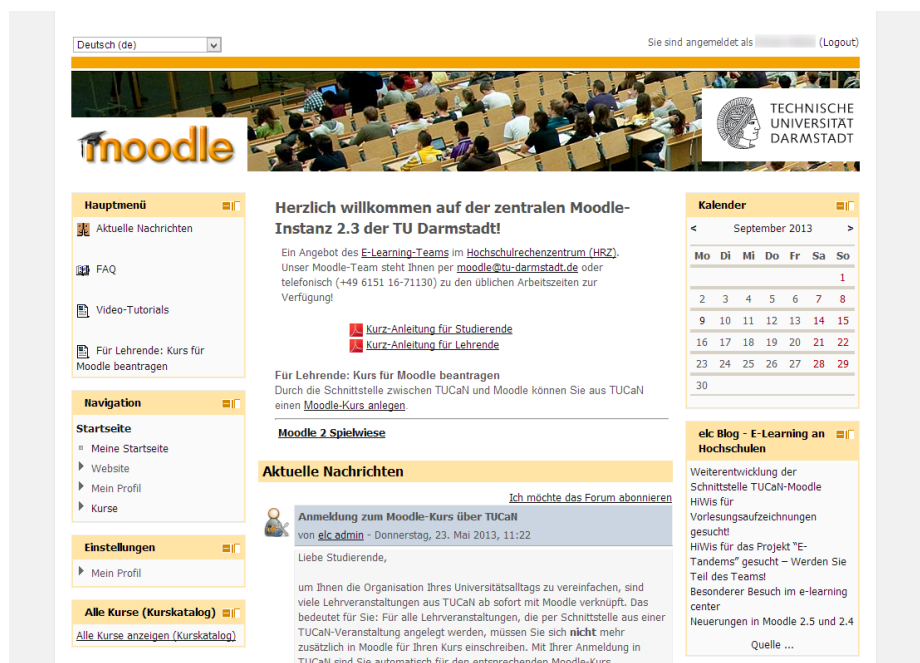


Abbildung 2.4.: Moodleinstanz der TU Darmstadt

Moodle wurde in der Programmiersprache PHP geschrieben und unterstützt die Datenbanken MySQL, PostgreSQL, MSSQL und Oracle. Die Installation von weiteren Funktionalitäten ist durch von Dritten geschriebenen Erweiterungen möglich. Seit Version 2.0 können für Moodle auch Webservices installiert werden, so können auch externe Anwendungen auf interne Funktionen und Daten zugreifen.

2.4.2 Canvas

Das von der Firma Instructure¹⁴ entwickelte *Canvas* ist ein unter Open Source Lizenz gestelltes LMS. Vom Funktionsumfang ist es Moodle nicht unähnlich. Es existiert eine Verwaltung einzelner Kurse. Innerhalb dieser Kurse können in einem Forum Diskussionen geführt und Lernmaterialien hoch- und heruntergeladen werden. Verteilung von Aufgaben, deren Benotung und ein Benachrichtigungssystem existiert ebenfalls. Canvas erlaubt auch das Einbinden von externen Diensten zum kollaborativen Lernen und Arbeiten wie Google Docs¹⁵ oder der Webkonferenz Anwendung BigBlueButton¹⁶.

¹⁴ <https://www.instructure.com/>

¹⁵ <https://drive.google.com>

¹⁶ <http://www.bigbluebutton.org>



Abbildung 2.5.: Instructure Canvas

Canvas wird mittels des Webframeworks *Ruby on Rails*¹⁷ entwickelt. Das Aussehen ist etwas moderner, als das von Moodle und es wird sehr stark auf die neuesten Webtechnologien wie HTML5 CSS3 und JQuery gesetzt. Eine Erweiterung der Funktionalität von Canvas ist durch das einbinden von Programmen möglich, die den *Learning Tools Interoperability*TM (LTI) Standard erfüllen. Einige solcher Programme finden sich auf der Webseite <https://www.edu-apps.org>. Unter anderem Programme zum Suchen und Einbinden von Youtube Videos, Wikipedia Artikeln, GitHub Gists¹⁸ und vielen weiteren.

2.4.3 Youtube

Die Youtube¹⁹ Webseite gehört wohl heute zu den beliebtesten Anlaufpunkten im Internet, wenn es um das Thema Videos geht. Monatlich nutzen über 1 Milliarde Nutzer die Seite und pro Minute werden 100 Stunden neuer Videos hochgeladen [38]. Doch nicht das komplette Videomaterial besteht aus Katzen, Musik oder Videos von Unfällen. Ein Teil der Benutzer die eigene Videos hochladen, wollen anderen Dinge beibringen, weil es sie schon immer interessierte oder früher selber Probleme damit hatten. Einer erklärt die Logarithmengesetze, ein anderer wie man Feuer ohne Feuerzeug macht und eine ganz andere gibt Schönheitstipps. Youtube ist also auch im E-Learning Bereich gut einsetzbar. Lehrende können eigene Videos hochladen, von anderen interessante Videos in Playlisten zusammenfassen und die Lernenden können über Kommentare Fragen zum Inhalt stellen.

¹⁷ <http://rubyonrails.org/>

¹⁸ <https://gist.github.com>

¹⁹ <https://www.youtube.com>

2.4.4 Facebook

Das soziale Online-Netzwerk Facebook²⁰ kann mit rund 699 Millionen aktiven Benutzern täglich [16] zu den aktuell beliebtesten Vertretern seiner Art bezeichnet werden. Facebook erlaubt es, wie alle sozialen Online-Netzwerke, bekannte Personen in Freundeslisten zusammenzufassen und mit ihnen private Nachrichten auszutauschen. Beiträge wie Texte, Fotos oder Videos können auf einer Art Pinnwand der „Wall“ öffentlich oder nur mit Freunden geteilt werden. Benutzer mit gemeinsamen Interessen können dazu eigene Gruppen bilden und dort auf einer eigenen Wall Beiträge veröffentlichen oder die anderer kommentieren. Wie in der Einleitung schon erklärt zeigt Qiyun Wang et. al. [37] das sich Facebook, wenn auch mit Einschränkungen, wunderbar zur Verwaltung und Nutzung durch Lernkurse und Lerngruppen eignet. Die gleiche Erfahrung teilte Anthony Fontana [19, 17], der Facebook als Alternative zum bestehenden System der Bowling Green State University in Ohio, USA verwendete.

2.4.5 Google+

Google+²¹ ist ein 2011 von Google gestartetes soziales Online-Netzwerk. Seit Anfang 2013 ist Google+, von der Anzahl der aktiven Benutzer her gesehen, auf Platz 2 hinter Marktführer Facebook [35]. Vom Funktionsumfang sind sich beide sehr ähnlich. Auf Google+ können andere Benutzer in sogenannten „Circles“ sortiert werden. Dies entspricht ungefähr den auf Facebook genutzten Freundeslisten. Jeder Benutzer hat einen eigenen „Stream“ in dem er Beiträge öffentlich oder nur für ein oder mehrere Circles verfassen kann. Das Gründen von Gruppen für bestimmte Interessensbereiche ist auch in Google+ möglich und werden dort als „Communities“ bezeichnet. Eines der interessantesten Funktionen von Google+ dürfte die Einführung von „Google Hangout“ sein. Hier können Benutzer neben Chats auch Videokonferenzen mit bis zu zehn anderen abhalten, ohne einen externen Service wie Skype²² zu nutzen. Diese Funktion wäre gut für den Einsatz in E-Learning nutzbar. Ein Tutor könnte so in kleiner Runde Fragestunden abhalten oder Gruppen Treffen abhalten.

2.5 Verwandte Arbeiten und Projekte

Verwandte Arbeiten Einleitung schreiben

2.5.1 What happens when Facebook is gone?

Frank McCown und Michael L. Nelson beschreiben in ihrem Bericht „What happens when Facebook is gone?“ [28], wie Möglichkeiten aussehen können, die unsere Daten von sozialen Online-Netzwerken (hier im speziellen Fall von Facebook) für uns und die Nachwelt archivieren können. Zum Beispiel, wenn eine Person einen großen Teil seines persönlichen Lebens auf

²⁰ <https://www.facebook.com/>

²¹ <https://plus.google.com>

²² <http://www.skype.com/>

Facebook verbringt und plötzlich stirbt. Wie sollen seine Angehörigen an nicht öffentliche Texte, Bilder, Videos heran kommen, wenn sie in der Regel keinen Zugriff auf das Benutzerkonto haben, da der Verstorbene so etwas nicht vorhersehen konnte. Oder wenn ein Benutzer mit seinen Daten in ein anderes soziales Online-Netzwerk umziehen will, sei dies bei Facebook zum damaligen Zeitpunkt nur schwer möglich.

„It is also likely he was not prepared to die at such a young age, and much of his personal life, which lies in the digital “cloud“, may never be accessible to his loved ones“ [28, S. 251]

Zum Anlegen eines solchen Archivs wurden mehrere Ansätze vorgestellt. Die einfachste Ansatz wäre die E-Mail-Benachrichtigung zu aktivieren und alle neuen Beiträge in einem E-Mail-Postfach zu sichern. So können aber nur neuen alle Beiträge erfasst werden, alte bleiben weiterhin in Facebook. Eine sehr aufwändige Möglichkeit wäre es Bildschirmfotos von den Beiträgen zu machen und diese durch ein Texterkennungsprogramm laufen zu lassen. Die dadurch erzeugten Dateien können dann in einer Datenbank gespeichert werden. Heutige Internetbrowser zusätzlich zum Anzeigen von Webseite auch der Herunterladen selbiger an. Dabei wird die HTML-Datei inklusive aller darin enthaltenen weiteren Dateien wie Bilder, Videos und CSS-Dateien gespeichert. Die so archivierte Seite hat dann im beschränkten Umfang genau das gleiche Aussehen und Verhalten wie die original Seite. Ebenfalls wäre eine Nutzung der von Facebook bereitgestellten API für Anwendungen eine Überlegung wert. 2009 war diese API noch sehr eingeschränkt. Gerade der Zugriff auf Beiträge und private Nachrichten war nicht möglich [28, S. 253, Table 1]. Für die Implementierung eines Beispiel Programms wurde ein fünfter Ansatz gewählt. Über einen sogenannten Webcrawler oder eine Erweiterung für den Browser werden relevante Seiten automatisch heruntergeladen und in einen Archiv abgelegt. Dynamische Inhalte sollen kein Problem darstellen, da Seite erst heruntergeladen wird, wenn alle Aufrufe dynamischer Funktionen abgeschlossen ist. Die archivierten Dateien können dann mittels Datamining Techniken verarbeitet und als Atom/RSS Feed²³ bereitgestellt werden.

2.5.2 Reclaim Social

Hat sich nicht jeder schon einmal vor den Rechner gesessen um, zum Beispiel, nach einen Bild gesucht das man irgendwann auf irgendeinem der unzähligen sozialen Netzwerke hochgeladen hat, einem aber partout nicht einfallen will wo? Wann und wo habe ich den Beitrag geschrieben, der perfekt zu meiner aktuellen Arbeit passen würde? Solche oder ähnliche Fragen wurden sicherlich schon mehrere Millionen mal von verschiedenen Menschen in der Welt des Internets gestellt. Wer hätte in so einen Fall nicht gerne alles was man über die letzten Jahre an verschiedenen Stellen im Netz geschrieben, hochgeladen oder als für ihn wichtig markiert hat zentral gespeichert um es durchsuchen zu können? Genau diesem Thema haben sich Sascha Lobo und Felix Schwenzel angenommen und auf der Netzkonferenz re:publica²⁴ 2013 ihr gestartetes Projekt „Reclaim Social“ [33] vorgestellt.

Das klingt ein wenig wie ein Werbetext ;-)

²³ <http://www.rssboard.org/rss-specification>

²⁴ <http://re-publica.de/>

Ziel mit diesem Projektes soziale Medien aus allen möglichen Quellen auf seinen eigenen Blog zu spiegeln und so einen zentrale Anlaufstelle für seine eigenen Inhalte schaffen. Aufbauend auf der weit verbreiteten Blogsoftware „WordPress“²⁵ und der dafür vorhandenen Erweiterung „FeedWordPress“²⁶. Diese Kombination ermöglicht alle Internetseiten, welche einen RSS Feed anbieten, in die Datenbank von WordPress zu spiegeln. Das Problem hierbei besteht darin, dass einige sehr beliebte Internetseiten solche RSS Feeds nicht anbieten (Facebook, Google+) oder eingestellt haben (<https://twitter.com>). Für einige solcher Seiten wurden „proxy-scripte“ [33, Tech Specs Details] implementiert, welche für diese einen RSS Feed emulieren. Zugleich können in den Feeds enthaltende Medien, wie Bilder und Videos(bisher nur als Referenz), heruntergeladen und in WordPress gespeichert werden. So ist es möglich alle gespiegelten Daten einfach zu durchsuchen oder nach bestimmten Kriterien zu filtern. Zusätzlich können alle Freunde, welche auch Reclaim Social einsetzen, in einen Kontaktliste eingetragen und so auch deren Inhalte eingebunden werden.

Aktuell befindet sich dieses Projekt noch im Alpha Stadium und die Installation ist relativ kompliziert. Es ist aber geplant eine eigene Erweiterung für WordPress zu schreiben „he goal is to build just one Reclaim Social-plugin for any wordpress user“ [33, How Does It Work]

²⁵ <http://wordpress.org/>

²⁶ <http://feedwordpress.radgeek.com/>

3 Analyse

Aufteilen in Szenarien: Datenaustausch, Übertragung...

Bevor nun ein System entwickelt werden kann, das die Synchronisation von Diskussionen auf Lernplattformen und sozialen Online-Netzwerken ermöglicht, muss erst analysiert werden, welche Aufgaben das System erfüllen muss und welche Techniken dafür geeignet sind. Alle diese Schritte sollen im folgenden Abschnitt an einen kleinen Beispiel gezeigt werden.

3.1 Ein Beispiel

Am Anfang muss jemand auf einer Webseite einen Beitrag schreiben der synchronisiert werden kann. Zum Beispiel geht ein Student in das Forum seiner Veranstaltung auf der Webseite A und in den Thread zur aktuellen Übung, weil er zu ihr eine Frage hat. Er schreibt also einen neuen Beitrag in das Eingabefeld und schickt es ab. Dieser wird dann an die Webseite A geschickt, im dortigen Format (Format A) in eine Datenbank oder Ähnliches geschrieben und als neuer Beitrag im Thread angezeigt (siehe Abbildung 3.1). Da eine andere Webseite B das Format von Webseite A in der Regel nicht versteht, muss in einen nächsten Schritt der Beitrag aus der Datenbank gelesen und auf irgendeine Art konvertiert werden.

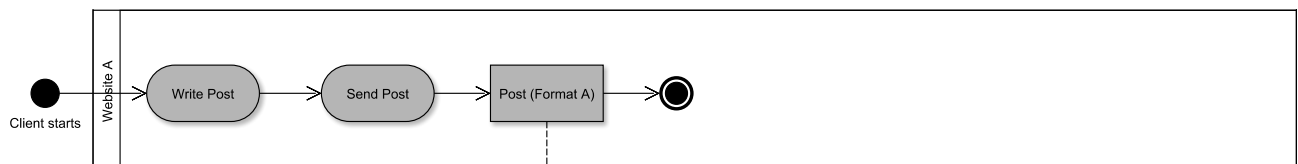


Abbildung 3.1.: Benutzer erstellt einen Beitrag im sozialen Netzwerk A.

3.1.1 Beiträge lesen und konvertieren

Abbildung 3.3 zeigt den Ablauf, wie Beiträge von der Webseite A gelesen, konvertiert und weiterverarbeitet werden. Dazu müssen zuerst die Daten über eine öffentliche API vom Server der Webseite A heruntergeladen werden. Da im Allgemeinen nicht automatisch bekannt ist, wann ein neuer Beitrag vorhanden ist, müssen die Server in zeitlichen Abständen abgefragt (*Polling* genannt) und die zurückgelieferten Daten nach neuen Beiträgen durchsucht werden. Sind ein oder mehrere neue Beiträge gefunden worden, können diese nicht direkt an die Webseite B geschickt werden, da sich diese in der Regel im verwendeten Datenformat unterscheiden. Diese müssen zuvor konvertiert werden.

Die einfachste Möglichkeit wäre die Daten von Webseite A, die in Format A vorliegen, in das Format B von Webseite B zu konvertieren. Bei zwei Formaten ist dies noch sehr einfach. Es müsste

lediglich ein Konverter von Format A nach Format B und einer in die umgekehrte Richtung implementiert werden. Für den Fall, dass ein weiteres Webseite C unterstützt werden soll, würde ich die Anzahl an notwendigen Konvertern , wie Tabelle 3.1 zeigt , auf Sechs erhöhen.

Tabelle 3.1.: Anzahl Konverter bei drei Webseiten

		Nach		
		Webseite A	Webseite B	Webseite C
Von	Webseite A	-	×	×
	Webseite B	×	-	×
	Webseite C	×	×	-

Nimmt man an n_{ws} sei eine beliebige Anzahl sozialer Netzwerke, entspricht die Anzahl der notwendiger Konverter $n_{k1} = n_{ws} * (n_{ws} - 1)$, da für jedes Netzwerk ein Konverter in alle anderen Netzwerke erzeugt werden muss. Sollen nur Zwei oder Drei Netzwerke unterstützt werden ist der Aufwand noch sehr überschaubar, bei mehr kann dies aber sehr Aufwendig werden.

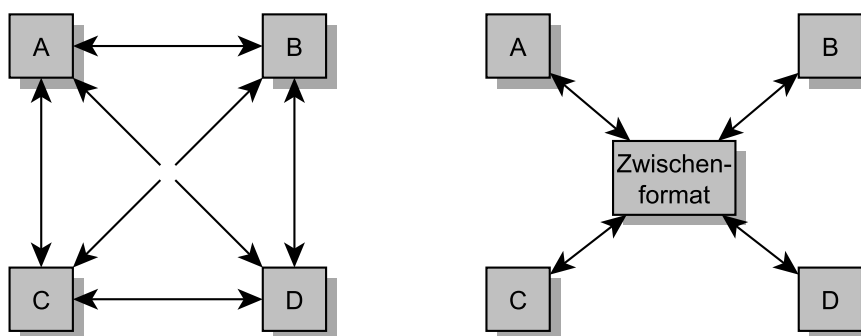


Abbildung 3.2.: Komplexität ohne und mit dem Einsatz eines Zwischenformats - Originalbild: [36]

Eine elegantere Methode für die Lösung dieses Problems, welche die Anzahl zu implementierender Konverter in Grenzen halten kann, wäre die Einführung eines Zwischenformates (auch Inter-Lingua, in Abbildung 3.3 als „IM Format“ bezeichnet) [36, S. 9]. Geht man davon aus, dass die Daten aller Webseiten nur in dieses Zwischenformat geschrieben und aus diesem gelesen werden müssen, würde sich der Aufwand auf maximal zwei Konverter je Webseite reduzieren. Für eine beliebige Anzahl Webseiten wären also $n_{k2} = n_{ws} * 2$ Konverter nötig. Nachteile hätte dieser Ansatz nur für $n_{ws} = 2$ und $n_{ws} = 3$, da in diesen Fällen mehr beziehungsweise gleich viele Konverter gegenüber der ersten Methode erforderlich wären. Erhöht man die Anzahl Webseiten jedoch nur geringfügig, sinkt die Menge an Konvertern sichtbar. Für $n_{ws} = 4$ wären es $n_{k2} = 8$ statt $n_{k1} = 12$ (siehe Abbildung 3.2) und für $n_{ws} = 5$ ergibt sich $n_{k2} = 10$ statt $n_{k1} = 20$ Konvertern. Gleichzeitig können so syntaktische Unterschiede in den einzelnen Formaten angeglichen werden, was sie leichter handhabbar macht.

Bei der automatischen Sammlung von benutzergenerierten Inhalten stellt sich immer die Frage der Privatsphäre. Nicht jeder möchte, dass vielleicht sensible Informationen von ihnen weitergegeben werden. Aus diesem Grund ist die Einführung eines Mechanismus sinnvoll mit

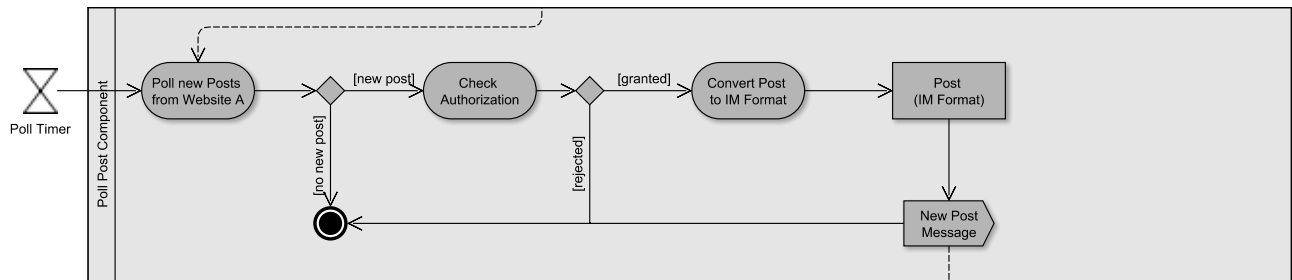


Abbildung 3.3.: Lesen des erstellten Beitrags und konvertieren in das Zwischenformat.

dem ein Benutzer das Lesen seiner Beiträge komplett erlaubt oder auf bestimmte Seiten, Foren oder nur für einzelne Threads beschränkt [6]. Sollte eine solche Erlaubnis nicht vorliegen, wird der gelesene Beitrag verworfen.

Da das Eintreffen neuer Beiträge nicht vorhersagbar ist, ist es angebracht beim Synchronisieren das Lesen und Schreiben zeitlich zu entkoppeln. Eine der weit verbreitetsten Techniken dazu ist das Versenden von Nachrichten über einen Nachrichtenkanal den die schreibende Komponente nach neuen Beiträgen abhört. Diesen Kanal können mehrere Gleichzeitig abhören und stellen so eine gute Flexibilität sicher.

3.1.2 Beiträge in eine andere Webseite schreiben

Empfängt nun eine Komponente, die für das Schreiben zuständig ist, eine Nachricht von einem neuen Beitrags der Webseite A ist der Ablauf ähnlich wie beim Lesen nur in umgekehrter Reihenfolge (Siehe Abbildung 3.4 links, oberer Ablauf). Der neue Beitrag wird vor dem Schreiben aus dem Zwischenformat in das Format B konvertiert. Da bei einer Synchronisation vom Vorteil wäre, wenn der synchronisierte Beitrag so aussehen würde, als hätte ihn der original Autor geschrieben. Hierzu muss das System Zugriff auf die Benutzerkonto haben und diese in einer Datenbank verwalten. Dort kann dann nach einem passenden Benutzerkonto des Autors gesucht und dieses dann zum Schreiben verwendet werden. Steht ein solches Benutzerkonto nicht zur Verfügung, ist das Ausweichen auf ein vorgegebenes Benutzerkonto hilfreich. Manche Benutzer sind vielleicht damit einverstanden, dass ihre Beiträge weitergegeben werden, aber nicht dass andere Programme automatisch in ihren Namen Beiträge schreiben. Deswegen ist wichtig davor noch einmal zu prüfen, ob eine entsprechende Erlaubnis gegen wurde. Sind alle Voraussetzungen gegeben, kann der Beitrag über das ausgewählte Benutzerkonto auf die Webseite B geschrieben werden.

3.2 Identifizierung der Komponenten

Anhand dieses kurzen Ablaufbeispiels kann man einige Komponenten ablesen, aus denen das neue System auf jeden Fall bestehen beziehungsweise enthalten muss und welche ergänzend dazu Wünschenswert wären:

- Eine Komponente muss Daten von einer Webseite über deren öffentlicher API in das System einlesen und diese in ein geeignetes Zwischenformat konvertieren können.

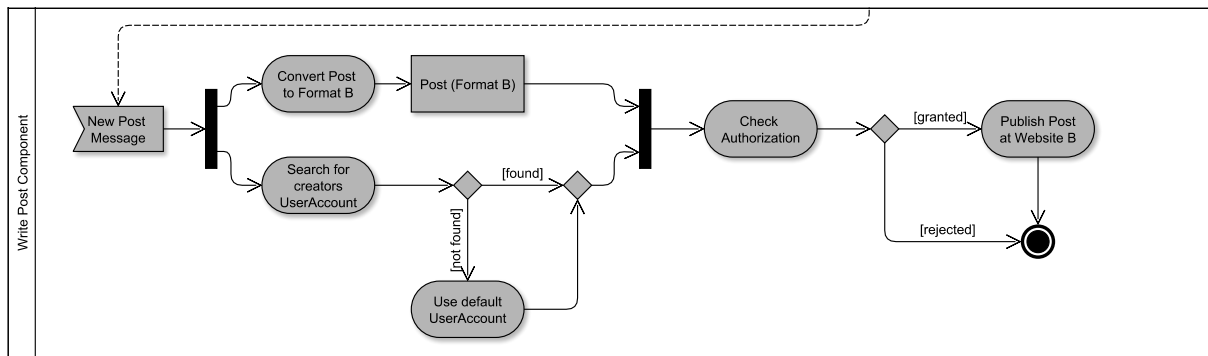


Abbildung 3.4.: Konvertierten des Beitrags in das Format B und schreiben n das soziale Netzwerk B

- Es muss ein passendes Nachrichtensystem zum Entkoppeln von Lesen und Schreiben gewählt und/oder erstellt werden.
- Eine weitere Komponente nimmt Beiträge im Zwischenformat entgegen, konvertiert diese in das Format des entsprechenden Netzwerkes und schreibt diese dorthin.
- Um stellvertretend für einen Benutzer schreiben zu können, muss es möglich sein nach dem Konto eines Benutzer zu einer bestimmten Webseite suchen zu können.
- Um die Privatsphäre der Benutzer zu wahren, wäre eine Mechanismus zum festlegen von Zugriffsrechten sinnvoll.

3.3 Wahl vorhandener Techniken

vielleicht doch eher an den Anfang von „Eigener Ansatz“!?

3.3.1 Welches Zwischenformat?

Ein passendes Zwischenformat wurde im Kapitel „2“ mit SIOC beziehungsweise die Kombination aus SIOC und FOAF bereits vorgestellt. Damit ist es nicht nur möglich einzelne Beiträge plattformunabhängig zu speichern, sondern es kann auch die Struktur der Plattform als auch der Benutzer mit eingebunden werden. Auch können Maschinen aus den RDF-Daten einfach Wissen für neue, hilfreiche Informationen ableiten. Doch warum sollte man auf RDF und Ontologien setzen anstatt etwas eigenes zu Entwickeln wie auf Basis des weitverbreiteten und vielseitigen XML Formats?

Hier gibt es mehrere Vorteile, warum man RDF einer mit XML-Schema definierten XML-Datei vorziehen sollte. XML hat das Problem, dass die selbe Aussage auf verschiedene Art repräsentiert werden kann. Listing 3.1 zeigt ein Beispiel für die Repräsentation der Aussage „Max Hiwi ist der Autor des Beitrags mit der ID 42“ in XML.

Listing 3.1: XML: Unterschiedliche Syntax, gleiche Semantik

1

```
<post>
```

```
2         <id>42</id>
3         <author>Max Hiwi</author>
4     </post>
5
6     <post id="42">
7         <author>Max Hiwi</author>
8     </post>
9
10    <post id="42" author="Max Hiwi" />
```

Alle drei XML-Elemente beschreiben die selbe Aussage, aber auf unterschiedliche Art und Weise. Für einen Menschen haben diese drei Schreibweisen die selbe Aussage, für eine Maschine ist dies aber nicht mehr ganz so offensichtlich. Auch kann nicht die Bedeutung der einzelnen Element und Tags von Maschinen erfasst werden. Dies muss der Programmierer erledigen. Zusätzlich erschweren diese unterschiedlichen Strukturen das ausführen von Abfragen auf diesen Daten, da ein Mapping zwischen einer logischen Abfrage und den Daten nicht eindeutig erfolgen kann (vgl. [32, s. 41]).

3.3.2 Welches Nachrichtensystem?

JMS

- zum reinen verschicken ausreichend
- fast alles muss selber gemacht werden.

Apache Camel

- erweitern der "Pipeline". B. mit Filtern einfach möglich

JMS + Apache Camel

- gute Kombination -> durable subscriber



4 Eigener Ansatz: Social Online Community Connectors (SOCC)

Aufbauend auf den in Kapitel 3 identifizierten Komponenten und Wahl der für ein System zur Synchronisation von Beiträgen passenden Techniken, soll nun der als *Social Online Community Connectors* (SOCC) benannter Ansatz vorgestellt werden. SOCC setzt für seine Aufgabe auf Techniken des in den Grundlagenkapitel beschriebenen Semantic Webs. Der Einsatz von offenen Datenformaten wie RDF und den darauf aufbauenden FOAF und SIOC nicht nur für den hier beschriebenen Zweck sondern auch für andere Projekte verwendet werden kann.

Den Aufbau von SOCC zeigt die Abbildung 4.1. Ein *Connector* des SOCC dient als Verbindungselement zwischen zwei oder mehreren Plattformen. SOCC stützt sich dabei auf die Prinzipien für die Verbindung von Daten im Web von Tim Berners-Lee (siehe [3]):

„Use URIs as names for things“ Seiten, Foren, Threads und Beiträge sollten immer über eine URI benannt werden, so dass andere Anwendungen diese verwenden können.

„Use HTTP URIs so that people can look up those names“ Die verwendeten URIs sollten dereferenzierbar sein, um auf die dahinter liegenden Daten zugreifen zu können.

„When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)“ Die Daten sollten in einem Standard für das Semantic Web formatiert werden. Wie schon erwähnt setzt SOCC dazu auf RDF und darauf aufbauenden Ontologien wie FOAF und SIOC. Dadurch können zum Beispiel andere Anwendungen mit Anfragen in SPARQL¹ nach Beiträgen suchen.

„Include links to other URIs. so that they can discover more things.“ Nicht nur die Struktur einer Diskussion kann über Links verfolgt werden, auch das Gewinnen zusätzlicher Informationen ist so möglich. Beiträge können auf Lernmaterialien wie Folien oder Videos verweisen oder über das FOAF-Profil eines Autor können weitere Beiträge von ihm gefunden werden.

URIs sind also das wichtigste Element mit denen SOCC arbeitet. Soll zum Beispiel eine Diskussion synchronisiert werden, muss einem Connector die URI übergeben werden, hinter der sich die Daten befinden. Ein Connector ist nur für eine einzige Plattform zuständig. Aber eine Plattform kann über mehrere Connectoren angesprochen werden, wodurch eine Arbeitsteilung möglich ist.

Intern besteht ein Connector aus drei Teilkomponenten die zum Einen für das Lesen (*PostReader*) und Schreiben (*PostWriter*) von Beiträgen verwendet werden, zum Anderen aus einem *StructureReader* der für das Auslesen der Struktur der einzelnen Diskussionen verantwortlich ist. Eine genaue Beschreibung dieser Komponenten folgt in Abschnitt 4.3.

¹ <http://www.w3.org/TR/sparql11-overview>

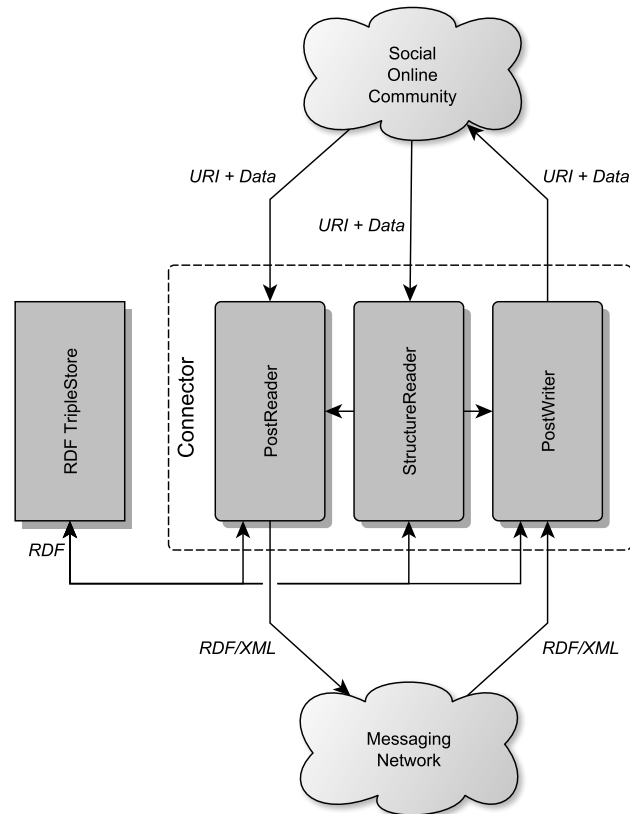


Abbildung 4.1.: Übersicht der Komponenten der SOCC

Jeder Connector hat Zugriff auf eine als *Triplestore* bezeichnete Datenbank in der RDF-Triples gespeichert und unter anderen mit SPARQL abgefragt werden. Der Connector benutzt diesen Triplestore als Speicher in dem seine Konfigurationsdaten lagern, aber auch zusätzliche Daten die er von Außerhalb, wie eine Liste von Benutzerkonten, benötigt. Er wird aber auch benutzt um Daten zu speichern die während des Betriebes anfallen, da sie so irgendwann wieder verwendet werden können ohne sie erneut zusammensuchen. Zum Beispiel die Daten über die Struktur der verwendeten SOC.

Die Beiträge werden dann zwischen den einzelnen Connectoren über ein Nachrichtennetzwerk auf der Basis von Apache Camel ausgetauscht. Die Beschreibung dieser als *SOCC-Camel* bezeichneten Komponente geschieht in Kapitel 4.4.

4.1 Datenformat

Auch begründen warum die Verwendung von Ontologien und nicht bspw. die Entwicklung eines XML Schemas (alternativ bei der Analyse Kap. 3.2)

4.2 Konfiguration

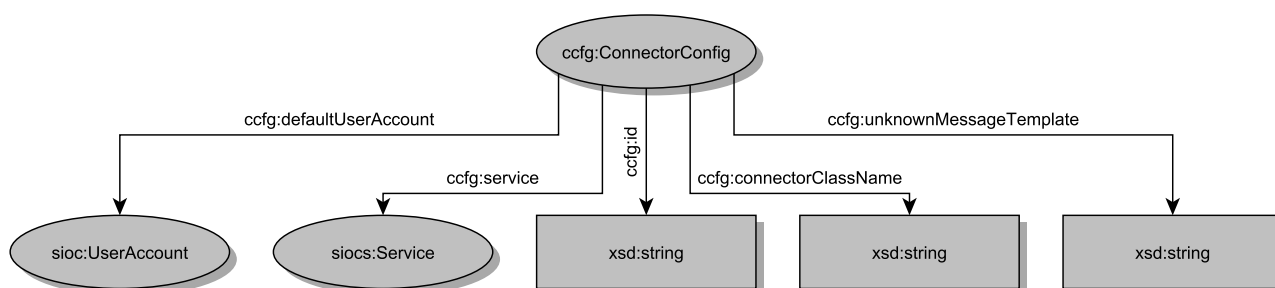
Dass ein Connector funktionieren kann, muss er von außen mit Informationen zugeführt bekommen welche er zum Betrieb braucht. Die sind zum Beispiel Informationen zu Benutzerkonten oder Parameter für die verwendete API. Da einige dieser Informationen werden nicht nur von einem Connector benutzt werden, ist es sinnvoll diese zusammen an einen Ort zu speichern und wiederverwenden zu können. Die wichtigsten Informationen für die Konfiguration der Connectoren stellen die Benutzerkonten dar. Sie enthalten unter anderem die Informationen um Zugriff auf die einzelnen APIs zu erhalten. Da die Benutzerkonten wie im Abschnitt 4.1 beschrieben im FOAF Format in einen Triplestore gespeichert werden, stellt es sich als Vorteil heraus die übrigen Informationen ebenfalls dort zu speichern und mit den schon vorhandenen zu verbinden.

Aus diesem Grund wurde für Konfiguration eines Connectors die *Connector Config Ontology* entwickelt. Diese Ontologie ist sehr einfach gehalten und baut auf schon vorhandenen Ontologien auf. Zusätzlich musste die SIOC Ontologie so erweitert werden, dass die Integration von Autorisierungs- und Authentifizierungsinformationen möglich war.

4.2.1 SOCC Connector Config Ontologie

Abbildung 4.2 zeigt die entwickelte *SOCC Connector Config* Ontologie. Sie besteht aus einer einzigen Klasse *ConnectorConfig* und fünf Eigenschaften für diese.

Einleitung umschreiben, klingt scheiße



@prefix ccfg: <http://www.m0ep.de/socc/config#> .
@prefix sioc: <http://rdfs.org/sioc/ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

Abbildung 4.2.: Schema der SOCC Connector Config Ontology

Jeder Connector erhält einen eindeutigen *id* zugewiesen, um jeden Connectoren später eindeutig identifizieren zu können. Die Eigenschaft *connectorClassName* beschreibt den vollständigen Klassennamen des beschriebenen Connector. Diese wird für das laden der richtigen Implementierung benötigt. Manchmal kann es passieren, dass keine passendes Benutzerkonto zum Weiterleiten eines Beitrags gefunden werden kann. Dadurch es es wünschenswert solche Beiträge dahingegen zu verändern, dass eine Verweis auf den original Autor und vielleicht wo der Beitrag gemacht wurde vorhanden ist. Durch die Eigenschaft *unknownMessageTemplate* kann eine Vorlage für das Aussehen des Verweises definiert werden. Innerhalb dieser Vorlage stehen einige

Variablen in der Form „{varName}“ zur Verfügung. Alle zur Zeit vorhandenen Variablennamen und deren Ersetzung sind in Tabelle 4.1 zu sehen.

Tabelle 4.1.: Variablennamen und Ersetzung innerhalb von `unknownMessageTemplate`

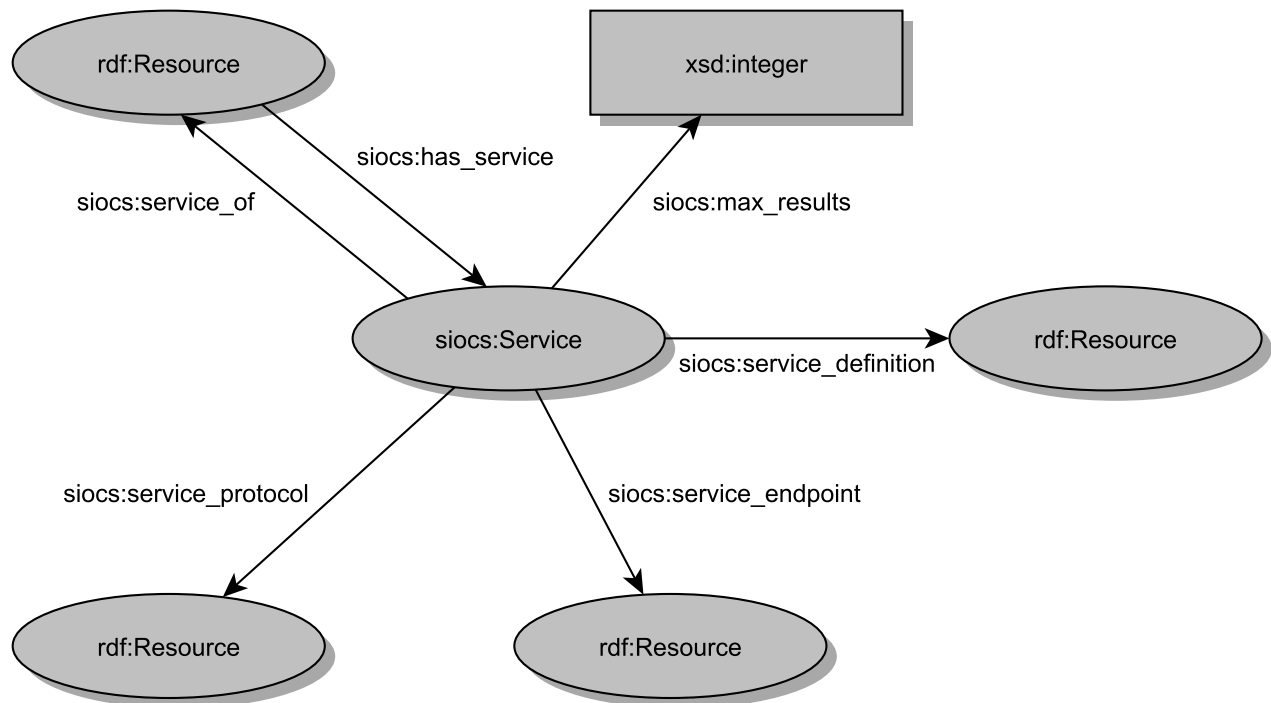
varName	Ersetzt durch
message	Original Beitrag
sourceUri	URI des original Beitrags
connectorId	ID des aktuellen Connectors
serviceName	Name des vom Connector verwendeten Service
creationDate	Erstelldatum des Beitrags (falls bekannt)
authorName	Name des Autors (falls bekannt)

Für die Nutzung einiger APIs müssen bestimmte Parameter angegeben werden. Dies könnte zum Beispiel die genau Adresse des Dienstes sein. Hierzu wird auf die schon Bestehende SIOC Services Ontologie zurückgegriffen. Diese stellt eine Klasse *Service* zur Verfügung und mittels der Eigenschaft `service` kann ein solcher Service einem Connector zugewiesen werden. Der genaue Aufbau eines solchen Services wird im Abschnitt 4.2.2 dargestellt. Die letzte Information für die Konfiguration eines Connectors ist eine vordefinierter Benutzer (im Folgenden Defaultuser genannt) und wird mit der Eigenschaft `defaultUserAccount` festgelegt. Dieser Defaultuser erfüllt im Großen und Ganzen zwei Aufgaben. Als Erstes wird er für lesende Zugriffe der API auf den verwendeten Dienst genutzt. Hierzu ist ein einzelnes Benutzerkonto vollkommen ausreichend, da nur die gelesenen Daten wichtig sind und nicht von welchen Konto sie kommen. Die zweite Aufgabe bezieht sich auf das stellvertretende Schreiben einzelner Benutzer. Nicht immer werden die dazu notwendigen Daten von den Benutzer zur Verfügung gestellt oder sind unbekannt. In diesem Fall wird der Defaultuser genutzt und der Beitrag mit einem Vermerk zum original Autor über diesen geschrieben.

4.2.2 Services

Wie eben schon beschrieben, existiert für SIOC ein Modul zur einfachen Modellierung von Diensten auf semantischer Ebene: Das SIOC Services Module (Präfix *siocs:*). Kernstück dieses Moduls ist die Klasse *Service*, wie auf Abbildung 4.3 zu sehen ist. Mit dieser Klasse kann durch eine Hand voll Eigenschaften ein Dienst beschrieben werden. Für diese Arbeit ist davon die wichtigste Eigenschaft `service_endpoint`. Durch diese kann die Adresse festgelegt werden, unter dem ein bestimmter Dienst erreichbar ist. Gerade bei Plattformen die nicht an eine feste Adresse (Foren, Blogs, ...) gebunden sind, ist diese Angabe unerlässlich. Die Eigenschaften `has_service` und `service_of` sind ideal zur Verbindung von einzelnen SIOC UserAccounts mit einem Service. Diese Verbindung hilft dabei für das stellvertretende Schreiben von Beiträgen schnell die passenden Benutzerdaten zu finden. Ebenfalls nützlich ist `max_results`. Manche Dienste erlauben es nur eine maximale Anzahl an Ergebnissen pro Aufruf zurückgeben zu lassen. Da sich diese Anzahl über die Zeit ändern kann ist es nicht sinnvoll diese fest im

Programm festzulegen, kann diese so im Nachhinein verändert werden. Für SOCC weniger interessant aber Vollständigkeit halber seien noch erwähnt `service_protocol` zum Angeben des verwendeten Übertragungsprotokolls (REST, SOAP, ...) und `service_definition` mit dem auf eine weiterführende Definition verwiesen werden kann.



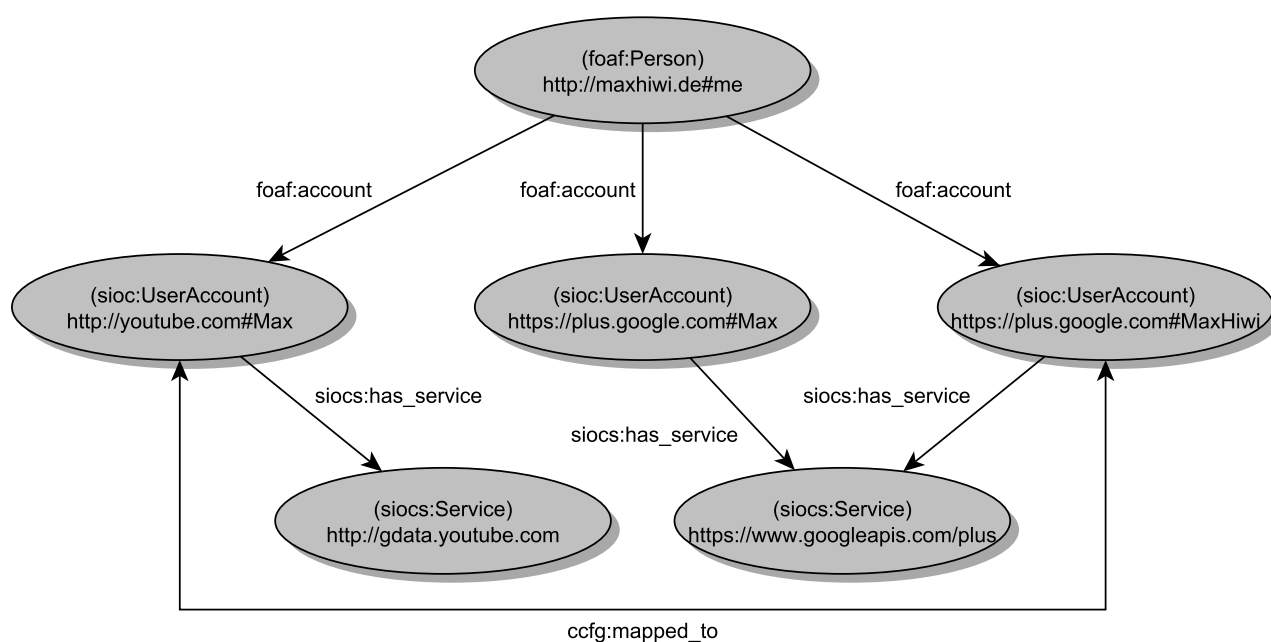
@prefix siocs: <http://rdfs.org/sioc/services#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

Abbildung 4.3.: SIOC Services Module

4.2.3 Benutzerdaten

Soll ein Beitrag eines Benutzers von Google+ nach Facebook synchronisiert werden und es so aussehen, als hat er diesen Beitrag selbst auf Facebook geschrieben, sind gute Kenntnisse über alle Benutzerkonten dieser einen Person notwendig. Als erstes muss die Existenz dieser Person dem System bekannt sein. Hierzu kann diese durch die Klasse `Person` aus der FOAF Ontologie dargestellt werden. Für ein einzelnes Benutzerkonto wurde in SIOC die Klasse `UserAccount` definiert. Da `UserAccount` eine Unterklasse von `OnlineAccount` aus FOAF ist, kann diese über die Eigenschaft `foaf:account` beziehungsweise `sioc:account_of` mit einer Person verbunden werden. Da es wichtig ist zu wissen zu welcher Webseiten ein Benutzerkonto gehört, wird der `UserAccount` mit einem Objekt der Klasse `Service` über die Eigenschaft `siocs:has_service` / `siocs:service_of` zusammengebracht. Diese Verbindung ist für manche APIs besonders bedeutend, da in dem Serviceobjekt relevante Daten für den Zugriff darauf enthalten sind. Nun

kann es vorkommen, dass eine Person private und geschäftliche Benutzerkonten besitzt. Um nicht private Beiträge auf Webseite A mit dem geschäftlichen Benutzerkonto auf Webseite B zu schreiben, muss ein Mapping zwischen den verschiedenen Benutzerkonten festgelegt werden. Dieses Mapping kann über ein in der „SOCC Connector Config Ontologie“ definierte Eigenschaft `mapped_to` realisiert werden. Diese Eigenschaft ist bijektiv, also jedes Benutzerkonto maximal auf ein anderes Konto gemappt werden darf. Ebenso gilt, falls Benutzerkonto A mit Benutzerkonto B verbunden ist, ebenfalls B nach A gemappt wird. Abbildung 4.4 zeigt den Zusammenhang zwischen der Klasse `Person`, `UserAccount` und `Service` sowie der Eigenschaft `mapped_to` noch einmal graphisch an einem Beispiel.



@prefix ccfg: <http://www.m0ep.de/socc/config#> .
 @prefix sioc: <http://rdfs.org/sioc/ns#> .
 @prefix siocs: <http://rdfs.org/sioc/services#> .
 @prefix foaf: <http://xmlns.com/foaf/0.1/> .

Abbildung 4.4.: Zusammenhang von `Person`, `UserAccount` und `Service`. Die inversen Eigenschaften `sioc:account_of` und `siocs:service_of` wurden zu einer besseren Übersicht weggelassen

4.2.4 Authentifizierung

Die Ontologien FOAF und SIOC sind hervorragend für die Abbildung von sozialen Netzwerken und Diskussionen, jedoch ist es mit ihnen nicht möglich Daten zur Authentifizierung (Feststellen ob jemand der ist, den er vorgibt zu sein). Wegen dem Schutz der Privatsphäre ist dies verständlich, jedoch um stellvertretend für einen Benutzer Beiträge zu schreiben, ist es wichtig Zugriff auf diese Daten zu haben und sie einem Benutzerkonto zuzuordnen zu können. Zuerst muss dazu aber festgestellt werden, welche verschiedenen Mechanismen es zum Anmelden ein solches Konto existieren.

Username/Passwort ist wohl eine der ersten und häufigsten Mechanismen, um den Zugriff sensibler Daten vor Dritten zu schützen. Das in Abschnitt 5.2.3 beschriebene LMS Moodle, setzt zum Beispiel den Username und Password eines angemeldeten Benutzers zu Authentifizierung ein.

OAuth ² stellt heutzutage den Standard der verwendeten Authentifizierungsmechanismen für hauptsächlich webbasierte API dar. Benutzer können so temporär Programmen den Zugriff auf ihre Daten erlauben und später wieder verbieten. Der aktuelle Standard stellt OAuth 2.0 dar und wird in dieser Version von den größten Seitenbetreibern wie Google, Facebook oder Microsoft eingesetzt³. Insgesamt sind für die Nutzung von OAuth vier Parameter wichtig. Für das Programm, dass Zugriff erhalten möchte sind die Parameter *client_id* und *client_secret* (Siehe [20][S. 8]). Sie weisen das Programm als autorisiert für die Benutzung der Schnittstelle aus. Soll nun beim Aufrufer einer von OAuth geschützten Funktion belegt werden ist ein sogenannter Accesstoken nötig (Siehe [20][S. 9]). Da dieser Accesstoken in der Regel nur eine bestimmte Zeit gültig ist, wird je nach Implementierung des Standards noch ein RefreshToken mitgeliefert. Mit diesem RefreshToken ist das Programm in der Lage ohne Zutun des Benutzers einen abgelaufen Accesstoken wieder zu aktivieren. Dies kann beliebig oft wiederholt werden, bis der Benutzer beide Token für ungültig erklärt.

vll. noch OAuth 1.0(a) einbauen

API Schlüssel sind eine dritte Möglichkeit Programmen Zugriff auf eine API zu gewähren. Der API Schlüssel entspricht ungefähr einer Kombination von *client_id* und *client_secret* von OAuth. Dieser Schlüssel schaltet in der Regel nicht den Zugriff auf persönliche Daten von Benutzer frei. Hier ist noch ein weiterer Mechanismus wie die Verwendung von einem Usernamen und Passwort nötig. Die in Abschnitt 5.2.6 beschriebene Google Youtube API hierzu ein gutes Beispiel.

Neben diesen drei Mechanismen wäre noch der Vollständigkeit halber die HTTP-Authentifizierung zu nennen. Hierbei handelt es sich um eine Form des Username/Passwort Verfahrens, welches auf dem HTTP Protokoll aufsetzt. Für einfachen Webseiten ist dies eine unkomplizierte Art die Datei vor fremden Zugriffen zu schützen. Für aktuelle öffentliche APIs ist diese Form der Authentifizierung nicht mehr Stand der Technik.

Die Suche nach einer bestehenden Ontologie, welche zusammen mit SIOC verwendet werden könnte, gestaltete sich als sehr schwierig. Ein Großteil der Ontologien in diese Richtung befasst sich eher mit dem Thema der Autorisierung wie zum Beispiel die *Web Access Control List* [24] mit Zugriffssteuerungsliste. Eine Ausnahme stellt die *Authentication Ontology*⁴ des *OmniVoke*⁵ Frameworks dar. Die Art der Authentifizierung wird darin durch die Klasse *AuthenticationMechanism* modelliert. Unterklassen für die wichtigsten Mechanismen wie OAuth, WebAPIs und Username/Passwort (dort *Direct* genannt) sind vorhanden. Jedem *AuthenticationMechanism* Objekt können dann sogenannte *Credentials* (engl. für Anmeldedaten) angehängt werden.

² OAuth Webseite: <http://oauth.net/>

³ OAuth Versionen im Einsatz: http://en.wikipedia.org/wiki/OAuth#List_of_OAuth_service_providers

⁴ Authentication Ontology: <http://omnivoke.kmi.open.ac.uk/authentication/>

⁵ OmniVoke: <http://omnivoke.kmi.open.ac.uk/framework/>

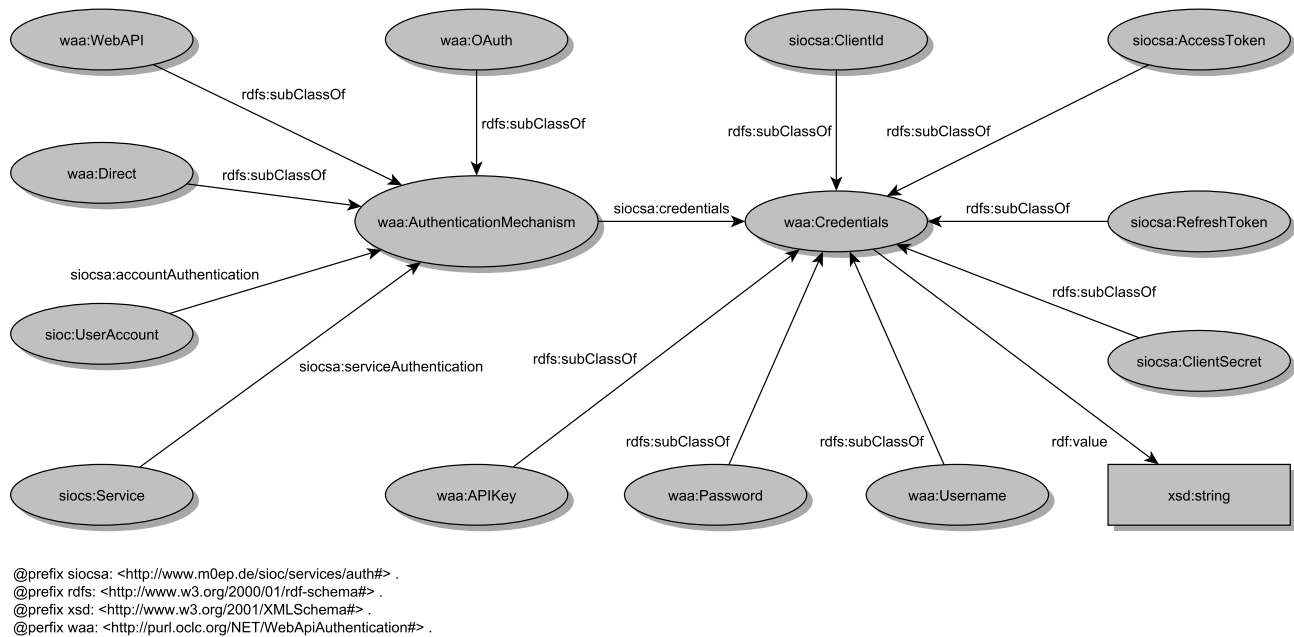


Abbildung 4.5.: SIOC Services Authentication Ontology

Das einzige Manko an dieser Ontologie war das Fehlen von Credentials für OAuth in der Version 2.0. In einzelnen waren dies Klassen für `client_id`, `client_secret` sowie für Access- und RefreshToken. Um auch diese OAuth Version unterstützen zu können, wurden hierfür die Klassen `ClientId`, `ClientSecret`, `AccessToken` und `RefreshToken` als Unterklassen von `Credentials` abgeleitet. Als Letztes musste noch eine Verbindung zwischen Authentication Ontology und SIOC hergestellt werden. Zum Einen war eine Erweiterung der Klasse `UserAccount` notwendig, so dass die Anmeldedaten der Benutzer zur Verfügung standen. Zum Anderen werden Daten wie ein API Schlüssel von einem Service benötigt, die von denen der Benutzer unabhängig sind. Für die Klasse `UserAccount` wurde die Eigenschaft `accountAuthentication` geschaffen. Diese erwartet als Subjekt einen `UserAccount` und als Objekt ein `AuthenticationMechanism`, welcher dann die Credentials enthält. Für die Klasse `Service` existiert das Äquivalent `serviceAuthentication`.

Diese Erweiterungen (Präfix `siocsa:`) und die übernommenen Teile der Authentication Ontology wurden danach im *SIOC Services Authentication Module* zusammengefasst. Graphisch ist sie in Abbildung 4.5 und im Anhang A.2 als OWL Schema zu sehen.

4.2.5 Autorisierung

Da für viele Menschen im Internet ihre Privatsphäre wichtig ist und sie nicht wollen, ist es wichtig von ihnen die Erlaubnis einzuholen, um ihr Beiträge zu lesen und weiterzuleiten. Ein verbreitetes Mittel für eine solche Zugriffssteuerung sind Access Control Lists (ACL) (engl. für Zugriffsteuerungsliste). Mit ihnen wird geregelt wer bestimmte Operationen auf eine Ressource durchführen darf. Für den Einsatz in dieser Arbeit wurde die Basic Access Control Ontologie [24, 1] ausgewählt (Im Folgenden nur als ACL bezeichnet). Da sie Teilweise auf FOAF aufbaut und so ließ sie sich sehr einfach in das bestehende System integrieren.



@prefix acl: <http://www.w3.org/ns/auth/acl#> .
 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
 @prefix gen: <http://www.w3.org/2006/gen/ont#> .
 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

Abbildung 4.6.: Basic Access Control Ontologie

Zugriffsrechte für Ressourcen werden in dieser ACL mit der Klasse `acl:Access` modelliert. Von dieser Klasse werden einzelne Rechte wie `acl:Read` für das Lesen und `acl:Write` für das Schreiben abgeleitet. Nebenbei existieren noch Ableitungen `acl:Control` zum Ausführen und `acl:Append` zum Anfügen von Daten an die Ressource. In SOCC werden aber nur Rechte für das Lesen von Ressourcen gebraucht.

Die Verbindung von einem Zugriffsrecht mit einer Ressource wird über die Klasse `acl:Authorization` erreicht. Der Besitzer dieser Autorisierung wird durch die Eigenschaft `acl:owner` festgelegt und gehört zur Klasse `foaf:Agent` beziehungsweise in SOCC zur Unterklasse `foaf:Person`. Mittels der Eigenschaft `acl:agent` wird festgelegt für welche Person/Agenten diese Autorisierung gilt. Das selbe wird auch über die Eigenschaft `acl:agentClass` festgelegt, wobei hier eine abstrakte Klasse statt einer Instanz davon gemeint ist. Soll zum Beispiel der öffentlicher Zugriff für eine Ressource definiert werden, wird bei `acl:agentClass` die Klasse `foaf:Agent` eingesetzt [1, „Public Access“]. Dies besagt, dass alle Agenten auf diese Ressource zugreifen können. Innerhalb von SOCC werden Ressourcen verarbeitet, die öffentlich zugänglich sind. Welches Recht für eine Ressource eingeräumt wird, wird über die Eigenschaft `acl:mode` festgelegt. Erlaubt ist hier die Angabe jeder beliebigen Klasse, SOCC testet aber nur ob das Recht `acl:Read` zum Lesen von Beiträgen eingeräumt wurde. Auf welche Ressource sich letztendlich eine Autorisierung bezieht, wird über die Eigenschaft `acl:accessTo` geregelt. Die Angabe von „`http://www.facebook.com`“ würde sich für SOCC zum Beispiel auf alle Beiträge des Besitzers auf Facebook beziehen, „`https://canvas.instructure.com/courses/798152`“ dahingegen nur auf alle Beiträge innerhalb eines Canvas Kurses mit der ID „798152“. Für einen Zugriff auf alle Beiträge prüft SOCC ob die Eigenschaft `acl:accessToClass` auf die Klasse `sioc:Post` verweist. So müsste nicht jede einzelne Webseite angegeben werden, wenn man ein generelles Zugriffsrecht einräumt.

Das Listing 4.1 zeigt ein Beispiel, wie eine Autorisierung mit der ACL aussehen könnte. Der Besitzer dieser Autorisierung mit der URI `http://example.org#john` wird in Zeile 6 beschrieben. Er erlaubt damit SOCC einen öffentlichen (Zeile 7), lesenden (Zeile 9) Zugriff auf all seine geschriebenen Beiträge (Zeile 8).

Listing 4.1: ACL Beispiel

```
1 @prefix sioc: <http://rdfs.org/sioc/ns#> .
2 @prefix foaf: <http://xmlns.com/foaf/0.1/>
3 @prefix acl: <http://www.w3.org/ns/auth/acl#> .
4
5 [] a acl:Authorization ;
6     acl:owner <http://example.org#john> ;
7     acl:agentClass foaf:Agent ;
8     acl:accessToClass sioc:Post ;
9     acl:mode acl:Read .
```

4.3 Design eines Connectors

Dieser Abschnitt widmet sich den internen Aufbau eines Connectors und beschreibt alle Komponenten aus denen er besteht und wie sie zusammenarbeiten. Im einzelnen sind dies der SOCC Context, die *AccessControl*, der *ClientManager*, der *StructureReader*, der *PostReader* und der *PostWriter*. Das vollständige UML Klassendiagramm mit allen Zusammenhängen ist in Abbildung 4.7 zu sehen.

4.3.1 SOCC Context

Der SOCC Context eines Connectors, beschreibt die Umgebung die er zum Arbeiten braucht. Über ihn bekommt der Connector Zugriff auf den sich außerhalb befindlichen Triplestore, der durch die Klasse *Model* der RDF2Go Bibliothek abstrahiert wird (Siehe Abschnitt 5.1.2). In ihm befinden sich alle Daten die der Connector für seinen Betrieb benötigt und benutzt ihn gleichzeitig als Lagerplatz für welche die während des Betriebs gespeichert werden müssen. Eine Referenz auf dieses Triplestore erhält der Connector über den Aufruf der Funktion *getModel()*. Durch die Methode *getAccessControl()* kann der Connector über die im folgenden Abschnitt beschriebene *AccessControl* Schnittstelle auf die Information für die Zugriffssteuerung zugreifen.

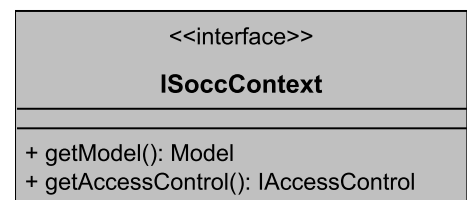


Abbildung 4.8.: SOCC Context

4.3.2 AccessControl

Die *AccessControl* Schnittstelle ist sehr einfach gehalten und dient für den Zugriff auf die in Abschnitt 4.2.5 beschriebenen ACL Information. Die Methode *checkAccessTo(...)* prüft,

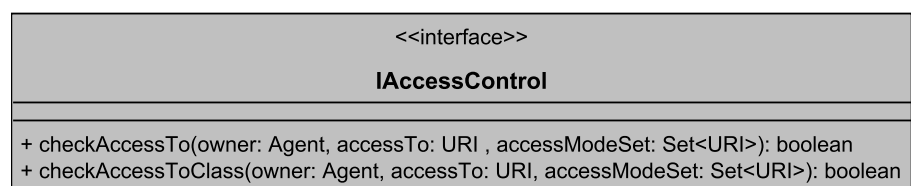


Abbildung 4.9.: AccessControl

ob der Zugriff auf eine Resource mit allen übergebenen Rechten erlaubt ist. Die andere Methode `checkAccessToClass` ist zur Überprüfung, ob die Rechte für den Zugriff auf eine komplette Klasse von Ressourcen vorhanden sind.

4.3.3 ClientManager

Der Zugriff auf eine API innerhalb des Programms erfolgt in der Regel über ein spezielles Clientobjekt (kurz Client). Dieser Client erlaubt es mit den Anmeldedaten für ein Benutzerkonto auf die Funktionen der API über verschiedene Methoden zu zugreifen. Da ein Client immer nur mit einem Benutzerkonto verknüpft ist und von diesen eine große Anzahl verwaltet werden müssen, enthält jeder Connector einen eigenen ClientManager für diese Aufgaben. Der interne Aufbau eines Client ist dabei stark von der verwendeten API abhängig und wird als generischer Parameter der Klasse beim Erzeugen eines Objekts übergeben. Für alle vom Benutzer unabhängigen Daten erhält der ClientManager ein, wie Abschnitt 4.2.2 beschriebenes, Objekt

der Klasse `Service`, dass zum Beispiel die Basis-URI für die API enthält. Das Erzeugen eines neuen Clients erfolgt dann durch den Aufruf der Methode `createClient(...)`. Als Parameter wird er ein Benutzerkonto, in Form eines SIOC `UserAccounts`, übergeben. Sind alle erforderlichen Authentifizierungsinformation aus Abschnitt 4.2.4 vorhanden, wird ein neuer Client erzeugt und zurück an den Aufrufer gegeben. Dieser Client wird aber dadurch nicht automatisch vom ClientManager verwaltet. Hierzu muss der im vorherigen Schritt erzeugte Client durch die Übergabe an `add(userAccount: UserAccount, client: T)` dauerhaft mit den angegebenen `UserAccount` verknüpft und intern gespeichert werden. In der aktuellen Implementierung ist es wichtig, dass die Eigenschaften `accountName` und `accountServiceHomepage` des `UserAccount` Objekts gesetzt sind. Aus diesen wird ein eindeutiger Schlüssel generiert, der zur Zuordnung von `UserAccount` und Client innerhalb des ClientManagers dient. Des weiteren stehen noch Methoden `remove(userAccount: UserAccount)` zum Entfernen und `get(userAccount: UserAccount)` zum Holen von Clients, sowie `contains(userAccount: UserAccount)` für Tests, ob ein Client zu einem `UserAccount` existiert, zur Verfügung. Sollen zum Beispiel am Ende der Laufzeit des Programms alle erzeugten Clients auf einmal abgemeldet und gelöscht werden, kann dies über die Methode `clear()` erfolgen. Der ClientManager verwaltet ebenfalls den Client für den in Abschnitt 4.2.1 angesprochenen Defaultuser. Dieser Defaultclient genannte Client kann über die Methode `setDefaultClient(client: T)` gesetzt und durch `getDefaultClient()` jederzeit wieder abgerufen werden.

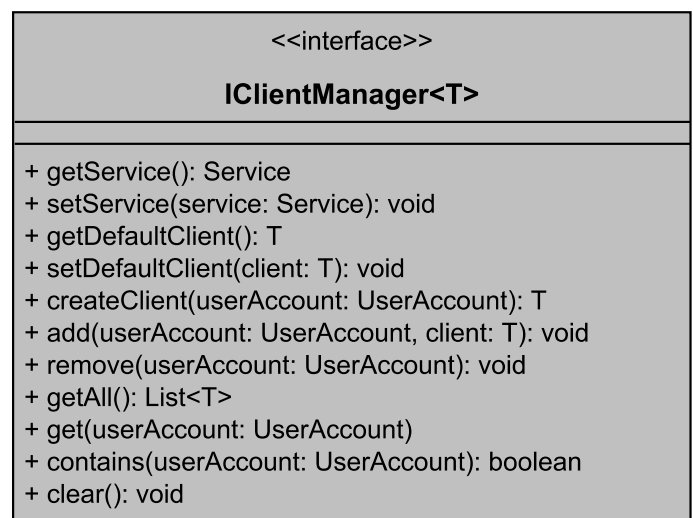


Abbildung 4.10.: ClientManager

4.3.4 StructureReader

Um auf Informationen über die Struktur von Foren, sozialen Online-Netzwerken und so weiter im SIOC-Format zugreifen zu können, implementiert jeder Connector dazu einen StructureReader. Die Struktur lässt sich, wie im Abschnitt 2.2.2 vorgestellt, durch die SIOC Klassen *Site* und *Container* (sowie Unterklassen davon) beschrieben. Für den Zugriff auf diese Struktur, enthält der StructureReader mehrere Methoden die in der UML Klasse rechts in Abbildung 4.11) zu sehen sind.

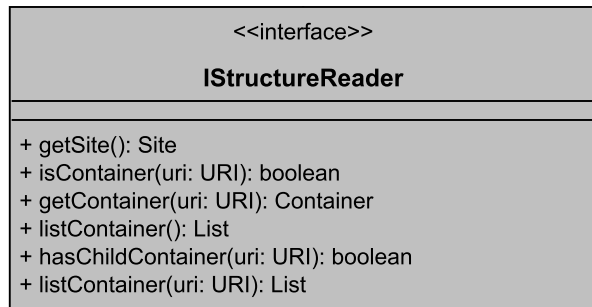


Abbildung 4.11.: StructureReader

`getSite()` ist eine Methode, welche die Beschreibung einer Seite (Forum, Blog, soziales Online-Netzwerk) als Objekt der SIOC-Klasse *Site* zurückliefert. Dieses wird relativ häufig gebraucht um die Zugehörigkeit anderer Objekte durch einen Verweis zu dieser Seite zu verdeutlichen. Dies kann bei einigen APIs nützlich sein, da dort manchmal keine Information zum Ursprungsort eines Beitrags mitgeliefert werden, über den man sonst eine Beziehung zwischen Seite und Beitrag herstellen könnte.

`isContainer(uri: URI)` wird zu Überprüfung verwendet, ob sich hinter einer URI ein potenzieller Container befindet.

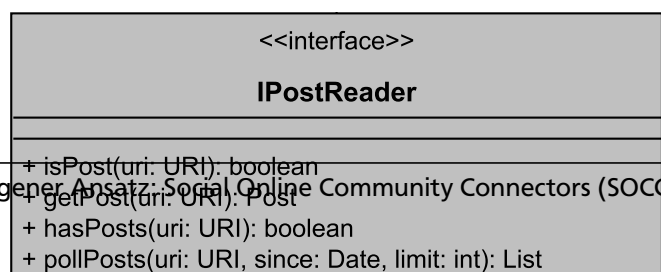
`getContainer(URI)` liefert Informationen zu einem Container im SIOC-Format, der sich hinter der übergebenen URI befindet.

`hasChildContainer(uri: URI)` überprüft, ob der Container hinter der übergebenen URI weitere Container als Kinder besitzt. Diese Methode wird dazu eingesetzt, um vorab zu testen, ob der Aufruf von `listContainer(URI)` das gewünschte Ergebnis liefert oder ein Fehler auftreten würde.

`listContainer(...)` sind Methoden, welche für den die Auflisten aller Container einer Seite aus der Sicht des Defaultusers zur Verfügung stehen. Die Methode ohne Parameter listet alle Container auf der ersten Ebene der Seitenstruktur auf. Dies könnten zum Beispiel alle Kurse innerhalb einer Canvasinstanz oder alle beigetretenen Gruppen auf Facebook sein. Die zweite Methode mit einer URI als Parameter gibt eine Liste aller Container, welche den Container hinter der übergebenen URI als Elternteil haben, zurück. Dies könnten im Falle von Canvas alle Diskussionsthemen innerhalb eines Kurses sein.

4.3.5 PostReader

Der PostReader dient als Schnittstelle für das Lesen geschriebener Beiträge innerhalb eines Containers oder der Kommentare auf einen anderen Beitrag. Er stellt nach außen



hin Funktionen bereit mit denen entweder ein einzelner Beitrag oder alle Beiträge, die bestimmte Kriterien erfüllen, gelesen werden können. Bevor ein Beitrag zurück gegeben wird, müssen die Methoden prüfen, ob der Autor eines Beitrag das Lesen für diesen erlaubt hat. Falls nicht, wird der Beitrag aus der Ergebnisliste gestrichen und nicht zurückgegeben. Die Funktionsweise der einzelnen Methoden ist wie folgt:

`isPost(uri: URI)` kann zur Überprüfung eingesetzt werden, ob sich hinter einer URI ein Beitrag befindet.

`getPost(uri: URI)` ist dazu gedacht einen einzelnen Beitrag anhand seiner URI zu lesen. Sie liefert dann den Beitrag als Objekt der SIOC-Klasse `Post` zurück.

`hasPost(uri: URI)` funktioniert ähnlich wie `isPost`, überprüft aber ob sich hinter der angegeben URI noch weitere Beiträge befinden können.

`pollPosts(uri: URI, since: Date, limit: int)` ist eine Methode die alle Beiträge hinter eine URI liest welche die gegebenen Kriterien erfüllen. Insgesamt erhält diese Methode drei Parameter. Der Erste ist eine URI die den Ort angibt von der die Beiträge gelesen werden sollen. Mit dem zweiten Parameter kann ein Zeitpunkt angegeben werden, ab dem ein zu lesender Beitrag geschrieben sein muss. Zum Beispiel der Zeitpunkt als diese Methode das letzte mal aufgerufen wurde, um alle Beiträge die danach geschrieben wurden zu lesen. Der letzte Parameter gibt eine obere Schranke an, wie viele Beiträge maximal pro Aufruf dieser Methode gelesen werden dürfen. Ist dieses Limit erreicht, werden keine weiteren Beiträge in die Ergebnisliste aufgenommen.

4.3.6 PostWriter

In Abbildung 4.13 ist ein Sequenzdiagramm der `PostWriter` Komponente zu sehen. Dort ist visualisiert, welche Schritte für das stellvertretende Schreiben von Beiträgen eines Benutzers unternommen werden müssen. Soll nun ein Beitrag in Plattform des Connectors geschrieben werden, wird die Methode `writePost(URI, String, Syntax)` mit dem Zielort als URI, dem Beitrag als serialisiertes RDF Objekt und dem verwendeten Serialisierungsformat aufgerufen. Es beginnt damit, dass zuerst nach einem Benutzerkonto des Beitragautors für die Plattform des aktuellen Connectors gesucht wird. Dieser sollte im Idealfall im Triplestore des Connectors nach dem im Abschnitt 4.2.3 Schema vorliegen. Mit diesem Benutzerkonto, in Form der Klasse `UserAccount` von SIOC, kann dann vom `ClientManager` ein `Client` für die verwendete API geholt werden. Sollte die Suche negativ verlaufen, steht noch der `Defaultclient` zur Verfügung. Mit dem `Client`, ob mit der des Autors oder dem `Defaultclient`, wird im letzten Schritt der Beitrag in das von der API verwendete Format konvertiert und an die richtige Stellen der Plattform geschrieben.

4.4 SOCC-Camel

SOCC-Camel ist ein Modul der SOCC für die Integration von Connectoren in Camel. Durch dieses ist auf flexible Weise möglich die gelesenen Beiträge von einer Plattform über die Connectoren in eine andere zu schreiben. Abbildung 4.14 zeigt SOCC-Camel als EIP-Diagramm.

Hauptklasse dieses Moduls ist `SoccComponent`. Sie ist für die Verwaltung von Connectoren und Erstellen der Endpunkte für Camel zuständig. Von dieser Klasse muss zuerst ein Objekt erzeugt und ein `SoccContext` mit dem Triplestore und alle Informationen für die Connectoren wie `ConnectorConfigs` Benutzerkonten und Services übergeben werden. Also alle Daten die in Abschnitt 4.2 beschrieben wurden. Das `SoccComponent` Objekt wird dann unter einen frei wählbaren Namen in Camel registriert. Als Namen ist „(“socc“) aber vorzuziehen. Nun kann wie in Abschnitt 2.3.2 mit dieser Komponente Routen erstellt werden. Die URIs für die Konfiguration der Endpunkte mit SOCC-Camel habe dazu den folgenden Aufbau:

Listing 4.2: SOCC-Camel Konfigurations URI

```
1 socc://{connectorId}?uri={targetUri}[&{parameter}...]
```

Der Anfang der URI mit „socc://“ sagt Camel, dass es für diesen Endpunkt die Komponente benutzen soll, die vorher unter dem Namen „socc“ registriert wurde - also die `SoccComponent`. Für den Platzhalter „{connectorId}“ muss eine gültige ID eines Connectors sein, dessen Konfigurationsdaten im TripleStore hinterlegt wurden. Über den Parameter „uri“ wird dann die Quellbeziehungsweise Ziel-URI für den Connector angegeben, von der er Beiträge lesen soll. Je nach Endpunkt können dann noch weitere Parameter angegeben werden.

4.4.1 SoccPostPollingConsumer

Wird ein Endpunkt mit der Absicht zum Lesen von Beiträgen erstellt, erzeugt die Klasse `SoccComponent` ein Objekt der Klasse `SoccPostPollingConsumer` und übergibt ihm die Parameter aus der Konfigurations-URI. Da `SoccPostPollingConsumer` sich von der Klasse `ScheduledPollConsumer` ableitet, ist es über dem Parameter `delay` möglich in periodischen Abständen das Lesen von Beiträgen auszuführen. Die Angabe erfolgt dabei in Millisekunden. Der Parameter `limit` entspricht dabei den gleichnamigen Argument der Methode `pollPosts(...)` des `PostReaders` aus Abschnitt 4.3.5. Das dort noch fehlende Argument `since`, für das Datum, ab wann ein Beitrag als neu gilt, holt sich der `SoccPostPollingConsumer` aus den im TripleStore gespeicherten Daten. Da dort in den Strukturdaten gespeichert wird, wann der letzte Beitrag eines Threads oder der letzte Kommentar auf einen Beitrag gespeichert wird. Alle neuen, gelesenen Beiträge werden am Ende in das RDF/XML-Format serialisiert und als Nachricht an Camel übergeben, welches es dann anhand der festgelegten Routen weiterleitet. Damit andere Komponenten diese Nachricht wieder korrekt de-serialisieren können, muss der Nachricht noch ein Header mit dem Schlüssel „Content-Type“ und dem MIME-Type⁶ vom RDF/XML „application/rdf+xml“ mitgegeben werden.

⁶ Multipurpose Internet Mail Extension - Types: <http://tools.ietf.org/html/rfc2046>

4.4.2 SoccPostProducer

Der SoccPostProducer ist das Gegenstück zum SoccPostPollingConsumer. Er ist der Endpunkt der zum Schreiben von Beiträgen in die Plattform des Connectors von der Klasse SoccComponent erzeugt wird. Intern verwendet er den PostWriter des betreffenden Connetors und leitet den Inhalt der Nachricht an diese weiter. Außer der Angabe über die Ziel-URI erhält der SoccPostProducer keine weiteren Parameter über die Konfigurations-URI.

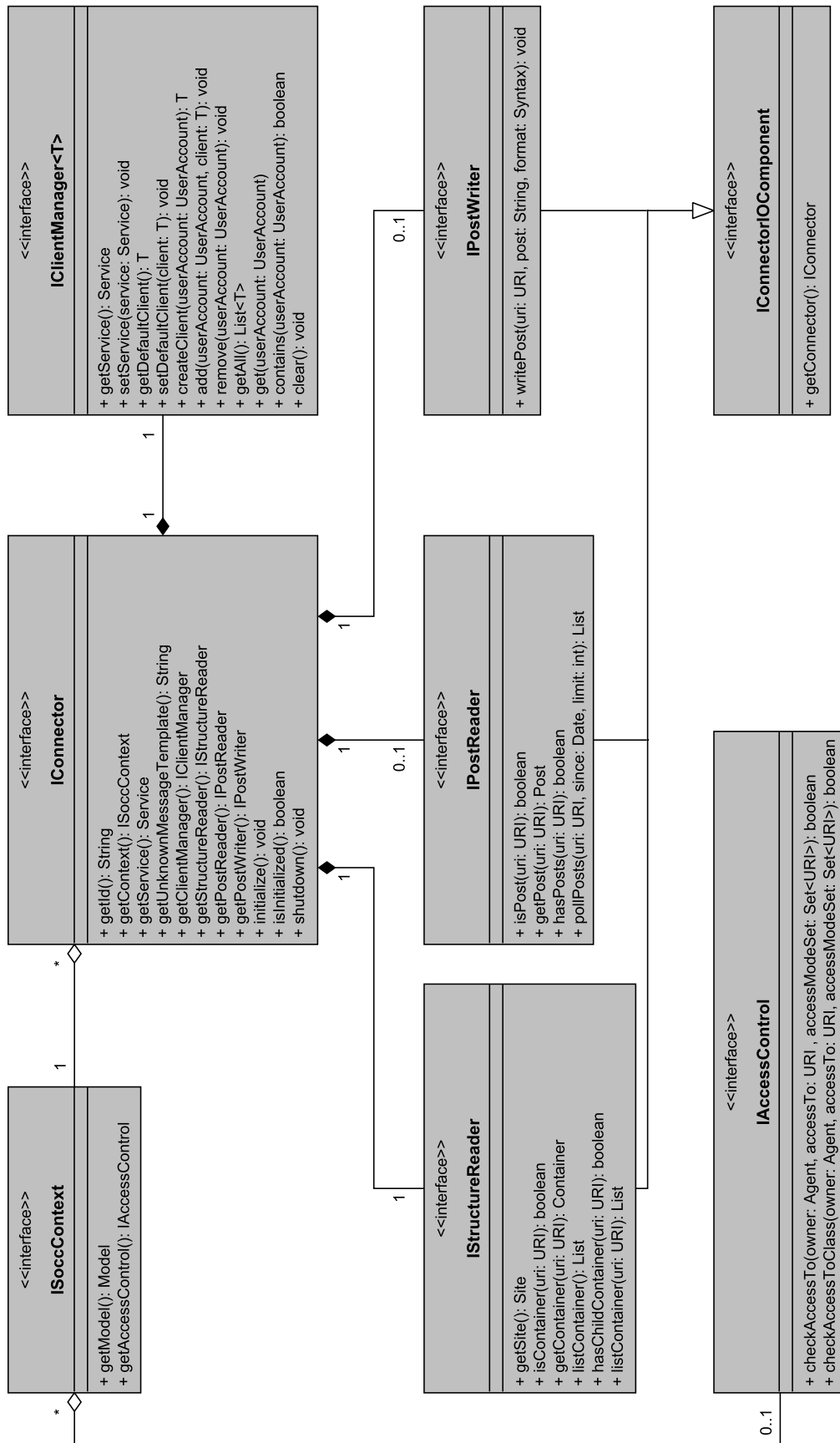


Abbildung 4.7.: UML Klassendiagramm eines Connectors

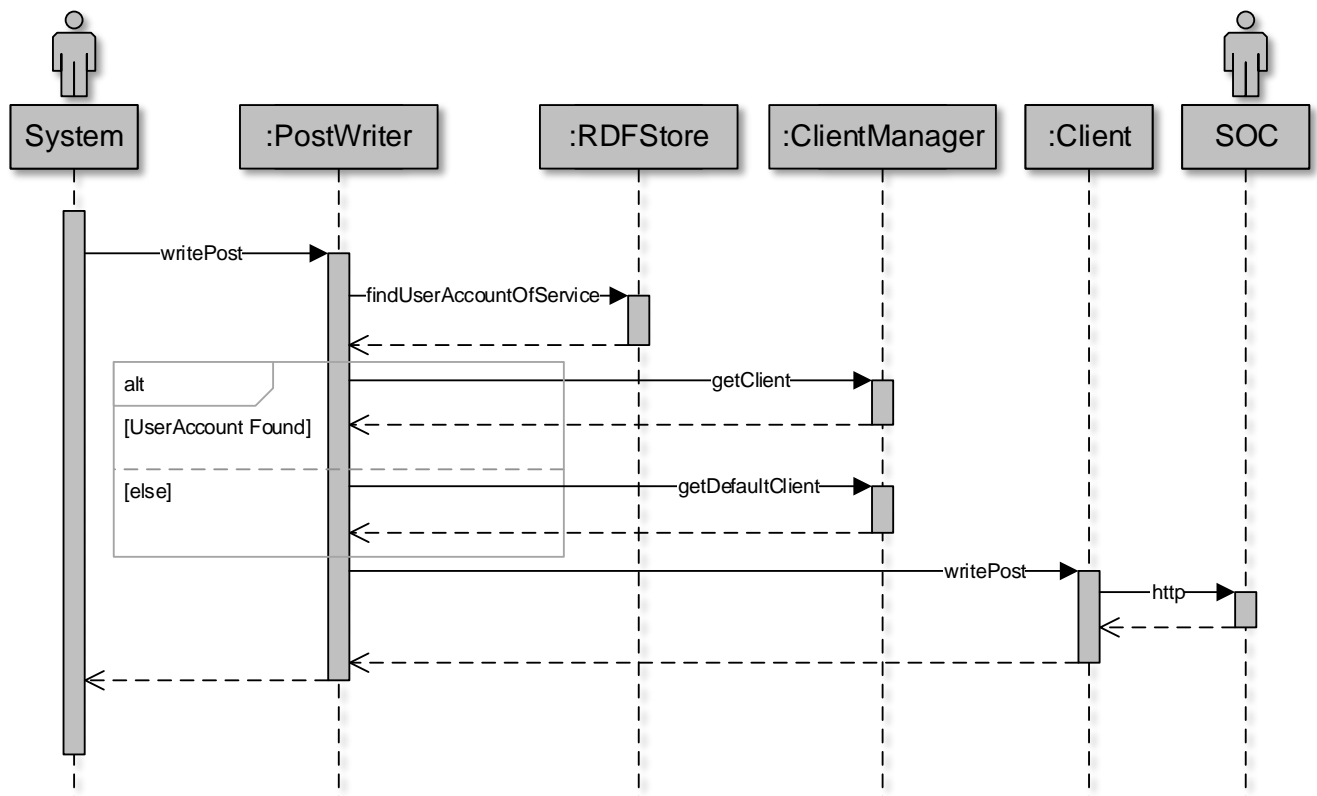


Abbildung 4.13.: UML Sequenzdiagramm eines PostWriters

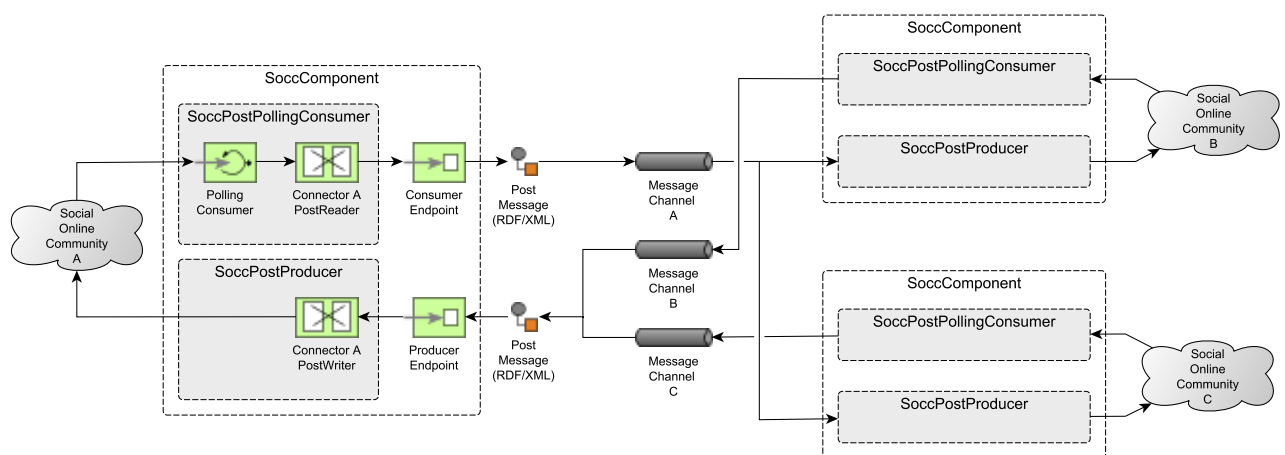


Abbildung 4.14.: Übersicht des Socc-Camel Moduls in EIP-Notation



5 Implementierung und Evaluation

Auf die verschiedenen Klassen von Facebook und Co eingehen, REST, SOAP, OAUTH, Client/Passwort, Art der Plattform, Forum, OSN, Videoplattform

5.1 Verwendete Programme und Bibliotheken

An dieser Stelle soll noch kurz auf zusätzliche Programme und Bibliotheken eingegangen werden, die für die Entwicklung der einzelnen Connectoren hilfreich waren.

5.1.1 Apache Maven

Für die Entwicklung der Programme und Bibliotheken dieser Arbeit wurde das Build-Management Programm Maven¹ von Apache eingesetzt. Mit diesem ist es auf einfache Art möglich Java-Programme zu verwalten und zu erstellen. So ist nur mit einer einzigen XML-Datei (`pom.xml`) die Abhängigkeiten festzulegen, das Programm zu kompilieren, zu testen und es auszuliefern. Maven schreibt eine feste Ordnerstruktur für ein Projekt vor wie der Produktivquellcode, dafür nötige Ressource (Bilder, Tests, ...) und der Code zum Testen geordnet werden sollen. Die Abhängigkeiten werden automatisch aus zentralen Repositories heruntergeladen und ins Projekt eingebunden. So muss dies nicht immer von Hand geschehen und spart so Arbeit.

5.1.2 RDF2Go

5.2 Implementierung der Connectoren

An dieser Stelle soll beschrieben werden wie Connectoren für die in Abschnitt 2.4 vorgestellten Plattformen implementiert werden können. Dafür wird darauf eingegangen mit welcher Art von API auf die Daten der einzelnen Plattformen zugegriffen werden und wie man deren Datenstruktur in SIOC abbilden kann. Ebenfalls werden auftretende Probleme und wenn möglich deren Lösung gezeigt.

5.2.1 Allgemeine Informationen zum Mapping nach SIOC

Innerhalb der Beschreibungen des Mappings nach SIOC werden zur besseren Übersicht und einfacheren Lesbarkeit Platzhalter für einige URIs benutzt. Welche Platzhalter dies sind und für welche URI sie stehen ist aus Tabelle 5.1 zu entnehmen.

¹ <http://maven.apache.org/>

Tabelle 5.1.: Allgemeine Platzhalter und deren Beschreibung für das Mapping nach SIOC

Platzhalter	Bedeutung
{rootUri}	Die WurzelUri einer SOC. Für Facebook wäre dies zum Beispiel <code>https://www.facebook.com</code>
{siteUri}	URI für eine SIOC Site
{serviceUri}	URI für einen SIOC Service
{userAccountUri}	URI für einen SIOC UserAccount
{forumUri}	URI für ein SIOC Forum
{threadUri}	URI für ein SIOC Thread
{postUri}	URI für ein SIOC Post

Da einige Eigenschaften der Klassen von SIOC bei allen Plattformen mit ähnlichen Werten belegt werden und sich nur vom Aufruf der API unterscheiden, sollen zuvor für die Klassen UserAccount, Forum, Thread und Post die Eigenschaften beschrieben werden, die überall verwendet werden.

UserAccount: Für die Klasse UserAccount ist die Angabe von vier Eigenschaften wichtig. Die erste ist `sioc:Id` mit der die ID des Benutzerkontos angegeben wird und `sioc:name` für den den Benutzernamen. Aus Kompatibilitätsgründen mit FOAF wird die ID des Benutzers noch mit `foaf:accountName` angegeben und die URI zum verwendeten Service mit `foaf:accountServiceHomepage`. Die Angabe der Daten zu Autorisierung erfolgt mit der Eigenschaft `siocsa:accountAuthentication`.

Forum: Jedes Forum hat seine eigene ID, die wie bei UserAccount unter der Eigenschaft `sioc:id` abgelegt wird. Ebenfalls bekommt jedes Forum einen Namen mit `sioc:name`. Jedes Forum hat eine Site von der es gehostet wird (`sioc:has_host`). Ein Forum dient als Container für Post, die mit der Eigenschaft `sioc:container_of` gekennzeichnet werden. Zugleich kann es auch Threads mit `sioc:parent_of` enthalten.

Thread: Ein Thread enthält ebenfalls wieder eine ID und einen Namen. Enthält nur Beiträger der Klasse Post mit `sioc:container_of` und ist immer ein Teil eines Forums mit `sioc:has_parent`.

Post: Jeder Post hat eine, für die Seite eindeutige ID mit `sioc:id`. Mit der Eigenschaft `sioc:creator` wird der UserAccount des Autors festgelegt und mit `sioc:content` der Inhalt des Beitrags. Besitzt ein Beitrag eine Titel, wird dieser mit `sioc:title` gespeichert. Der Zeitpunkt der Erstellung kommt nach `dcterms:created` und der bei eine Modifikation nach `dcterms:modified`. Beiträge gehören immer zu einen Forum oder Thread welches mit `sioc:has_container` angegeben wird. Auch können Beiträge Kommentare mit `sioc:has_reply` enthalten oder selber Kommentare auf andere Beiträge mit `sioc:reply_of` sein.

5.2.2 Allgemeine Probleme und Lösungen bei der Implementierung

Bei der Implementierung der Connectoren treten ab und zu Probleme auf, die unabhängig von der verwendeten API gelöst werden müssen. Hier sollen diese Probleme und deren Lösung beschrieben werden.

Doppelte Beiträge beim Synchronisieren

Sollen zwei Threads auf unterschiedlichen Plattformen synchronisiert werden, tritt automatisch das Problem auf, dass Beiträge die von A nach B geschrieben wurden beim auslesen von B wieder bei A als „neuer“ Beitrag landen würde, obwohl er ursprünglich von dort kam. Dies würde ohne Eingreifen zu einer Dauerschleife führen, bei der die selben Beiträge immer hin und her geschrieben werden.

Aus diesem Grund wird vor dem Schreiben eines Beitrags eine Art Wasserzeichen an diesen Angebracht. Diese Wasserzeichen besteht aus der Zeichenkette - forwarded by SOCC siteUri -. Hierzu muss jedem Beitrag mit der Eigenschaft `dc:terms:isPartOf` die URI des Site Objekts zu dem der Beitrag gehört, hinzugefügt werden. Wird nun eine Beitrag gelesen, der ursprünglich von der Site des Connectors kam, wird dieser ignoriert. Enthält dagegen der Beitrag schon eine Wasserzeichen mit der URI einer anderen Site, wird kein neues Wasserzeichen hinzugefügt.

Rekonstruktion der Kommentarstruktur

Bei der Synchronisation ist es ebenso wichtig die Struktur von Kommentaren beizubehalten. Wenn Beitrag A2 ein Kommentar auf Beitrag A1 auf Seite A muss diese Struktur beim Synchronisieren mit Seite B mit den Beiträgen B1 und B2 erhalten bleiben.

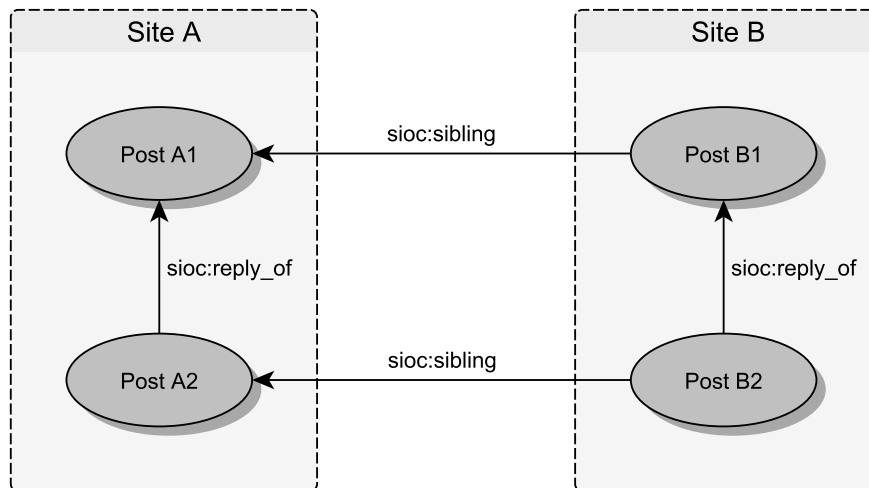
Hierzu wird nach dem Schreiben eines Beitrags mit der Eigenschaft `sioc:sibling` (Englisch für Geschwister) die URI zum original Beitrag beim geschriebenen Beitrag festgehalten. Taucht nun beim Schreiben ein Beitrag auf, der ein Kommentar auf den mit `sioc:sibling` gespeicherten Beitrags ist, kann dieser an die richtige Stelle als Kommentar geschrieben werden. Der RDF-Graph nach dem Synchronisieren ist in Abbildung zu sehen.

5.2.3 Moodle Connector

Moodle Webservice API

Moodle bringt von Haus aus schon eine Webservice-Schnittstelle für einen Zugriff auf seine Daten mit. Diese API unterstützt dabei Protokolle für REST, SOAP und oder XML-RPC². Mit ihr können Kurse, Foren und deren Threads (in Moodle Discussion Topics genannt), Benutzerprofile, Gruppen und viele andere Ressourcen ausgelesen oder geschrieben werden. Leider gilt dies nicht

² <http://xmlrpc.scripting.com/>



@prefix sioc: <http://rdfs.org/sioc/ns#> .

Abbildung 5.1.: Erhalten der Kommentarstruktur beim Schreiben von Beiträgen

für Beiträge innerhalb von Foren Threads. Die aktuelle Version (Beim verfassen dieser Arbeit war dies 2.5.1) erlaubt es nicht Beiträge weder zu lesen noch zu schreiben.

Doch existiert ein Plugin für Moodle, dass auf dieser Webservice API aufbaut und erweitert - MoodleWS³. MoodleWS bietet den Zugriff über REST und SOAP, wobei bei der Verwendung von REST JSON als Datenformat verwendet wird und nicht XML wie mit SOAP. Mit diesem Plugin ist ein Programm nun auch in der Lage Beiträge aus Threads zu lesen. Die REST-Schnittstelle hat aber noch eine paar Fehler und so können damit noch keine neuen Beiträge geschrieben werden. Mit SOAP ist dies aber problemlos möglich.

Für die Benutzung der muss sich ein Client mit dem Benutzernamen und Passwort eines vorhandenen Benutzers anmelden. Ist dies erfolgreich, bekommt der Client einen zufällig ID und einen Sitzungsschlüssel die er bei jeder Abfrage angeben muss. Dieser Sitzungsschlüssel ist nur eine begrenzte Zeit gültig, so kann es passieren, dass eine Operation fehlschlägt und der Client sich neu anmelden muss.

Für die Verwendung der API in Java wird auch eine Bibliothek moodle_ksoap2⁴ zur Verfügung gestellt. Sie abstrahiert komplett das SOAP-Protokoll im Hintergrund und enthält für alle Moodle-Ressource Java-Klassen und Methoden für den Zugriff. Bevor moodlews_ksoap2 benutzt werden kann, ist darauf zu achten in Moodle unter „Site administration“, „Development“, „OK Tech Webservices (aka wspp)“ die Einstellung „Use an auto generated wsdl“ auszuschalten, da es sonst zu Problemen mit moodlews_ksoap2 kommen kann.

Listing 5.1: MoodleWS mit moodlews_ksoap2: Lesebeispiel

```

1 Mdl_soapserverBindingStub client = new Mdl_soapserverBindingStub(
2     MOODLE_URI + "/wspp/service_pp2.php",

```

³ <https://github.com/patrickpollet/moodlews>

⁴ https://github.com/patrickpollet/moodlews_ksoap2

```

3      MOODLE_URI + "/wspp/wsdl2",
4      false );
5
6  LoginReturn loginReturn = client.login(username, password);
7
8  ForumRecord[] forumRecords = client.get_all_forums(
9      loginReturn.getClient(),
10     loginReturn.getSessionKey(),
11     "",
12     "" );
13
14  for(ForumRecord forum : forumRecords){
15      System.out.println(forum.getId() + " " + Forum.getName());
16  }

```

Wie in der ersten Zeile von Listing 5.1 zu sehen wird als ist der Client eine Objekt der Klasse `Mdl_soapserverBindingStub` aus `moodlews_ksoap2`. Dem Konstruktor wird die URI auf die Datei „service_pp2.php“ (Zeile 3) und der XML-Namespace der WSDL-Datei des MoodleWS-Plugins übergeben. Das dritte Argument besagt, ob der Client im Debugmodus laufen und mehr Logging-Ausgaben macht über die einzelnen internen Abläufe ausgeben soll. In Zeile 6 meldet sich der Client mit einem Benutzernamen und einem Passwort bei MoodleWS an und bekommt ein Objekt der Klasse `LoginResult` mit der oben angesprochenen ID für den Client und einen Sitzungsschlüssel zurück. Mit der Methode `get_all_forums(...)` können nun ein Array mit allen Foren angefordert werden, auf die der angemeldete Benutzer Zugriff hat. Die ersten zwei Argumente dieser Methode sind die ID des Clients und der Sitzungsschlüssel. Mit den letzten zwei Argumenten kann das Ergebnis gefiltert werden. Dazu wird als erstes ein Spaltenname aus der Foren-Tabelle der Moodle-Datenbank angeben und als zweites welcher Wert dieser haben soll. Am Ende werden dann alle Ergebnisse mit der ID und Name des Forums ausgegeben.

Listing 5.2: MoodleWS mit `moodlews_ksoap2`: Beitrag schreiben

```

1  ForumPostDatum postDatum = new ForumPostDatum(client
2      .getBindingStub()
3      .getNAMESPACE());
4
5  postDatum.setMessage("War kam bei der Aufgabe raus?");
6  postDatum.setSubject("Aufgabe 2.1");
7
8  client.forum_add_reply(
9      client.getAuthClient(),
10     client.getSessionKey(),
11     postId,
12     postDatum);

```

Beiträge können nur als Kommentare geschrieben werden und dessen ID bekannt sein. Alle Daten für den Beitrag werden in einem Objekt der Klasse `ForumPostDatum` gespeichert, wie in der ersten Zeile des Listings 5.2 zu sehen ist. Das Argument ist der beim Client angegebene

XML-Namespace der WDSL-Datei. Als Daten können wie in Zeile 5 und 6 nur eine Nachricht (Message) und ein Thema (Subject) übergeben werden. Dieses Objekt wird dann an die Methode `forum_add_reply` mit der ID des Clients, dem Sitzungsschlüssel und der ID des Beitrags wohin geschrieben werden soll übergeben und an Moodle übertragen.

Moodle Mapping nach SIOC

Da Beiträge in Moodle in einer Forumsstruktur geschrieben werden, ist das Mapping nach SIOC problemlos möglich. Die Struktur von Moodle besteht aus der Klasse `Forum` und kann der gleich lautenden Klasse aus SIOC zugeordnet werden. Jedes Moodle Forum besteht aus einzelnen Objekten der Klasse `ForumDiscussions` welche der Klasse `Thread` aus SIOC entsprechen. Letztendlich werden die Beiträge mit der Klasse `ForumPost` aus Moodle modelliert, welche der Klasse `Post` aus SIOC zugeordnet werden kann.

Die Tabelle 5.2 zeigt das Format der URIs, mit denen die Objekte der SIOC-Klassen in RDF benannt werden.

Tabelle 5.2.: URI Format für SIOC-Objekte aus Moodle

SIOC Klasse	URI-Format
Site	{rootUri}
Service	{rootUri}/wspp
UserAccount	{rootUri}/user/profile.php?id={userId}
Forum	{rootUri}/mod/forum/view.php?id={forumId}
Thread	{rootUri}/mod/forum/discuss.php?d={discussionId}
Post	{rootUri}/mod/forum/discuss.php?d={discussionId}#p{postId}

Besonderheiten bei der Implementierung des MoodleConnectors

Der Connector für Moodle war wohl einer der herausforderndsten bei der Implementieren von allen fünf Connectoren. Nicht nur weil erst nach mehreren Test sich herausstellte, dass die favorisierte REST-Schnittstelle von MoodleWS kein schreiben von Beiträgen möglich macht sondern dass auch die Java-Bibliothek `moodlews_ksoap2` so ihre kleinen Eigenheiten hat.

Zuerst mussten die Klasse für `moodlews_ksoap2` komplett neu generiert werden. In der Dokumentation wird beschrieben, dass dieser Schritt nötig ist, wenn sich an der WSDL-Datei von MoodleWS was geändert hat. Der passende Generator wird vom Entwickler gleich mitgeliefert und ist mit dem folgenden Befehl schnell bewerkstelligt.

Listing 5.3: Neugenerierung von `moodlews_ksoap2`

```
1 java org.ksoap2.wsdl.WSDL2ksoap -p de.m0ep.moodlews.soap -o /tmp
http://localhost/moodle/wspp/wsdl_pp2.php
```

Mit dem Parameter „-p“ wird das Java-Package für die Java-Klassen festgelegt und mit „-o“ der Ordner in dem die Klassen geschrieben werden. Die URI am Ende gibt den Pfad zur Datei „wsdl_pp2.php“ vom MoodleWS-Plugin. Diese ist dann an die jeweilige Moodleinstanz anzupassen.

Eine sehr nervende Eigenheit von moodlews_ksoap2 ist die Tatsache, dass Java-Exceptions und Gründe für Fehler gerne verschluckt und nicht weitergegeben werden. So ist es zum Beispiel nicht möglich zu wissen ob eine Operation aufgrund eines Netzwerkfehlers, eines abgelaufenen Sitzungsschlüssels oder falschen Daten fehlschlug. Aus diesem Grund wurde der Client von moodlews_ksoap2 für den ClientManager (siehe Abschnitt 4.3.3) in eine extra Klasse Moodle2ClientWrapper gepackt und um Mechanismen zur Fehlererkennung und Wiederherstellung erweitert. Alle Funktionsaufrufe des Clients wie `get_all_forums(...)` werden dazu zuerst in die Methode `call` eines Objekts des Java-Interfaces `Callable<V>` verpackt, so dass diese Funktion, ohne zu wissen um welche es sich genau handelt, mehrfach aufrufen werden kann. Dieses Objekt wird dann an die Methode `callMethod(...)` der Klasse `Moodle2ClientWrapper` übergeben. Diese Methode ruft dann zuerst die `call()` Methode des `Callable`-Objekts und prüft ob ein gültiges Ergebnis zurückgegeben wurde. Dies ist der Fall wenn es ungleich `null` ist. Ist dies nicht der Fall, wird geschaut ob die Moodleinstanz erreichbar ist. Sollte hier alles in Ordnung sein, muss geschaut werden ob der Sitzungsschlüssel abgelaufen ist. Als Test reicht hier es hier aus, ob die Methode `get_my_id(...)` des moodlews_ksoap2 Clients eine Zahl ungleich 0 (`Null`) zurück gibt. Diese liefert die ID des aktuell angemeldeten Benutzers zurück und im Fehlerfall ist dieser gleich 0. Ist dieser Test positiv, also der Sitzungsschlüssel abgelaufen, wird versucht sich neu anzumelden und die Anfrage erneut ausgeführt.

Eine Funktion mit der nur ein Beitrag mit einer bestimmten ID abgefragt werden könnte fehlt ebenfalls. Wenn nur ein Beitrag geladen werden soll, müssen zunächst alle innerhalb des selben Threads angefordert und diese nach dem passenden Beitrag durch werden.

Ebenfalls ungewöhnlich ist der Inhalt der Klasse `ForumDiscussionRecord`. Normalerweise würde man erwarten, dass die Methode `getId()` die ID des Diskussionsthreads zurück geben wird. Doch stattdessen entspricht der Wert der ID des ersten Beitrags dieses Threads. An die richtige ID kommt man erst dann, wenn man sich mit der Methode `getPost()` ein Objekt des ersten Beitrags und darin den Rückgabewert von `getDiscussion()` benutzt.

5.2.4 Facebook Connector

[

Keine Schreiben mit fremden Benutzerkonten auf die Wall eines Benutzers]

Facebook Graph API

Die primäre Weg an die von Facebook gespeicherten Daten zu gelangen, geht über die *Graph API*⁵. Sie erlaubt einen REST basierten Zugriff auf den von Facebook sogenannten *Social Graph* [15]. Dieser Graph enthält alle Beziehungen die ein Person auf Facebook zwischen anderen Personen,

⁵ Graph API Dokumentation: <https://developers.facebook.com/docs/reference/api>

Gruppen, Seiten, Beiträge, Fotos und so weiter besitzt. Der einfachste Weg zum Kennenlernen der Graph API ist der *Graph API Explorer*⁶. Er ist eine Onlineanwendung REST-Abfragen an die Graph API schnell ausprobiert und das Ergebnis begutachtet werden kann.

Jede Ressource im Social Graph von Facebook hat seine eigen, eindeutige ID. Wegen diesem Merkmal besteht eine URI für den REST-Zugriff nur aus `https://graph.facebook.com/` gefolgt von der ID und optionalen Parametern. Dies macht die Form der URIs zwar sehr einheitlich, aber es ist so unmöglich anhand der URI herauszufinden welcher Typ von Ressource sich dahinter befindet. Dies widerspricht zwar nicht die am Anfang von Kapitel 4 beschreiben Prinzipien von Berners-Lee macht aber die spätere Verarbeitung etwas komplizierter.

Graph API Datenformat

Also Datenformat setzt Facebook mit der Graph API auf das unter Webanwendungen weit verbreitete JSON. Neben den eigentlichen Daten enthalten Ressourcen noch Verbindungen zu anderen Ressourcen. Diese Verbindungen werden „Connections“ genannt. Um zu lesen welche Connections eine Ressource besitzt muss bei der Abfrage der Parameter `metadata=1` angehängt werden. Ist die geschehen enthalten die JSON Daten einen neuen Eintrag `metadata` und darin unter dem Eintrag `connections` eine Liste der vorhandenen Connections. Die wichtigsten Connections für diese Arbeit sind `feed` und `comments`. Die Connection `feed` sagt über die Ressource aus, dass sie eine Liste von Beiträgen enthält. Das könnten zum Beispiel die Wall einer Person oder die einer Gruppe sein. Dahingegen zeigt die Connection `comments`, dass die betreffende Ressource Kommentare enthält und auch kommentiert werden kann. Um die Beiträge eines Feeds lesen zu können, muss nur die URI zu der Ressource mit den Pfad `/feed` erweitert werden: `http://graph.facebook.com/me/feed`. Die ID `me` ist dabei ein spezielle ID die sich auf den aktuell angemeldete Benutzer bezieht. Möchte man nun nicht alle Daten einer Ressource abfragen sondern zum Beispiel nur den Name und das Datum der letzten Änderung, kann der Parameter `fields` für die URI eingesetzt werden. Als Wert wird dann eine Liste mit den Namen der gesuchten Daten angegeben, wobei jeder Name durch ein Komma getrennt wird.

Facebook Login

Die Graph API verwendet zur Autorisierung der Anfragen OAuth 2.0 auch *Facebook Login*⁷ genannt. Accesstoken gelten bei Facebook nur für zwei Stunden. Es ist aber möglich einen sogenannten *Extended Accesstoken*⁸ zu erstellen der zwei Monate gültig ist. Ist ein Accesstoken irgendwann abgelaufen, muss ein neuer erstellt werden. Da aber die Graph API hauptsächlich nur für Webanwendungen gedacht ist, kann dies ein Desktopprogramm nur mit dem Umweg über einen Internetbrowser machen. Für das Anlegen eines neuen Accesstokens werden vom Programm eine OAuth 2.0 ClientID und das ClientSecret benötigt. Diese bekommt man von Facebook, wenn man für sein Programm auf der Webseite `https://developers.facebook.com/apps` eine „App“ anlegt. Die Daten befinden sich dann in den „Einstellungen“ unter „Grundlegendes“ und werden dort „App ID“ und „App Secret“ genannt.

⁶ Graph API Explorer: <https://developers.facebook.com/tools/explorer>

⁷ Facebook Login: <https://developers.facebook.com/docs/facebook-login/>

⁸ Graph API Accesstokens: <https://developers.facebook.com/docs/facebook-login/access-tokens/>

Jeder Accesstoken hat nur Zugriff auf bestimmte Bereiche der Graph API die beim Erstellen des selbigen als „Permission“⁹ (engl. für Erlaubnis oder Befugnis), beziehungsweise von OAuth 2.0 „Scope“ (engl. für Geltungsbereich) genannt, angegeben werden muss. Die für SOCC wichtigen Permissons sind in der Tabelle 5.3 beschrieben.

Tabelle 5.3.: Für SOCC wichtige Facebook Permissons

Permission	Beschreibung
<i>read_stream</i>	Lesen von Beiträgen und Kommentaren eines Benutzers
<i>publish_actions</i>	Schreiben von Beiträgen und Kommentaren für einen Benutzer
<i>user_groups</i>	Zugriff auf die Lister der beigetretenen Gruppen eines Benutzers

Um nun einen Accesstoken für einen Benutzer zu Erstellen muss in einen Internetbrowser folgende URI aufgerufen werden:

Listing 5.4: Facebook Login: OAuth URI

```
1 https://www.facebook.com/dialog/oauth?client_id={appId} &
   redirect_uri={redirectUri}&responseType=code&scope={scopeList}
```

Anstatt den Platzhalter {appId} muss dann natürlich die oben angesprochene ClientID beziehungsweise App ID eingetragen werden. Mit dem Parameter redirect_uri muss bei dieser URI eine Adresse angegeben werden, auf die der Internetbrowser nach der Anfrage mit der Antwort umgeleitet wird. Dies kann die Adresse eines externen Webservers oder die eines Webservers den die Anwendung nur für diesen Zweck laufen lässt. Einen solchen extra Webserver könnte man mit der Jetty¹⁰ Bibliothek erstellen und würde dann als redirect_uri die URI http://localhost:port/ angeben. Der Parameter responseType=code besagt, dass diese Anfrage einen Code zurück liefert mit dem dann der eigentliche AccessToken geholt werden kann. Dies ist der Standardweg. Mit dem letzten Parameter scope können dann noch die oben erwähnten Permissons übergeben werden, wobei alle durch eine Komma getrennt hintereinander zu schreiben sind. Wird dann diese URI im Browser aufgerufen, wird er danach bei Erfolg auf die Adresse redirectUri?code=code umgeleitet. Der Platzhalter {code} würde dann einen beliebigen Code darstellen, mit dem über eine weitere Anfrage mit der folgenden URI der Accesstoken geholt werden kann:

Listing 5.5: Facebook Login: URI für den Tausch von Code gegen Accesstoken

```
1 https://www.facebook.com/oauth/access_token?client_id={appId}&
   redirect_uri={redirectUri}&client_secret={appSecret}&code={code}
```

Der Platzhalter {appId} und {appSecret} entsprechen wieder der ClientID und Client Secret. Bei dieser URI ist es sehr wichtig, dass {redirectUri} den selben Wert hat, wie er weiter oben angegeben wurde. Für {code} muss dann noch der oben zurückgelieferte Code angegeben

⁹ Graph API Permissions: <https://developers.facebook.com/docs/reference/login/#permissions>

¹⁰ Jetty Webserver: <http://www.eclipse.org/jetty>

werden. Die Anfrage mit dieser URI muss dann nicht über einen Internetbrowser geschehen sondern kann normal aus einem Programm erfolgen. Als Antwort wird dem Programm dann `access_token=accessToken&expires=secondsTilExpiration` zurück geliefert. Der Wert des Parameters `access_token` ist der gewollte Accesstoken und der Wert von `expires` ist die Haltbarkeit des Accesstokens in Sekunden.

Soll nun die Haltbarkeitsdauer des Tokens auf zwei Monate verlängert werden muss eine neue Anfrage an die folgende URI gemacht werden:

Listing 5.6: Facebook Login: Extended Accesstoken

```
1 https://www.facebook.com/oauth/access_token?grant_type=
  fb_exchange_token&client_id={appId}&client_secret={app-secret}&
  fb_exchange_token={accessToken}
```

Die Werte für die einzelnen Parameter dürften nun offensichtlich sein. Als Antwort kommt dann ein Accesstoken mit verlängerter Haltbarkeit zurück. Dabei kann es passieren, dass er genau so aussieht wie der alte, was aber kein Problem darstellt.

RestFB

Facebook selber bietet direkt keine API für die Verwendung der Graph API für Desktopprogramme an. Speziell für die Programmiersprache Java existiert von Facebook nur eine API für das mobile Betriebssystem Android¹¹. Als Alternative dazu kann die *RestFB*¹² Java-Bibliothek von Mark Allen verwendet werden. RestFB bietet dabei die Möglichkeit von Java aus die Graph API von Facebook zu benutzen. Eine Abfrage könnte wie in Listing 5.7 aussehen.

Listing 5.7: RestFB Beispielprogramm

```
1 FacebookClient client = new DefaultFacebookClient(USER_ACCESS_TOKEN);
2
3 User user = client.fetchObject("me", User.class);
4 System.out.println("User name: " + user.getName());
5
6 Connection<Post> myFeed = client.fetchConnection("me/feed", Post.
  class);
7 for (List<Post> feedPage : myFeed){
8     for (Post post : feedpage){
9         System.out.println("Post: " + post);
10    }
11 }
```

In Zeile 1 wird eine Objekt der Klasse *FacebookClient* mit einem Accesstoken als Argument erstellt über dem die Methoden für die Abfragen aufgerufen werden können. Die Zeile 3 zeigt eine solche Abfrage nach den Benutzerdaten für den aktuell angemeldeten Benutzer mit der speziellen ID „me“. Der Anfangsteil der URI für die Graph API `https://graph.facebook.com/`

¹¹ Facebook Android API: <https://developers.facebook.com/docs/android/>

¹² RestFB Dokumentation: <http://restfb.com/>

muss dabei weggelassen werden. Zurück gegeben wird ein Objekt der Klasse User und damit wird dann der Benutzername ausgegeben. Die Angabe der Klasse mit `User.class` ist ebenfalls wichtig, so dass RestFB weiß in welche Klasse es die von der Graph API zurückgelieferten JSON-Daten konvertieren soll. Zeile 6 zeigt noch wie man an die Daten einer Connection, hier der Beiträge auf der Wall des aktuellen Benutzers, gelangen kann. Enthält eine Antwort eine Liste mit zu vielen Einträgen, wird die Antwort auf mehrere Seiten aufgeteilt und ein Verweis auf die nächste Seite der aktuellen Seite mitgegeben. Die Klasse Connection sorgt dann alleine dafür, dass die neue Seite gelesen wird, wenn die alte abgearbeitet wurde. Die erste For-Schleife in Zeile 7 iteriert dann über alle vorhandenen Seiten und die zweite in Zeile 8 über alle Einträge auf der aktuellen Seite.

Listing 5.8: RestFB: Schreiben von Beiträgen auf eine Wall

```
1 FacebookType result = client.publish(  
2     userId + "/feed",  
3     FacebookType.class,  
4     Parameter.with("message", "Tolles Wetter heute \o/"));
```

Beiträge können bei Facebook nur an Ressourcen geschrieben werden die eine „feed“ oder „comment“ Connection haben. In Listing 5.8 ist zu sehen wie ein neuer Beitrag auf die Wall eines Benutzers geschrieben werden kann. Zu diesem Zweck enthält der Client aus RestFB die Methode `publish(...)` zum Schreiben von Daten nach Ressourcen. Das erste Argument ist die Zielressource, die in Zeile 2 aus der ID des Benutzers gefolgt von einem „/“ und dem Namen der Connection, hier „feed“. Da `publish(...)` bei Erfolg die ID des erstellten Beitrags zurückliefert, wird als Klasse für das erwartete Ergebnis die Basisklasse `FacebookType` angegeben. Als letztes muss noch der Inhalt des Beitrags festgelegt werden. Dies geschieht nicht im JSON-Format sondern wird als Parameter in der URI für die HTTP-POST-Operation angegeben. RestFB bietet dazu die Klasse `Parameter` an, mit der ein Parameter mit dem Namen „message“ und dem Inhalt als Wert erzeugt wird.

Facebook Mapping nach SIOC

Besonderheiten bei der Entwicklung des FacebookConnectors

Wie im Abschnitt über die Graph API schon beschrieben, ist es nahezu unmöglich an der URI einer Ressource von Facebook zu erkennen welche Typ sich dahinter befindet. Deshalb muss bei jeder unbekannten URI erst einmal alle Informationen mit dem Parameter `metadata=1` heruntergeladen werden und anhand der vorhandenen Connections ermittelt werden, ob es sich um einen Beitrag oder um einen Container handelt. Deshalb ist es wichtig solche Erkenntnisse sofort in den Triplestore zu speichern um weitere Abfragen zu vermeiden.

5.2.5 Google+ Connector

Die API von Google für den Zugriff auf sein soziales Online-Netzwerk Google+¹³ baut wie viele auf andere auf den REST-Prinzipien auf. Die URIs für die Anfragen fangen mit `https://www.googleapis.com/plus/v1/` an und danach folgt der Pfad zu den Ressourcen und etwaige Parameter. Die für hier wichtigsten Parameter sind `access_token` zum Angeben eines OAuth 2.0 Accesstoken und `fields` mit dem nur die dort angegebenen Datenfelder im Ergebnis zurück geliefert werden. Als Datenformat wird, wie schon bei Facebook, JSON eingesetzt, aber mit einen von Facebook vollkommen unterschiedlichen Aufbau. Activitys und Comments bestehen im Grunde aus drei Teilen:

Actor: Der Actor-Eintrag sagt über die Ressource aus von wem sie erstellt wurde.

Verb: Das Verb beschreibt auf welche Weise diese Ressource erstellt wurde. Für Activitys ist die Angabe von „post“, wenn ein Beitrag vom Actor selber geschrieben oder „share“, wenn dieser nur geteilt wurde und von jemand anderem stammt.

Object: Der Object-Eintrag enthält den eigentlichen Inhalt.

Neben diesen drei enthält eine Ressource noch Metadaten wie ID, Zeitpunkt der Veröffentlichung oder Änderung, Ortsdaten und andere mehr.

Die Google+ API teilt zu lange Ergebnislisten, ebenfalls wie Facebook, in mehrere Teillisten die auf einzelne Seiten aufgeteilt werden. Ob eine Seite der aktuellen folgt, kann über die Existenz des `nextPageToken` Eintrags im JSON-Objekt erkannt werden. Dieser *PageToken* muss für die Abfrage der nächsten Seite an die gleiche URI als Parameter `pageToken` angehängt werden.

Wie schon erwähnt setzt die Google+ API für die Autorisierung auf OAuth 2.0. Ein Accesstoken ist aber nur nötig, wenn Operationen stellvertretend für einen Benutzer ausgeführt werden müssen. Für alle anderen Fälle ist ein API-Schlüssel vollkommen ausreichend. Zusätzlich zum Accesstoken liefert Google+ auch einen Refresh token mit. So können abgelaufene Accesstoken vom Programm automatisch wieder reaktiviert werden, ohne dass der Benutzer dies von Hand machen muss. Die für die Erstellung von Access- und Refresh token gebrauchten ClientID und ClientSecret können über die Google API Console¹⁴ erstellt werden. Wichtig dabei ist eine „Client ID for installed applications“ auszuwählen.

Sowohl für den Zugriff auf die REST-API, als auch für die Erstellung der Token, stellt Google eine Java-Bibliothek zur Verfügung.

Listing 5.9: Google+ API Java: Access- und Refresh token

```
1 MemoryCredentialStore credentialStore = new MemoryCredentialStore();
2 GoogleAuthorizationCodeFlow flow = new GoogleAuthorizationCodeFlow.
   Builder(
3     new NetHttpTransport(),
4     new JacksonFactory(),
5     CLIENT_ID,
6     CLIENT_SECRET,
```

¹³ Google+ API Dokumentation: <https://developers.google.com/+/api/>

¹⁴ <https://code.google.com/apis/console>

```

7         Arrays.asList("https://www.googleapis.com/auth/plus.login"))
8         .setCredentialStore(credentialStore)
9         .build();
10
11 Credential credentials = new AuthorizationCodeInstalledApp(
12     flow,
13     new LocalServerReceiver())
14     .authorize("user");

```

Listing 5.9 zeigt wie man sich Access- und RefreshToken für Google+ in Java holen kann. In der ersten Zeile wird ein `MemoryCredentialStore` erstellt, mit dem die Token im flüchtigen Speicher des Systems zwischen gespeichert werden können. Es gibt aber auch Klassen die alles in einer Datei auf der Festplatte ablegen. So können sie auch zwischen Programmstarts benutzt werden. Von Zeile 2 bis 9 wird ein sogenannter „`AuthorizationCodeFlow`“ erzeugt. Dieser abstrahiert für den Benutzer den OAuth-Mechanismus über den man die Token von Google bekommt. In der Zeile 3 und 4 bekommt dieser `CodeFlow` externe Objekte übergeben mit dem er HTTP-Operationen ausführen und die zurückgelieferten Daten verarbeiten kann. In den zwei Folgezeilen werden die `ClientID` und das `ClientSecret` übergeben. In der siebten Zeile muss, wie schon bei Facebook, ein Geltungsbereich (Scope) für den AccessToken festgelegt werden. In diesem Falle wollen wir mit `https://www.googleapis.com/auth/plus.login` Operationen stellvertretend für einen Benutzer ausführen. In Zeile 8 wird noch der oben erstellte `textttMemoryCredentialStore` übergeben und das `Code-Flow-Objekt` in Zeile 9 über die `build()` Methode erzeugt. Für Desktopprogramme wird wieder ein externer Internetbrowser zum Holen der Token benötigt. Doch auch dafür bietet Google eine fertige Klasse an, die (fast) alle Schritte automatisiert. Diese Klasse ist `AuthorizationCodeInstalledApp`. Ihr werden der obige `CodeFlow` und ein lokaler Webserver (ebenfalls schon fertig vorhanden) übergeben und über die Methode `authorize(...)` die Token erzeugt. Die Zeichenkette „user“ legt dabei fest, unter welchen Namen die Token im `textttMemoryCredentialStore` abgelegt werden.

Listing 5.10: Google+ API: Zugriff auf Activity-Feed mit Java

```

1 Plus plus = new Plus.Builder(
2     new NetHttpTransport(),
3     new JacksonFactory(),
4     credential)
5     .setApplicationName(APPLICATION_NAME)
6     .build();
7
8 ActivityFeed activityFeed = plus.activities()
9     .list( "me","public" )
10    .execute();
11
12 for(Activity activity : activityFeed.getItems()){
13     System.out.println(
14         activity.getActor().getDisplayName()
15         + ": "
16         + activity.getObject().getContent());

```

Das kleine Beispiel in Listing 5.10 zeigt wie auf einen Activity-Feed, also die Liste der geschriebenen Beiträge eines Benutzers, in Google+ mit Java zugegriffen werden kann. Von Zeile 1 bis 6 wird das Clientobjekt für den Zugriff auf Google+ mit einem Builder¹⁵ gebaut. Als Argument wird wieder das Objekt einer Klasse für die Verwendung des HTTP und eines für das Arbeiten mit JSON übergeben. Als dritter Parameter in Zeile 4 kommen noch die Credentials aus Listing 5.9 mit dazu. Mit der Methode `setApplicationName(...)` kann noch ein Name für die Anwendung angegeben werden bevor der Client mit der Methode `build()` in Zeile 6 erstellt wird. Das Clientobjekt enthält mehrere Methoden mit denen man auf die verschiedenen Ressource wie Activitys, Kommentare und Personen zugreifen kann. In Zeile 8 wird die Methode `activities()` eingesetzt, um an eine Hilfsklasse für alle Activitys zu gelange, welche die Operationen zum Lesen eines Activity-Feeds, nur eines einzelnen Activitys oder der Suche nach anderen Activitys. Mit `list("me", "public")` wird eine Abfrage für den „public“ Activity-Feed (nur dieser ist aktuell verfügbar) des Benutzers mit der ID „me“ (Steht immer für den aktuell angemeldeten Benutzer) erstellt, die dann mit dem Aufruf der Methode `execute()` des zurückgelieferten Objekts ausgeführt wird. Das Ergebnis dieser Abfrage enthält ein Array von Activitys, das von der Methode `getItems()` zurückgegeben wird. Zum Schluss werden von allen Activitys der Name des Autors und der Inhalt ausgegeben. Ist ein oben erwähnter Pagetoken für eine weiter Ergebnisseite vorhanden, kann dieser an das Objekt, was von `list("me", "public")` zurückgegeben wird, mit der Methode `setPageToken(...)` übergeben werden.

Leider ist es mit der aktuellen Version der Google+ API nur möglich auf Ressourcen zuzugreifen, die von den Benutzer auf als „öffentlich“ markiert wurde. Zugleich sind keine Operationen vorhanden um irgendwelche Beiträge zu Schreiben.

Google+ Mapping nach SIOC

SIOC Klasse	URI-Format
Site	
Service	
UserAccount	
Forum	
Thread	
Post	

5.2.6 Youtube Connector

¹⁵ „Builder“ Entwurfsmuster: <http://www.oodesign.com/builder-pattern.html>

Für Youtube existieren zur Zeit zwei verschiedene APIs, eine Version 3 und die ältere Version 2. Die *Youtube Data API v3* befindet sich noch im Entwicklungsstadium. Dort ist es zum Beispiel noch nicht möglich ist Kommentare von Video zu lesen, geschweige denn zu schreiben. Aus diesem Grund muss für einen YoutubeConnector die älter *Youtube Data API v2* benutzt werden, welche diese Funktionen noch bietet.

Die Youtube Data API v2 baut noch auf dem Google Data Protokoll auf, eine API nach dem REST-Prinzip mit Atom und JSON als Datenformat. Für Youtube wird aber primär Atom eingesetzt (umschalten auf JSON ist möglich). Die einzelnen Ressourcen können als Atom-Feed, also eine Liste einzelnen Einträgen wie die einer Playliste, oder als ein einzelnes Entry-Element, zum Beispiel Informationen zu einen Benutzerkonto, zurückgegeben werden. Youtube verwendet aber nicht nur die vom Atom-Standard definierten Element, sondern erweitert das Format durch eine große Zahl eigener Erweiterungen wie Spieldauer eines Videos, Statistiken, Bewertungen oder Informationen zu Benutzerkonten. Jeder Feed/Eintrag enthält mehrere Link-Element die auf weiterführenden Ressourcen verweisen. Ist die Menge der Einträge eines Feedes zu groß, wird das Ergebnis auf mehrere Feeds aufgeteilt. Um zwischen diesen Teilergebnissen zu navigieren enthalten die Teilfeeds Links auf ihren Vorgänger und/oder Nachfolge-Feed (siehe Listing 5.11).

Listing 5.11: Youtube Data API v2: Navigation durch die Ergebnisse

```
1 <link rel='previous' type='application/atom+xml'
2   href='http://gdata.youtube.com/feeds/api/videos?start-index=1&max-
   results=25...' />
3 <link rel='next' type='application/atom+xml'
4   href='http://gdata.youtube.com/feeds/api/videos?start-index=51&max-
   results=25...' />
```

Alle Ressourcen auf die öffentlich zugegriffen werden kann brauchen keine vorherige Anmeldung. Sollen aber zum Beispiel Kommentare geschrieben werden ist eine Anmeldung notwendig. Für der Programm muss ein ein API-Schlüssel (von Google „Developer Key“ genannt) im *Product Dashboard*¹⁶ erstellt werden. Dieser API-Schlüssel wird dann jeder Anfrage als URI Parameter `key=apiKey` angehängt. Soll auch stellvertretend für einen Benutzer Kommentare geschrieben werden, ist noch zusätzlich der Username und das Password von ihm Voraussetzung.

Google bietet selber eine Java API für die Youtube Data API v2 an, mit der das Auslesen, Verarbeiten und Senden der Atom Feeds und Einträge sehr einfach ist. Das Programmbeispiel in Listing 5.12 zeigt eine Beispiel für die Verwendung der Youtube Data API in Java. Ausgangspunkt in Zeile 1 ist die Klasse `YouTubeService` die als Client eingesetzt wird. Bei der Erzeugung eines Objektes dieser Klasse wird eine beliebige Zeichenkette als `ClientID` (nicht zu verwechseln mit der `ClientID` von OAuth 2.0) und der oben beschriebene API-Schlüssel benötigt. In der zweiten Zeile werden der Benutzername und ein Password an den Client übergeben. Für das Auslesen eines Atom-Feeds wird die Methode `getFeed(...)` eingesetzt. Ihr werden in Zeile 5 die URI zum gesuchten Feed und in Zeile 6 die Klasse in die der Feed aus dem XML-Format konvertiert werden

¹⁶ <http://code.google.com/apis/youtube/dashboard/>

soll. In diesen Falls steckt hinter der URI eine Playliste mit mehreren Videoeinträgen. In Zeile 8 bis 11 werden dann die Titel und der Link auf das Video ausgegeben.

Listing 5.12: Youtube Data API v2: Java YouTubeService

```
1 YouTubeService service = new YouTubeService(clientId, apiKey);
2 service.setUserCredentials(username, password);
3
4 VideoFeed videoFeed = service.getFeed(
5     new URL("http://gdata.youtube.com/feeds/api/playlists/
6         PL59AFCB0F92BB89A9"),
7     VideoFeed.class);
8 for(VideoEntry videoEntry : videoFeed.getEntries() ) {
9     System.out.println(videoEntry.getTitle()
10         + ": "
11         + videoEntry.getSelfLink());
12 }
```

Wurde das Ergebnis in mehrere Teilfeeds aufgeteilt dann ist der Rückgabewert der Funktion `videoFeed.getNextLink()` ungleich `null`. Ist dies der Fall kann der nächste Teilfeed wieder mit der Methode `service.getFeed(...)` und der von `URI` des von `getNextLink()` zurückgegebenen Links geholt werden.

Listing 5.13: Youtube Data API v2: Kommentar schreiben

```
1 commentUrl = "http://gdata.youtube.com/feeds/api/videos/oHg5SJYRHA0/
2     comments";
3 CommentEntry newComment = new CommentEntry();
4 newComment.setContent(new PlainTextConstruct("Tolles Video!! :));
5 service.insert(new URL(commentUrl), newComment);
```

Das Schreiben eines Eintrags ist in Listing 5.13 beschrieben. Die URI zum Schreiben in Zeile 1 zeigt auf einen Kommentar-Feed für das Video mit der ID „oHg5SJYRHA0“. In der zweiten Zeile wird dann zunächst eine Objekt der Klasse `CommentEntry` erstellt und in Zeile 3 mit einer Kommentarnachricht befüllt. Dieses Objekt wird dann zusammen mit der URI aus Zeile 1 der Methode `insert(...)` übergeben. Für diese Methode ist das vorherige übergeben von Username und Passwort Pflicht, da sonst Youtube mit einen Fehler antwortet.

Youtube Mapping nach SIOC

Besonderheiten bei der Implementierung des YoutubeConnectors

Vorteil bei der Youtube Data API v2 ist die Basis von Atom als Datenformat. So sind schon viele URIs auf weiterführende Ressourcen vorhanden. Leider wird eine Anwendung sehr schnell von Youtube blockiert, wenn diese zu oft hintereinander Anfragen sendet. Aus diesem Grund muss zwischen zwei Abfragen immer eine bestimmte Zeit vergangen bevor die nächste abgeschickt

SIOC Klasse	URI-Format
Site	
Service	
UserAccount	
Forum	
Thread	
Post	

wird. Am besten hat sich eine Zeitspanne von 500 ms gezeigt, welche aber ab und zu immer noch Probleme macht. Diesen Wert aber höher zu stellen verlangsamt das komplette System, da manchmal einige Abfragen nötig sind bis alle gesuchten Informationen vorhanden sind. Zusätzlich sind nicht alle Informationen explizit über die Java-API zugänglich. Zum Beispiel erhält die Klasse für Kommentare keine Methode für den Zugriff auf die Video und Kommentar ID. Diese muss aus der URI des Kommentar erste extrahiert werden.

5.2.7 Canvas Connector

Canvas REST-API

Der öffentliche Zugriff auf die Daten von der Lernplattform Canvas basiert auf einer REST-API und OAuth 2.0 zur Autorisierung der Zugriffe. Die Adresse, über die auf die API zugegriffen werden kann, besteht aus der URI für die verwendete Canvasinstanz (`{rootUri}`) und gefolgt von dem Pfad „/api/v1“. Für die Demoinstanz von Instructure würde dies der URI `https://canvas.instructure.com/api/v1` entsprechen. An diese können dann weitere Pfade für den Zugriff auf die einzelnen Ressourcen angehängt werden. Zur Autorisierung wird von OAuth nur der Accesstoken benötigt, der bei der REST-Abfrage als HTTP Authorization Header mitgeschickt wird. Listing 5.14 zeigt eine Beispielanfrage und Angabe des Accesstokens mit dem Programm *curl*¹⁷. Der Platzhalter `{accessToken}` muss dann natürlich erst durch einen validen Accesstoken ersetzt werden. Diesen kann jeder Benutzer in seinem Canvas Profil unter „Settings“ und im Abschnitt „Approved Integrations“ selbst erstellen. Die Angabe von `ClientId` und `ClientSecret` von OAuth 2.0 sind nicht nötig.

Listing 5.14: Canvas Authorization Header

```
1 curl -H "Authorization: Bearer {accessToken}" https://canvas.instructure.com/api/v1/courses
```

Als Datenformat für die zurückgelieferten Daten wird von der Canvas API auf JSON gesetzt. Für die Verwendung von POST und PUT Operationen zum Schreiben nach Canvas können die

¹⁷ <http://curl.haxx.se/>

Daten entweder nach dem HTML Form Encoding¹⁸ Standard oder ebenfalls in JSON angegeben werden.

Da einige Anfragen eine Liste von Ergebnissen zurückliefern und diese möglicher Weise lang werden können, teilt die Canvas API diese Listen auf mehrere Seiten auf, die jede einzeln abgefragt werden müssen. Für jede Seite schickt die Canvas API mehrere URIs als HTTP Link Header¹⁹ der Antwort mit. Diese URIs erhalten zusätzlich noch ein Attribut *rel*, das beschreibt in welcher Relation die URI zu dieser Seite steht. Als Wert für diese Relation können „current“ für eine URI auf die aktuelle, „next“ auf die nächste, „prev“ auf die vorherige, „first“ auf die erste und „last“ für eine URI auf die letzte Seite vorkommen. Um also die nächste Seite vom Ergebnis zu bekommen, muss eine neue REST-Anfrage mit der Relation „next“ ausgeführt werden. Fehlt eine URI mit dieser Relation, ist das die letzte Seite erreicht.

CanvasLMS4J

Ein Problem mit der API von Canvas war es, dass es zwar eine gute REST Anbindung gab, aber noch keine Bibliothek um sie mit der Programmiersprache Java anzusprechen. Es musste also erst eine eigne Java API dazu entwickelt werden, die den Namen *CanvasLMS4J* (Kurzform für „Canvas LMS API für Java“) bekam.

Anhand der für die REST API verwendeten URIs ist auffällig, dass die einzelnen Bestandteile aufeinander aufbauen. Zum Beispiel ist der Ablauf für den REST-Zugriff auf DiscussionTopics in Gruppen und Kursen der gleiche, nur die verwendete URI unterscheidet sich. Aus diesem Grund wurden die einzelnen Ressource (Course, Groupe, DiscussionTopic, Entries, ...) als einzelne Endpunkte implementiert die sich von der Klasse *EndPoint* ableiten. Jeder Endpunkt kann einen Eltern-Endpunkt haben, wobei sich die endgültige URI für die REST-Abfrage aus dem Pfad des Eltern-Endpunktes und dem des aktuellen Endpunktes zusammensetzt. Zur Verdeutlichung sei hier die URI `https:{canvasUri}/api/v1/courses/1/discussion_topics` als Demonstration genannt. Sie besteht aus den statischen Teil `https:{canvasUri}/api/v1/` der den Ort für die verwendete Canvasinstanz angibt. Darauf folgt ein Kurs als erster Endpunkt mit der Kurs-ID „1“. Für diesen Kurs sollen nun alle Diskussionen abgefragt werden. Dies geschieht durch die Angabe des zweiten Endpunktes `/discussion_topics`. „`/courses/1`“ bildet hier also den Eltern-Endpunkt von `/discussion_topics`. Sollen aber nun alle Diskussionen in einer Gruppe abgefragt werden, reicht es aus den Kurs-Endpunkt durch einen Gruppen-Endpunkt auszutauschen.

Ausgangspunkt für CanvasLMS4J ist die Klasse *CanvasLmsClient* (siehe Abbildung 5.2). Über sie werden alle Endpunkte verwaltet die keinen Eltern-Endpunkt besitzen. Bei der Erzeugung eines Objektes dieser Klasse, werden ihr die URI zur verwendenden Canvasinstanz und der AccessToken des Benutzerkontos übergeben. Im aktuellen Stadium können von einem Client aus nur Endpunkte für Kurse oder Gruppe erstellt werden.

Endpunkte können dann über die Klasse *CanvasLmsRequest* REST-Anfragen an die Canvas API stellen. Hierzu verwendet CanvasLMS4J die *HttpClient*²⁰ Bibliothek von Apache mit die einzelnen HTTP-Operationen ausgeführt werden. Um die im JSON-Format zurück gelieferten

¹⁸ <http://www.w3.org/TR/html4/interact/forms.html#h-17.13.4>

¹⁹ <http://www.w3.org/Protocols/9707-link-header.html>

²⁰ <http://hc.apache.org/httpclient-3.x/>

Antworten der Canvas API in Java verwenden zu können, wir auf die Funktionen der von Google entwickelten *Gson*²¹ Bibliothek zurück gegriffen. Sie erlaubt es Java Objekte in das JSON-Format zu konvertieren und genauso aus Daten in JSON ein Java Objekt zu machen. Dadurch verringert sich der Aufwand für das Verarbeiten der Daten von Canvas auf das Erstellen der entsprechenden Java Klassen. CanvasLmsRequest können mit zwei verschiedenen Methoden ausgeführt werden. Die erste `execute()` wird dazu verwendet, wenn die REST-Abfrage nur die Rückgabe eines Objektes zu Folge hat. Also zum Beispiel, wenn nur die Daten eines bestimmten Kurses abgefragt werden soll. Die zweite Methode ist `executePagination`. Diese Methode dient für Abfragen die eine in Seiten aufgeteilte Liste von Ergebnissen zurückliefert und gibt für eine einfache Handhabbarkeit ein Objekt der Klasse `Pagination` zurück.

Die Klasse `Pagination` ist nach dem Iterator Muster aufgebaut und liefert vom kompletten Ergebnis bei jedem Iterationsschritt eine Seite zurück. Diese Seite enthält dann je eine Teilliste der angefragten Objekten. Ist eine Seite fertig ausgelesen, holt `Pagination` über eine vorgegeben REST Abfrage die nächste Seite. Dies geschieht solange bis keine Seiten mehr geladen werden können oder das Programm das Lesen abbricht.

Listing 5.15: CanvasLMS4J Beispielprogramm

```
1 CanvasLmsClient client = new CanvasLmsClient(  
2     "https://canvas.instructure.com",  
3     "7~LUpV7B3lJY...");  
4  
5 Pagination<DiscussionTopic> discussionPages = client.courses()  
6     .discussionTopics( 798152 )  
7     .list()  
8     .executePagination();  
9  
10  
11 for ( List<DiscussionTopic> discussions : discussionPages ) {  
12     for ( DiscussionTopic discussion : discussions ) {  
13         System.out.println( discussion );  
14     }  
15 }
```

In Listing 5.15 ist ein Beispiel für die Anwendung der CanvasLMS4J API zu sehen. In den ersten drei Zeilen wird eine neuer `CanvasLmsClient` für die Canvasinstanz auf `https://canvas.instructure.com` und einem `AccessToken` erstellt. Als nächstes soll für einen Kurs mit der ID „798152“ alle Diskussionen aufgelistet werden. Mit dem Aufruf der Methode `courses()` des Clients wird der Endpunkt für die Kurse und von diesem aus der Endpunkt für die Diskussionen im Kurs „798152“ über die Methode `.discussionTopics(798152)`. In der siebten Zeile wird die Art der Abfrage genauer festgelegt. Da eine Liste alle Diskussionen gesucht ist, wird die Methode `list()` aufgerufen die ein `CanvasLmsRequest` Objekt für die gewünschte Abfrage erstellt. Da die Antwort aus mehrere Objekten mit der Beschreibung der einzelnen Diskussionen besteht, wird die Anfrage in Zeile Acht mit der Methode `executePagination()` ausgeführt. Die einzelnen Seiten des Ergebnisses werden dann in der elften Zeile mit einer For-

²¹ <https://code.google.com/p/google-gson/>

Schleife durchlaufen. Das nachladen der Seiten erfolgt dabei automatisch durch die Pagination Klasse. Jede Seite besteht nun aus einer Liste mit den Beschreibung der Diskussionen in einem DiskussionTopic Objekt, welche dann wieder in einer weiteren Schleife ausgegeben werden.

Canvas Mapping nach SIOC

SIOC Klasse	URI-Format
Site	
Service	
UserAccount	
Forum	
Thread	
Post	

Tabelle 5.4.: Format der URIs für Canvas

URI Platzhalter	URI-Format
{serviceUri}	{rootUri}/api/v1
{userAccountUri}	{rootUri}/about/{userId}
{siteUri}	{rootUri}
{forumUri}	{rootUri}/courses/{courseId}, {rootUri}/groups/{groupId}
{threadUri}	{forumUri}/discussion_topics/{topicId}
{postUri}	{threadUri}/entries/{entryId}, {threadUri}#discussion_topic

5.3 Evaluation

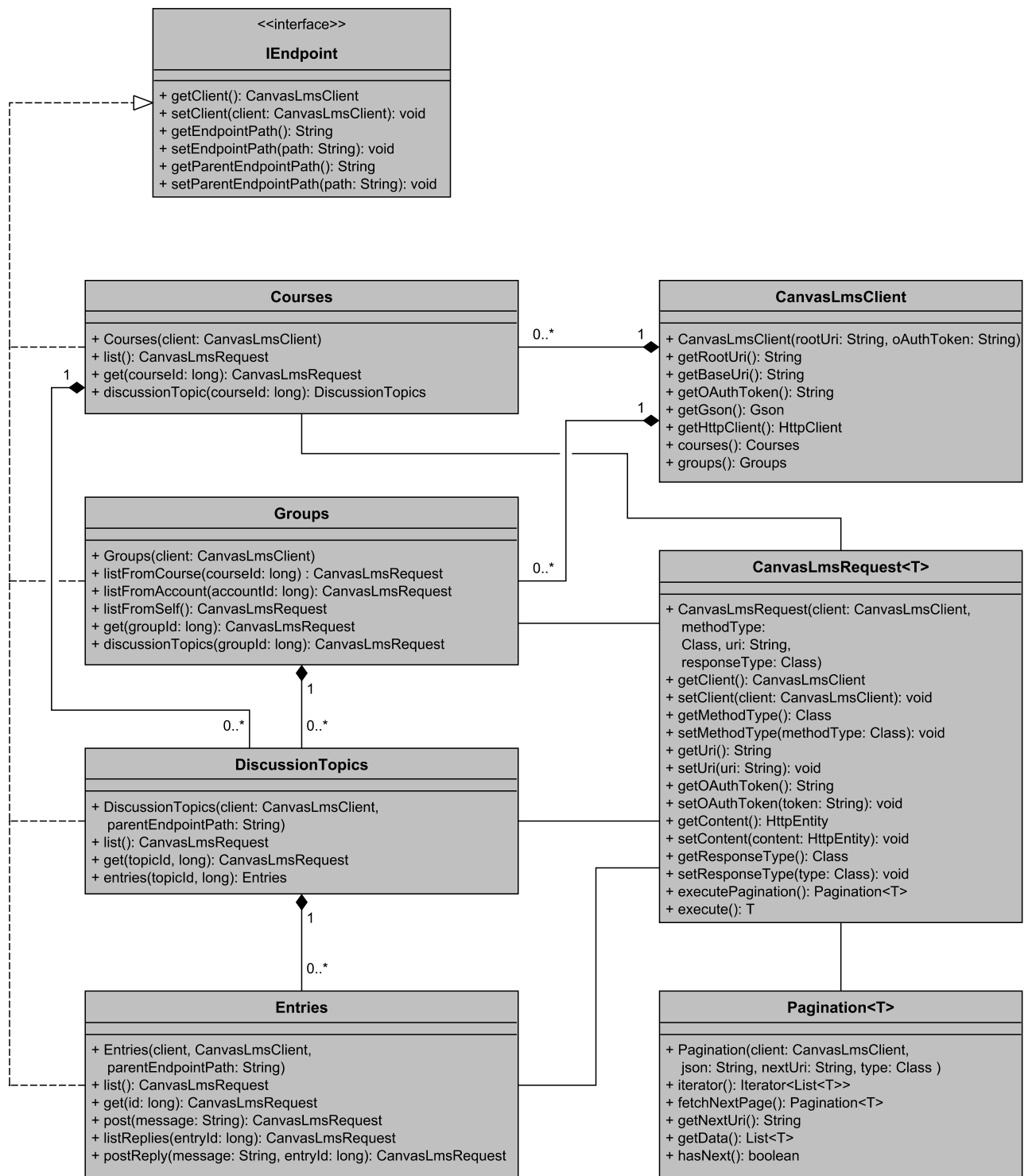


Abbildung 5.2.: UML Klassendiagramm von CanvasLMS4J



6 Zusammenfassung und Ausblick

Im letzten Kapitel sollen noch einmal alle Ergebnisse diese Arbeit zusammengefasst und einen Ausblick auf weiterführende Themen gegeben werden. In der Einleitung wurde beschrieben, dass Diskussionen ein wichtiger Bestandteil des E-Learnings ist, aber diese oftmals auf verschiedene Plattformen verteilt stattfinden. Dadurch verpassen Personen, die nur auf einer dieser Plattformen präsent sind, vielleicht für sie wichtige Informationen auf einer anderen. Außerdem werden die selben Diskussionsthemen immer wieder an unterschiedlichen Orten an den unterschiedlichsten Orten doppelt und dreifach behandelt. Aus diesem Grund sollte ein System entwickelt werden das den Austausch von Diskussionen zwischen den unterschiedlichen Plattformen ermöglicht.

In Kapitel 3 wurde analysiert, welche Schritte für eine solche Synchronisation nötig sind. Zuerst musste eine Zwischenformat gefunden werden in das die Daten der unterschiedlichen Plattformen konvertiert werden konnten, da dieser Weg effizienter ist, als die einzelnen Formate untereinander zu konvertieren. Also ein solches Zwischenformat bat sich die SIOC Ontologie in Verbindung mit FOAF wunderbar an. Ontologien bietet nicht nur einen guten Ansatz für die Integration von unterschiedlichen Datenformaten, mit RDF ist es sogar möglich dass Programme diese Daten verstehen und aus ihnen neues Wissen ableiten können. Um die Daten letztendlich in das Zwischenformat konvertieren und verarbeiten zu können, braucht es eine einheitliche Schnittstelle mit der dies umgesetzt werden kann. Zusätzlich wurde noch ein System gebraucht, dass die konvertierten Daten zwischen den Schnittstellen für die einzelnen Plattformen austauscht. Zu eignete sich der Austausch über Nachrichten am besten, da dadurch die einzelnen Schnittstellen zeitlich und räumlich entkoppelt werden konnten und nicht voneinander abhängig sind. Als Basis für dieses Nachrichtensystem wurde Camel ausgewählt. Mit dieser Java-Bibliothek können die Routen, welche die Nachrichten nehmen, flexibel konfiguriert werden und sind so leicht erweiterbar. Privatsphäre spielt in der heutigen Zeit ebenfalls eine wichtige Rolle. Deswegen muss es Benutzern ermöglicht werden das automatischen Lesen und/oder Schreiben für seine Beiträge zuzustimmen oder abzulehnen.

Das Kapitel 4 widmet sich der Beschreibung des Systems, welches die Anforderungen aus Kapitel 3 erfüllen soll. Dieses SOCC genannte System definiert dazu Conectoren, welche die Schnittstellen zu den einzelnen Plattformen abstrahiert und die unterschiedlichen Formate in das SIOC-Format konvertiert. Ein Connector bestehen aus drei Komponenten die je für eine bestimmte Teilaufgabe des Systems verantwortlich sind. Der StructureReader hilft dabei die Struktur der jeweiligen Plattform zu lesen und in SIOC abzubilden, so dass die Beiträge von den richtigen Stellen gelesen und an die richtigen Stellen geschrieben werden können. Der PostReader ist dabei für das Lesen von Beiträgen und deren Konvertierung vom Format der gelesenen Plattform in das SIOC-Format. Für dem umgekehrten Weg wird dann der PostWriter eingesetzt. Für den Betrieb der Connectoren brauchen diese einige Informationen von den Benutzern, für den Zugriff auf die einzelnen APIs. Für diese Informationen konnte auf der vorhandene Service Modul von SIOC für API Beschreibungen, die Basic Access Control Ontologie für die Autorisierung, sowie auf die aus vorhandenen Ontologien zusammengesetzte und erweiterte Services Authentication Ontologie zur Authentifizierung aufgebaut werden. Gespeichert wird alles in einen RDF-Triplestore und erlaubt

so einen einfachen Zugriff auf die untereinander verbunden Datensätze. Für die Integration der Connectoren in Camel wurde die Komponente SOCC-Camel entwickelt. Mit den EIP aus Camel können so die Connectoren auf verschiedenste Weisen verbunden werden.

Im letzten Kapitel

6.1 Ausblick

A Anhang

A.1 SOCC Connector Config Ontologie

```
1  <?xml version="1.0"?>
2  <!--
3      Author: Florian Mueller
4      Date: 2013-09-05
5      Version: 1.2
6  -->
7
8  <!DOCTYPE rdf:RDF [
9      <!ENTITY sioc "http://rdfs.org/sioc/ns#" >
10     <!ENTITY foaf "http://xmlns.com/foaf/0.1/" >
11     <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
12     <!ENTITY siocs "http://rdfs.org/sioc/services#" >
13     <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
14     <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
15     <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
16 ]>
17
18 <rdf:RDF xmlns="http://www.m0ep.de/socc/config#"
19     xml:base="http://www.m0ep.de/socc/config#"
20     xmlns:sioc="http://rdfs.org/sioc/ns#"
21     xmlns:dcterms="http://purl.org/dc/terms/"
22     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
23     xmlns:siocs="http://rdfs.org/sioc/services#"
24     xmlns:foaf="http://xmlns.com/foaf/0.1/"
25     xmlns:owl="http://www.w3.org/2002/07/owl#"
26     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
27     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
28
29     <owl:Ontology
30         rdf:about="http://www.m0ep.de/socc/config#"
31         rdf:type="http://www.w3.org/2002/07/owl#Thing">
32         <dcterms:title>SOCC Connector Configuration Ontology</
dcterms:title>
33         <dcterms:description>
34             Ontology for connector configurations of the SOCC (
Social Online Community Connectors) Framework.
35         </dcterms:description>
```

```

36
37     <owl:imports rdf:resource="http://rdfs.org/sioc/services#"/>
38     <owl:imports rdf:resource="http://rdfs.org/sioc/ns#"/>
39 </owl:Ontology>
40
41
42 <!--
43 ////////////////////////////////////////////////////
44 // Object Properties
45 ////////////////////////////////////////////////////
46 -->
47
48 <!-- http://www.m0ep.de/socc/config#defaultUser -->
49 <owl:ObjectProperty rdf:about="http://www.m0ep.de/socc/config#
defaultUserAccount">
50     <rdfs:range rdf:resource="&sioc;UserAccount"/>
51     <rdfs:domain rdf:resource="http://www.m0ep.de/socc/config#
ConnectorConfig"/>
52 </owl:ObjectProperty>
53
54 <!-- http://www.m0ep.de/socc/config#service -->
55 <owl:ObjectProperty rdf:about="http://www.m0ep.de/socc/config#
service">
56     <rdfs:domain rdf:resource="http://www.m0ep.de/socc/config#
ConnectorConfig"/>
57     <rdfs:range rdf:resource="&siocs;Service"/>
58 </owl:ObjectProperty>
59
60 <!--
61 ////////////////////////////////////////////////////
62 // Data properties
63 ////////////////////////////////////////////////////
64 -->
65
66
67 <!-- http://www.m0ep.de/socc/config#connectorClass Name
68      Java class of the connector
69 -->
70 <owl:DatatypeProperty rdf:about="http://www.m0ep.de/socc/config#
connectorClassName">
71     <rdfs:domain rdf:resource="http://www.m0ep.de/socc/config#
ConnectorConfig"/>
72     <rdfs:range rdf:resource="&xsd:string"/>
73 </owl:DatatypeProperty>
74
75 <!-- http://www.m0ep.de/socc/config#id -->

```

```

76     <owl:DatatypeProperty rdf:about="http://www.m0ep.de/socc/config#
id">
77         <rdfs:domain rdf:resource="http://www.m0ep.de/socc/config#
ConnectorConfig"/>
78         <rdfs:range rdf:resource="&xsd:string"/>
79     </owl:DatatypeProperty>
80
81 <!-- http://www.m0ep.de/socc/config#unknownMessageTemplate -->
82     <owl:DatatypeProperty rdf:about="http://www.m0ep.de/socc/config#
unknownMessageTemplate">
83         <rdfs:domain rdf:resource="http://www.m0ep.de/socc/config#
ConnectorConfig"/>
84         <rdfs:range rdf:resource="&xsd:string"/>
85     </owl:DatatypeProperty>
86
87     <!--
88     //////////////////////////////////////
89     // Classes
90     //////////////////////////////////////
91     -->
92
93     <!-- http://www.m0ep.de/socc/config#ConnectorConfig -->
94     <owl:Class rdf:about="http://www.m0ep.de/socc/config#
ConnectorConfig"/>
95 </rdf:RDF>
96
97 <!-- Generated by the OWL API (version 3.4.2) http://owlapi.
sourceforge.net -->

```

assets/listings/socc-config.owl

A.2 SIOC Services Authentication Module

```

1  <?xml version="1.0"?>
2  <!--
3      Author:      Florian Mueller
4      Date:        2013-08-07
5      Version:     2.0
6  -->
7
8  <!DOCTYPE RDF [
9      <!ENTITY sioc "http://rdfs.org/sioc/ns#" >
10     <!ENTITY dcterms "http://purl.org/dc/terms/" >
11     <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
12     <!ENTITY sioc_services "http://rdfs.org/sioc/services#" >

```

```

13      <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
14      <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
15      <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
16      <!ENTITY waa "http://purl.oclc.org/NET/WebApiAuthentication#" >
17  ]>
18  <rdf:RDF
19      xml:base="http://www.m0ep.de/sioc/services/auth#"
20      xmlns="http://www.m0ep.de/sioc/services/auth#"
21      xmlns:dcterms="http://purl.org/dc/terms/"
22      xmlns:owl="http://www.w3.org/2002/07/owl#"
23      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
24      xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
25      xmlns:sioc="http://rdfs.org/sioc/ns#"
26      xmlns:siocs="http://rdfs.org/sioc/services#"
27      xmlns:waa="http://purl.oclc.org/NET/WebApiAuthentication#"
28      xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
29
30      <owl:Ontology
31          rdf:about="http://www.m0ep.de/sioc/services/auth#"
32          rdf:type="http://www.w3.org/2002/07/owl#Thing">
33          <dcterms:title>SIOC Service Authentication Module</
dcterms:title>
34          <dcterms:description>
35              Extends the SIOC Core and Service Module to add
information about authentication mechanisms and their required
credentials. It reuses some parts from the WebApiAuthentication
Ontology.
36          </dcterms:description>
37          <rdfs:seeAlso rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#"/>
38          <rdfs:seeAlso rdf:resource="http://rdfs.org/sioc/services#"/
>
39
40          <owl:imports rdf:resource="http://rdfs.org/sioc/services#"/>
41      </owl:Ontology>
42
43      <!--
44      //////////////////////////////////////
45      // Object Properties
46      //////////////////////////////////////
47      -->
48
49      <!-- http://purl.oclc.org/NET/WebApiAuthentication#
hasInputCredentials -->
50      <owl:ObjectProperty

```

```

51         rdf:about="http://purl.oclc.org/NET/WebApiAuthentication#
hasInputCredentials">
52         <owl:equivalentProperty
53             rdf:resource="http://www.m0ep.de/sioc/services/auth#
credentials"/>
54         </owl:ObjectProperty>
55
56         <!-- http://www.m0ep.de/sioc/services/auth#serviceAuthentication
-->
57         <owl:ObjectProperty rdf:about="http://www.m0ep.de/sioc/services/
auth#serviceAuthentication">
58             <rdfs:range
59                 rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
60             <rdfs:domain rdf:resource="http://rdfs.org/sioc/services#
Service"/>
61             </owl:ObjectProperty>
62
63             <!-- http://www.m0ep.de/sioc/services/auth#accountAuthentication
-->
64             <owl:ObjectProperty rdf:about="http://www.m0ep.de/sioc/services/
auth#accountAuthentication">
65                 <rdfs:range
66                     rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
67                 <rdfs:domain rdf:resource="http://rdfs.org/sioc/ns#
UserAccount"/>
68                 </owl:ObjectProperty>
69
70             <!-- http://www.m0ep.de/sioc/services/auth#credentials -->
71             <owl:ObjectProperty rdf:about="http://www.m0ep.de/sioc/services/
auth#credentials">
72                 <rdfs:domain
73                     rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
74                 <rdfs:range rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
75                 </owl:ObjectProperty>
76
77             <!--
78             //////////////////////////////////////
79             // Classes
80             //////////////////////////////////////
81             -->
82
83             <!-- http://purl.oclc.org/NET/WebApiAuthentication#APIKey -->

```

```

84     <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#APIKey">
85         <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
86     </owl:Class>
87
88     <!-- http://purl.oclc.org/NET/WebApiAuthentication#
AuthenticationMechanism -->
89     <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
90
91     <!-- http://purl.oclc.org/NET/WebApiAuthentication#Credentials
-->
92     <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
93
94     <!-- http://purl.oclc.org/NET/WebApiAuthentication#Direct -->
95     <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#Direct">
96         <rdfs:subClassOf
97             rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
98     </owl:Class>
99
100    <!-- http://purl.oclc.org/NET/WebApiAuthentication#OAuth -->
101    <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#OAuth">
102        <rdfs:subClassOf
103            rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
104    </owl:Class>
105
106    <!-- http://purl.oclc.org/NET/WebApiAuthentication#Password -->
107    <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#Password">
108        <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
109    </owl:Class>
110
111    <!-- http://purl.oclc.org/NET/WebApiAuthentication#Username -->
112    <owl:Class rdf:about="http://purl.oclc.org/NET/
WebApiAuthentication#Username">
113        <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
114    </owl:Class>
115

```

```

116     <!-- http://www.m0ep.de/sioc/services/auth#AccessToken -->
117     <owl:Class rdf:about="http://www.m0ep.de/sioc/services/auth#
AccessTokens">
118         <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
119         <owl:equivalentClass
120             rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#OAuthToken"/>
121     </owl:Class>
122
123     <!-- http://www.m0ep.de/sioc/services/auth#ClientId -->
124     <owl:Class rdf:about="http://www.m0ep.de/sioc/services/auth#
ClientId">
125         <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
126         <owl:equivalentClass
127             rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#OAuthConsumerKey"/>
128     </owl:Class>
129
130     <!-- http://www.m0ep.de/sioc/services/auth#ClientSecret -->
131     <owl:Class rdf:about="http://www.m0ep.de/sioc/services/auth#
ClientSecret">
132         <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
133         <owl:equivalentClass
134             rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#OAuthConsumerSecret"/>
135     </owl:Class>
136
137     <!-- http://www.m0ep.de/sioc/services/auth#RefreshToken -->
138     <owl:Class rdf:about="http://www.m0ep.de/sioc/services/auth#
RefreshToken">
139         <rdfs:subClassOf rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#Credentials"/>
140     </owl:Class>
141
142     <!-- http://www.m0ep.de/sioc/services/auth#WebAPI -->
143     <owl:Class rdf:about="http://www.m0ep.de/sioc/services/auth#
WebAPI">
144         <rdfs:subClassOf
145             rdf:resource="http://purl.oclc.org/NET/
WebApiAuthentication#AuthenticationMechanism"/>
146     </owl:Class>
147 </rdf:RDF>

```

148 | `<!-- Generated by the OWL API (version 3.4.2) http://owlapi.
sourceforge.net -->`

`assets/listings/sioc-service-auth.owl`

Literaturverzeichnis

- [1] WebAccessControl. <http://www.w3.org/wiki/WebAccessControl>, Zugriff: 2013-09-12.
- [2] Apache Software Foundation. Apache Camel. <http://camel.apache.org>, Zugriff: 2013-09-15.
- [3] Tim Berners-Lee. Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>, Zugriff: 2013-09-17, 2009.
- [4] Tim Berners-Lee. Socially Aware Cloud Storage. <http://www.w3.org/DesignIssues/CloudStorage.html>, Zugriff: 2013-08-26, 2011.
- [5] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.
- [6] Uldis Bojars, John G. Breslin, and Stefan Decker. Porting social media contributions with SIOC. *Recent Trends and Developments in Social Software*, 6045:116–122, 2011.
- [7] Uldis Bojars, Alexandre Passant, John G. Breslin, and Stefan Decker. Social Network and Data Portability using Semantic Web Technologies. (1):5–19, 2008.
- [8] John G. Breslin, Andreas Harth, Uldis Bojars, and Stefan Decker. Towards semantically-interlinked online communities. *The Semantic Web: Research and Applications*, pages 71–83, 2005.
- [9] Dan Brickley and R.V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. <http://www.w3.org/TR/rdf-schema/>, Zugriff: 2013-09-14, 2004.
- [10] Dan Brickley and Libby Miller. FOAF Vocabulary Specification 0.98. <http://xmlns.com/foaf/spec/>, Zugriff: 2013-09-13, 2010.
- [11] Erik Christensen, Francisco Curbera, Greg Meredith, and Sanjiva Weerawarana. Web Services Description Language (WSDL). <http://www.w3.org/TR/wsdl>, Zugriff: 2013-09-09, 2001.
- [12] Jo Davies and Martin Graff. Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, 36(4):657–663, 2005.
- [13] Digital Enterprise Research Institute. SIOC - Semantically-Interlinked Online Communities. <http://sioc-project.org/>, Zugriff: 2013-08-15.
- [14] Stephen Downes. E-learning 2.0. *eLearn Magazine*, 2005.
- [15] Facebook. Graph API. <https://developers.facebook.com/docs/reference/api>: Zugriff: 2013-09-20.
- [16] Facebook. Key Facts. <http://newsroom.fb.com/Key-Facts>, Zugriff: 2013-09-11, 2013.

-
- [17] Facebook in Education and Anthony Fontana. Using a Facebook Group As a Learning Management System. <https://www.facebook.com/notes/facebook-in-education/using-a-facebook-group-as-a-learning-management-system/10150244221815570>, Zugriff: 2013-09-14, 2010.
- [18] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. PhD thesis, University of California, 2000.
- [19] Anthony Fontana. The Multichronic Classroom : Creating an Engaging Environment for All Students. *FOUNDATIONS IN ART: THEORY AND EDUCATION, FATE IN REVIEW*, 30:13—18, 2009.
- [20] D. Hardt. The OAuth 2.0 Authorization Framework. <http://tools.ietf.org/html/rfc6749>, Zugriff: 2013-08-30, 2012.
- [21] Hans-Werner Heinzen. Primer: Getting into RDF & Semantic Web using N3 Deutsche Übersetzung. <http://www.bitloeffel.de/DOC/2003/N3-Primer-20030415-de.html>, Zugriff: 2013-08-19.
- [22] Pascal Hitzler, Arkus Krötzsch, Sebastian Rudolph, and York Sure. *Semantic Web: Grundlagen*. Springer, 2008 edition, 2008.
- [23] Gregor Hohpe and Bobby Woolf. *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2003.
- [24] James Hollenbach, Joe Presbrey, and Tim Berners-Lee. Using RDF Metadata To Enable Access Control on the Social Semantic Web. *Proceedings of the Workshop on Collaborative Construction, Management and Linking of Structured Knowledge (CK2009)*, 514, 2009.
- [25] Iwen Huang. The effects of personality factors on participation in online learning. In *Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication, ICUIMC '09*, pages 150–156, New York, NY, USA, 2009. ACM.
- [26] Graham Klyne and Jerme J. Carrol. Resource Description Framework (RDF): Concepts and Abstract Syntax. <http://www.w3.org/TR/rdf-concepts/>, Zugriff: 2013-08-19, 2004.
- [27] Frank Manola and Eric Miller. RDF Primer. <http://www.w3.org/TR/rdf-primer/>, Zugriff: 2013-08-19, 2004.
- [28] Frank McCown and Michael L. Nelson. What happens when facebook is gone. *Proceedings of the 2009 joint international conference on Digital libraries JCDL 09 (2009)*, pages 251–254, 2009.
- [29] Nilo Mitra and Yves Lafon. SOAP Version 1.2. <http://www.w3.org/TR/2007/REC-soap12-part0-20070427/>, Zugriff: 2013-09-09, 2007.
- [30] Stuart Palmer, Dale Holt, and Sharyn Bray. Does the discussion help? The impact of a formally assessed online discussion on final student results. *British Journal of Educational Technology*, 39(5):847–858, 2008.
- [31] Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. <http://www.w3.org/TR/owl-semantics/>, Zugriff: 2013-

09-14, 2004.

- [32] Andrea Schröder. Web der Zukunft. *XML & Web Services Magazin*, pages 40—43, 2003.
- [33] Felix Schwenzel and Sascha Lobo. Reclaim Social. <http://reclaim.fm/>, Zugriff: 2013-08-14.
- [34] Sun/Oracle. Java Message Service. <http://www.oracle.com/technetwork/java/jms/index.html>, Zugriff: 2013-09-15.
- [35] Watkins Thomas. Suddenly, Google Plus Is Outpacing Twitter To Become The World's Second Largest Social Network. <http://www.businessinsider.com/google-plus-is-outpacing-twitter-2013-5>, Zugriff: 2013-09-11, 2013.
- [36] Mike Uschold and Michael Gruninger. Ontologies: Principles, methods and applications. *Knowledge engineering review*, (February), 1996.
- [37] Qiyun Wang, Huay Lit Woo, Choon Lang Quek, Yuqin Yang, and Mei Liu. Using the Facebook group as a learning management system: An exploratory study. *British Journal of Educational Technology*, 43(3):428–438, May 2012.
- [38] Youtube. Statistik. <http://www.youtube.com/yt/press/de/statistics.html>, Zugriff: 2013-09-10.