

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...



Inhaltsverzeichnis

1. Einleitung	3
2. Grundlagen	5
2.1. Zugriff auf Webanwendungen	5
2.1.1. Representational State Transfer (REST)	5
2.1.2. Simple Object Access Protocol (SOAP)	6
2.2. Datenintegration	7
2.2.1. Semantic web und das Resource Description Framework	7
2.2.2. Friend of a Friend (FOAF)	10
2.2.3. Semantically-Interlinked Online Communities (SIOC)	10
2.3. Datenverteilung	11
2.3.1. Java Messaging Service	11
2.3.2. Enterprise Integration Pattern und Apache Camel	11
2.4. Lernplattformen und soziale Online-Netzwerke	11
2.4.1. Moodle	11
2.4.2. Canvas	12
2.4.3. Youtube	13
2.4.4. Facebook	13
2.4.5. Google+	14
2.5. Verwandte Arbeiten und Projekte	14
2.5.1. What happens when Facebook is gone?	14
2.5.2. Reclaim Social	15
3. Analyse	17
3.1. Neuen Beitrag verfassen	17
3.2. Beiträge von sozialen Netzwerk A lesen	17
3.3. Beitrag in soziales Netzwerk B schreiben	19
3.4. Identifizierung der Komponenten	20
4. Eigener Ansatz	21
4.1. Social Online Community Connectors	21
4.1.1. Datenformat	21
4.1.2. Konfiguration	21
4.1.3. Aufbau der SOCC	27
4.1.4. Design eines Connectors	27
4.2. SOCC-Camel	31
4.2.1. SoccComponent	31
4.2.2. SoccPostPollingConsumer	31
4.2.3. SoccPostProducer	31

5. Implementierung	35
5.1. Implementierung der Connectoren	35
5.1.1. Moodle	35
5.1.2. Facebook	36
5.1.3. Google+	36
5.1.4. Youtube	37
5.1.5. Canvas	37
6. Evaluation	39
7. Abschlussbetrachtung	41
7.1. Fazit	41
7.2. Ausblick	41
A. Anhang	45
A.1. SOCC Connector Config Ontologie	45
A.2. SIOC Services Authentication Module	47

Abbildungsverzeichnis

2.1. Einfacher RDF-Graph	8
2.2. Aufbau von SIOC (modifiziert) - Originalquelle: [9]	11
2.3. Moodle Instanz der TU Darmstadt	12
2.4. Instructure Canvas	13
3.1. Benutzer erstellt einen Beitrag im sozialen Netzwerk A.	17
3.2. Lesen des erstellten Beitrags und konvertieren in das Zwischenformat.	18
3.3. Konvertierten des Beitrags in das Format B und schreiben in das soziale Netzwerk B .	19
4.1. Connector Config Ontology	22
4.2. SIOC Services Module	23
4.3. Zusammenhang von Person, UserAccount und Service. Die inversen Eigenschaften sioc:account_of und siocs:service_of wurden zu einer besseren Übersicht weggelassen	24
4.4. SIOC Services Authentication Ontology	25
4.5. Basic Access Control Ontologie	26
4.8. SOCC Context	27
4.6. Übersicht der Komponenten der SOCC	28
4.9. AccessControl	28
4.10. ClientManager	28
4.11. StrukturReader	29
4.12. PostReader	30
4.7. UML Klassendiagramm eines Connectors	32
4.13. UML Sequenzdiagramm eines PostWriters	33
4.14. Übersicht des Apache Camel Moduls Socc-Camel	33



Tabellenverzeichnis

2.1. Wichtigsten HTTP Operationen mit REST	6
3.1. Anzahl Konverter bei drei sozialen Netzwerken	18
4.1. Variablennamen und Ersetzung innerhalb von <code>unknownMessageTemplate</code>	22



Liste der noch zu erledigenden Punkte

Kapitelbeschreibung	4
mehr QUELLEN!!!	5
Idee semantic web einbauen	7
Das klingt ein wenig wie ein Werbetext ;-)	15
Inhalt: wass gibt es, was muss neu gemacht werden	17
Begriff soziales Netzwerk anpassen	17
Auch begründen warum die Verwendung von Ontologien und nicht bspw. die Entwicklung eines XML Schemas (alternativ bei der Analyse Kap. 3.2)	21
vll. noch OAuth 1.0(a) einbauen	25



1 Einleitung

Durch die Omnipräsenz des Internets im heutigen Alltag haben sich viele Bereichen 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 [10]. Es wurden Studien durchgeführt, die zeigen dass Studenten welche sich an online Diskussionen teilnehmen dazu tendieren gute Noten zu bekommen, als solche die nicht teilnahmen [8, 23]. 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 [18].

Qiyun Wang et. al. [26] 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“ [26, 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 abgeschottete „Datensilos“ [3] handelt auf die nur über zum Großteil heterogene Schnittstellen

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

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!!!

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 der Zugriff über eine vordefinierte Schnittstelle gestattet. Für Anwendungen und Dienste im Web sind die folgenden zwei Ansätze für die Architektur einer solchen Schnittstellen besonders verbreitet.

2.1.1 Representational State Transfer (REST)

Eine sehr beliebte Architektur für den öffentlichen Zugriff auf Webanwendungen ist *Representational State Transfer* (REST) [13, 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.“ [13, 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 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.

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.

2.1.2 Simple Object Access Protocol (SOAP)

Das *Simple Object Access Protocol*[22] (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.

Web Services Description Language

Gebräuchlich ist der Einsatz von SOAP bei sogenannten *Remote Procedure Calls* (RPC). Unter RPC versteht man den Aufruf eine Funktion von einem entfernten Dienst und das Zurückliefern einer eventuell vorhandenen Antwort. Welche Funktionen von einem Dienst zur Verfügung stehen wird

ein einer *Web Services Description Language*[7] (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 einem einzigen Dienst.

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

2.2 Datenintegration

2.2.1 Semantic web und das Resource Description Framework

Idee semantic web einbauen

Eine der bekanntesten Umsetzungen der Vision des semantischen Webs ist wohl das *Resource Description Framework* (RDF). Wie der Name schon suggeriert dient RDF zur Beschreibung von einzelnen Ressourcen innerhalb des Internets. Nach [19, 20] bestand die Motivation bei der Entwicklung von RDF Information über Ressource in einem 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“[19].

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 [16]. Für Subjekt, Prädikat und Objekt werden in RDF *Uniform Resource Identifier* (URI), *Literale* oder *leere Knoten* (im englischen *Blank Nodes* genannt) verwendet.

¹ <http://www.w3.org/XML/Schema>

URIs sind eindeutige Bezeichner die eine beliebige reale oder abstrakte Ressource darstellen und werden wie in RFC 2396² beschrieben formatiert. Relative URIs sollten aber nach [19] 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"sd: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 Sprache als auch der Datentyp müssen übereinstimmen.

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.

Darstellung von RDF

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.

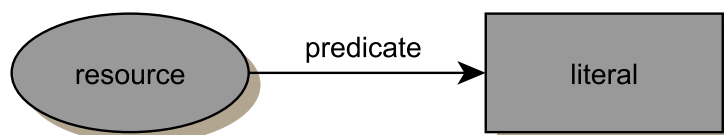


Abbildung 2.1.: Einfacher RDF-Graph

RDF/XML

RDF/XML[20, Abschnitt 3.2] 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 zusehen. 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

² <http://www.isi.edu/in-notes/rfc2396.txt>

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:ressource` 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

Turtle (Ausgeschrieben: *Terse RDF Triple Language*) ist eine weitere Möglichkeit RDF-Graphen darzustellen und ist eine ging aus der Sprache N3 (Kurzform für Notation 3) hervor [2]. In Turtle wird das Triple aus Subjekt, Prädikat und Objekt hintereinander geschrieben und zwischen jeden mindestens ein Leerzeichen gelassen. Als Abschluss folgt nach jedem Tripel 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 Spitzenkammern 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#`. Dieser Präfix kann nun überall innerhalb des Dokumentes verwendet werden, wobei die Spitzenkammern 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 die Frage für die Dose angegeben, das Triple aber mit einem Semikolon abgeschlossen und nicht mit einem Punkt. Durch das Semikolon ist es möglich, das 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. Ein Beispiel für einen solchen leeren Knoten in Turtle ist in Listing 4.1 zu sehen.

Ontologien

2.2.2 Friend of a Friend (FOAF)

*Friend of a Friend*³ (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 [6] 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 ist 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-Bereich enthält alle Teile, die für das Web interessant wären. Das wären zum Beispiel, welche E-Mail-Adresse eine Person besitzt, welche Benutzerkonten, 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 besseren Austausch zwischen den einzelnen Seiten [6, Abstract].

2.2.3 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 [9, 5]. 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 designed. 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 Terms⁴ eingesetzt.

³ <http://www.foaf-project.org>

⁴ <http://dublincore.org/documents/dcmi-terms>

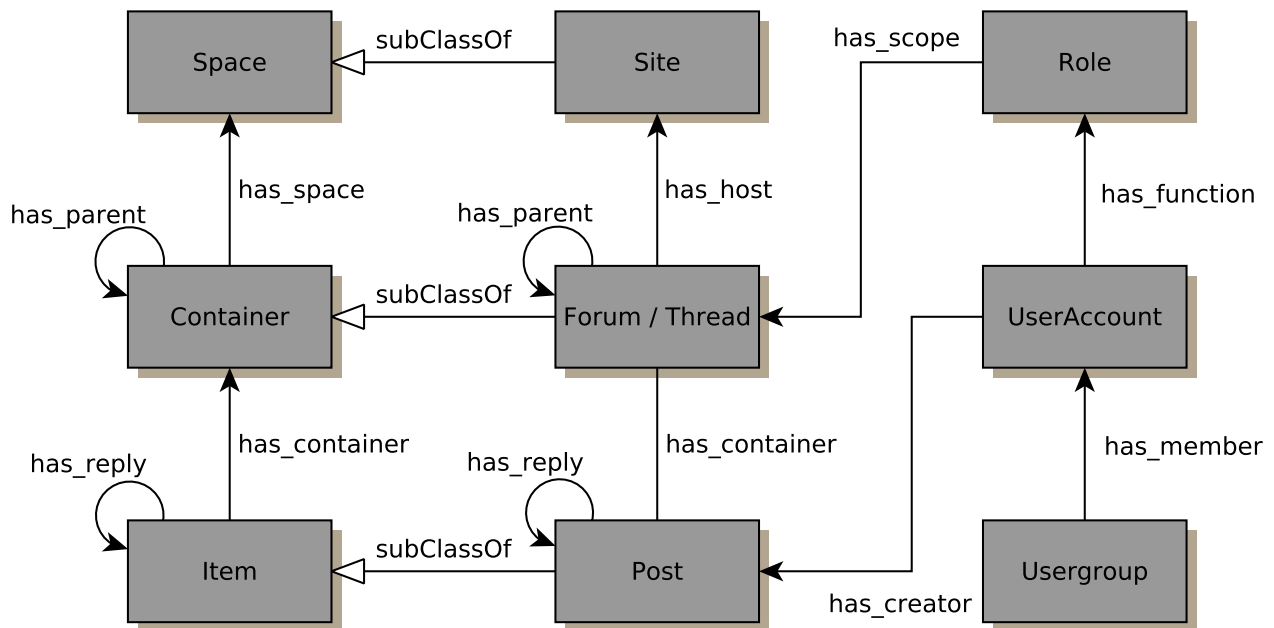


Abbildung 2.2.: Aufbau von SIOC (modifiziert) - Originalquelle: [9]

2.3 Datenverteilung

2.3.1 Java Messaging Service

2.3.2 Enterprise Integration Pattern und Apache Camel

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

Moodle⁵ 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 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

⁵ <https://moodle.org/>

ebenfalls vorhanden. Teilnehmer können Lerngruppen bilden, sich über persönliche Nachrichten austauschen, gemeinsam an Wikis arbeiten oder in Foren diskutieren.

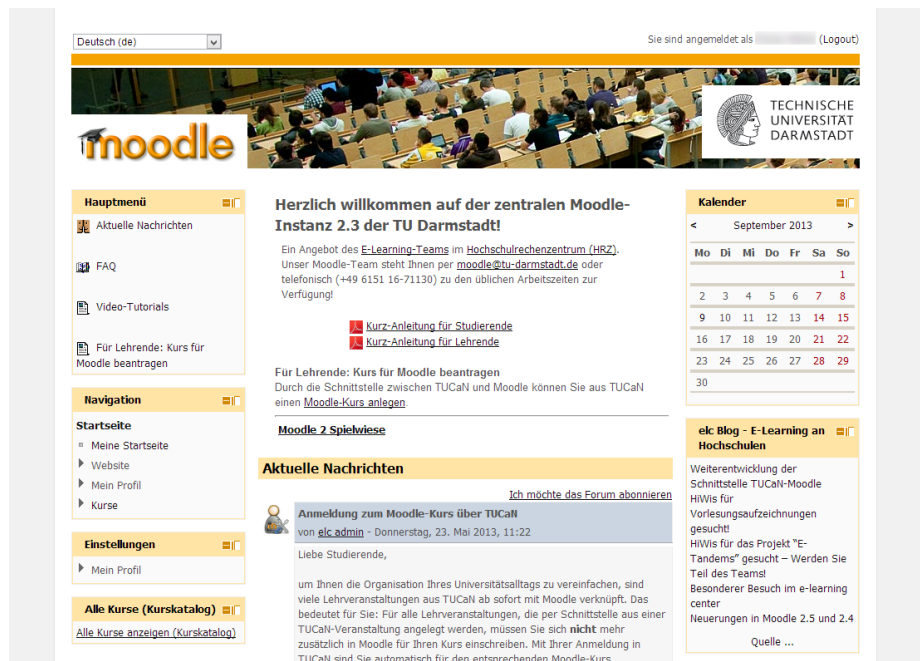


Abbildung 2.3.: Moodle Instanz der TU Darmstadt

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 Instructure⁶ 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 Docs⁷ oder der Webkonferenz Anwendung BigBlueButton⁸.

Canvas wird mittels des Webframeworks *Ruby on Rails*⁹ 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

⁶ <https://www.instructure.com/>

⁷ <https://drive.google.com>

⁸ <http://www.bigbluebutton.org>

⁹ <http://rubyonrails.org/>

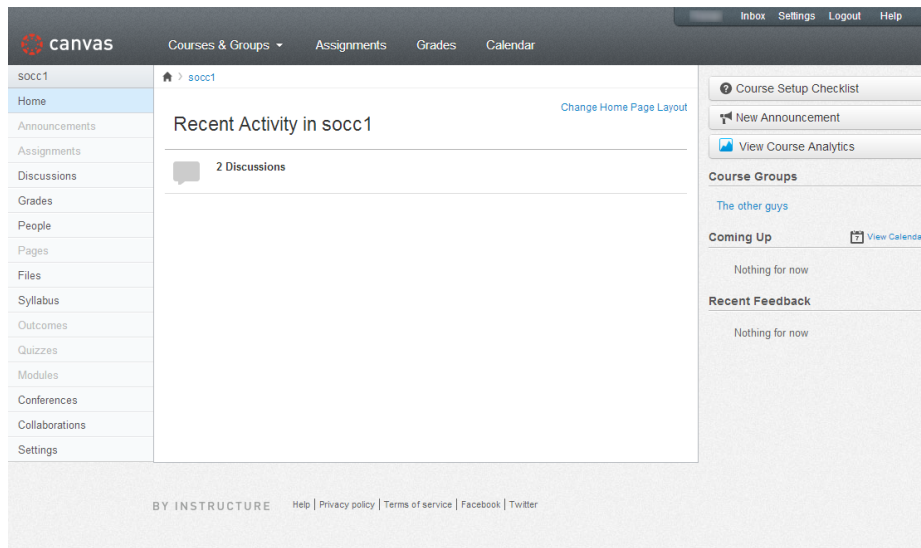


Abbildung 2.4.: Instructure Canvas

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 Gists¹⁰ und vielen weiteren.

2.4.3 Youtube

Die Youtube¹¹ 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 [27]. 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.

2.4.4 Facebook

Das soziale Online-Netzwerk Facebook¹² kann mit rund 699 Millionen aktiven Benutzern täglich [11] 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

¹⁰ <https://gist.github.com>

¹¹ <https://www.youtube.com>

¹² <https://www.facebook.com/>

Beiträge veröffentlichen oder die anderer kommentieren. Wie in der Einleitung schon erklärt zeigt Qiyun Wang et. al. [26] das sich Facebook, wenn auch mit Einschränkungen, wunderbar zur Verwaltung und Nutzung durch Lernkurse und Lerngruppen eignet. Die gleiche Erfahrung teilte Anthony Fontana [14, 12], der Facebook als Alternative zum bestehenden System der Bowling Green State University in Ohio, USA verwendete.

2.4.5 Google+

Google+¹³ 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 [25]. 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 Skype¹⁴ 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

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?“[21], 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 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“ [21, 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

¹³ <https://plus.google.com>

¹⁴ <http://www.skype.com/>

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 [21, 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 Feed¹⁵ 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:publica¹⁶ 2013 ihr gestartetes Projekt „Reclaim Social“ [24] vorgestellt.

Das klingt ein wenig wie ein Werbetext ;-)

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 „WordPress¹⁷“ und der dafür vorhandenen Erweiterung „FeedWordPress“¹⁸. 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“ [24, 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.

¹⁵ <http://www.rssboard.org/rss-specification>

¹⁶ <http://re-publica.de/>

¹⁷ <http://wordpress.org/>

¹⁸ <http://feedwordpress.radgeek.com/>

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“[24, How Does It Work]

3 Analyse

Inhalt: wass gibt es, was muss neu gemacht werden

Begriff soziales Netzwerk anpassen

Um sich eine Vorstellung davon zu machen, wie ein System auszusehen hat und welche Komponenten dazu nötig sind um zwei oder mehrere sozialen Netzwerke zu verbinden, soll hierzu ein kleines Ablaufbeispiel konstruiert werden.

3.1 Neuen Beitrag verfassen

Alles beginnt damit, dass zum Beispiel ein Student im sozialen Netzwerk A, im Forum zur Veranstaltung Telekooperation 1, eine Frage zur aktuellen Übung stellen will. Er geht zuerst in den passenden Thread und beginnt einen neuen Beitrag zu schreiben. Sobald er fertig ist, klickt er auf „Absenden“ und sein Beitrag wird in der Datenbank des sozialen Netzwerkes A gespeichert und als neuer Eintrag im Thread angezeigt (siehe Abbildung 3.1).

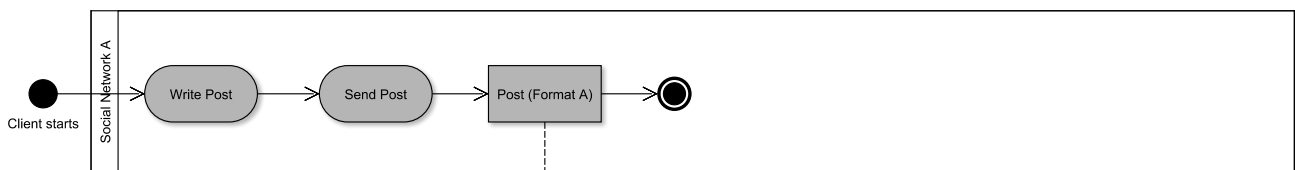


Abbildung 3.1.: Benutzer erstellt einen Beitrag im sozialen Netzwerk A.

3.2 Beiträge von sozialen Netzwerk A lesen

Um diese Beitrag in das soziale Netzwerk B transferieren zu können, müssen zuerst die Daten über eine öffentliche Schnittstelle vom Server des sozialen Netzwerkes A heruntergeladen werden. Da in der Regel 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 das soziale Netzwerk B geschickt werden, da sich diese in der Regel im verwendeten Datenformat unterscheiden. Diese müssen zuvor konvertiert werden.

Die einfachste Möglichkeit wäre nun die Daten von Format A nach Format 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 nun ein weiteres Netzwerk C unterstützt werden soll, würde ich die Anzahl an nötigen Konvertern auf Sechs erhöhen, wie Tabelle 3.1 zeigt.

Nimmt man an n_{sn} sei eine beliebige Anzahl sozialer Netzwerke, entspricht die Anzahl der notwendiger Konverter $n_{k1} = n_{sn} * (n_{sn} - 1)$, da für jedes Netzwerk ein Konverter in alle anderen

Tabelle 3.1.: Anzahl Konverter bei drei sozialen Netzwerken

		Nach		
		Netzwerk A	Netzwerk B	Netzwerk C
Von	Netzwerk A	-	×	×
	Netzwerk B	×	-	×
	Netzwerk C	×	×	-

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.

Eine elegantere Methode, welche die Anzahl zu implementierender Konverter in Grenzen halten kann, wäre die Einführung eines Zwischenformates. Geht man davon aus, dass die Daten aller Netzwerke nur in dieses Zwischenformat geschrieben und aus diesem gelesen werden müssen, würde sich der Aufwand auf maximal Zwei Konverter pro neuem Netzwerk reduzieren. Für eine beliebige Anzahl Netzwerke wären also $n_{k2} = n_{sn} * 2$ Konverter nötig. Nachteile hätte dieser Ansatz nur für $n_{sn} = 2$ und $n_{sn} = 3$, da in diesen Fällen mehr beziehungsweise gleich viele Konverter gegenüber der ersten Methode erforderlich wären. Erhöht man die Anzahl Netzwerke jedoch nur geringfügig, sinkt die Menge an Konvertern sichtbar. Für $n_{sn} = 4$ wären es $n_{k2} = 8$ statt $n_{k1} = 12$ und für $n_{sn} = 5$ 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. Abbildung 3.2 verdeutlicht den Ablauf unter Verwendung des eben beschriebenen Zwischenformats.

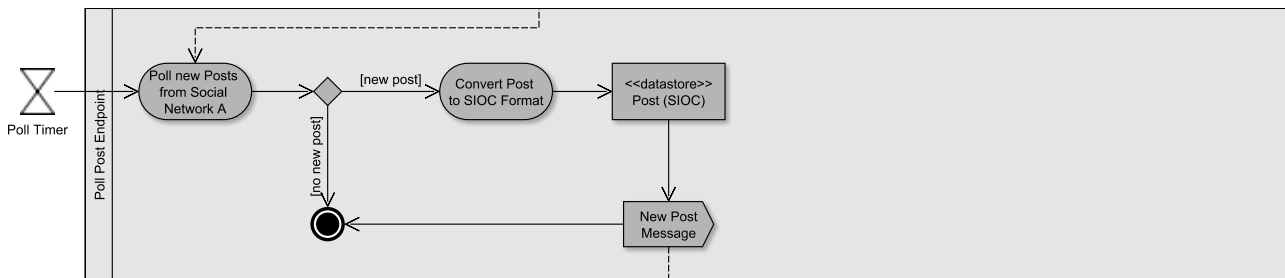


Abbildung 3.2.: Lesen des erstellten Beitrags und konvertieren in das Zwischenformat.

Liegen nun alle Daten in diesem Zwischenformat vor, könnten diese an einer Datenbank gespeichert werden. Diese Datenbank macht es möglich auf einfache Art und Weise auf die gespeicherten Daten von außerhalb zu zugreifen und diverse Abfragen ausführen zu können. Dies könnte zum Beispiel sein, in den Daten nach inhaltlich gleichen oder ähnlichen Beiträgen zu suchen oder passende externe Materialien aus Vorlesungen zu einem Thema vorzuschlagen. Die Möglichkeiten sind hier vielfältig.

Da ein Teil dieser Arbeit darin besteht Beiträge zwischen zwei oder mehr sozialen Netzwerken zu synchronisieren, muss das Eintreffen neuer Beiträge an die einzelnen Komponenten mitgeteilt werden. Da es unmöglich ist diesen Zeitpunkt vorher zu sagen ist der beste Weg einen ereignisorientierten Ansatz zu wählen. Hierbei wird beim lesen eines neuen Beitrags eine Ereignisnachricht

erzeugt, welche die interessierten Komponenten über das Eintreffen informiert. So wird eine zeitliche Entkopplung von Lesen und Schreiben möglich, gleichzeitig können sich einzelnen Komponenten an diesen Nachrichtenstrom an- oder abhängen ohne Andere zu stören. Dies erhöht die Flexibilität des Systems.

3.3 Beitrag in soziales Netzwerk B schreiben

Hat sich eine Komponenten und empfängt eine Ereignisnachricht von einen neuen Beitrag, wird ähnlich Verfahren wie schon beim Lesen, nur umgekehrt (Siehe Abbildung 3.3 links, oberer Ablauf). Der neue Beitrag wird vor den schreiben in das soziale Netzwerk B aus dem Zwischenformat in der Zielformat B konvertiert. Alle dazu nötigen Information können direkt aus der Ereignisnachricht oder zusätzlich aus der oben genannten Datenbank geholt werden.

Wünschenswert wäre es hierbei, wenn es so aussehen würde der Beitrag nicht vom hier entwickelten System, sondern von Autor des ursprünglichen Beitrags geschrieben worden. Hierzu es unerlässlich auf das Konto des Autor im entsprechenden Netzwerk zugreifen zu können, da im allgemeinen der Schreiben stellvertretend für dritte Personen nicht möglich ist. Hierzu muss zunächst für den betreffenden Autor das Konto für das soziale Netzwerk B herausgesucht werden (Siehe Abbildung 3.3 links, unterer Ablauf). Wurde ein passendes Konto gefunden, wird der konvertierte Beitrag über diesen geschrieben. Sollten kein Passender gefunden werden, müsste dies über ein vorher definiertes Konto geschehen. Dabei sollte aber auf den ursprünglichen Autor beziehungsweise Beitrag in irgendeiner Form hingewiesen werden. Dies kann zum Beispiel durch anbringen eines Links an den Text des Beitrags erfolgen.

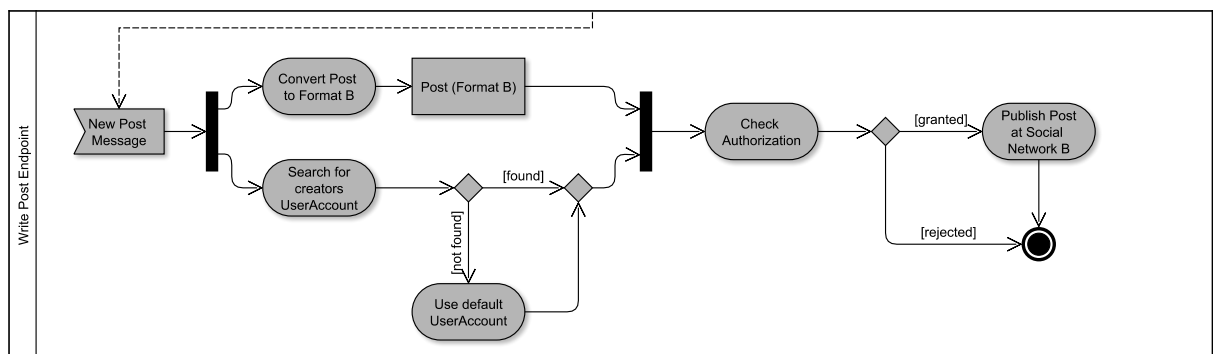


Abbildung 3.3.: Konvertierten des Beitrags in das Format B und schreiben in das soziale Netzwerk B

Gleichzeitig mit den Sammeln von Daten kommt immer auch das Thema zum Schutz der Privatsphäre auf. Wie soll damit umgegangen werden, wenn ein Benutzer nicht möchte dass seine Beiträge gesammelt werden oder automatisch weiter geleitet werden sollen? Hierzu wäre es sinnvoll die eben beschriebenen Abläufe so zu erweitern, dass der Benutzer festlegen kann bestimmte Quellen oder einzelne Teile für das automatische Lesen und Schreiben zu blockieren, wie es von Uldis Bojars et al. in [4] vorgeschlagen wird. Dies könnte durch eine Access Control List (ACL) realisiert werden. Hier könnte der Benutzer Lese und Schreibrechte für einzelne Bereiche festlegen, welche dann das System benutzt um einzelne Abläufe auszuführen oder abubrechen.

3.4 Identifizierung der Komponenten

Anhand dieses kurzen Ablaufbeispiels können wir nun einige Komponenten ablesen die das System unbedingt beinhalten muss und welche ergänzend dazu Wünschenswert wären:

- Eine Komponente muss Daten von einem sozialen Netzwerk in das System einlesen und diese in ein geeignetes Zwischenformat konvertieren können.
- 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 Konto eines Benutzer in einem Netzwerk suchen zu können.
- Um die Privatsphäre der Benutzer zu wahren, wäre eine ACL Mechanismus sinnvoll.

4 Eigener Ansatz

4.1 Social Online Community Connectors

4.1.1 Datenformat

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

4.1.2 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 4.1.1 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.

Connector Config Ontologie

Abbildung 4.1 zeigt die entwickelte Connector Con Ontologie. Sie besteht aus einer einzigen Klasse `ConnectorConfig` und fünf Eigenschaften für diese.

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 eins 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 des Verweises definiert werden. Innerhalb dieser Vorlage stehen einige

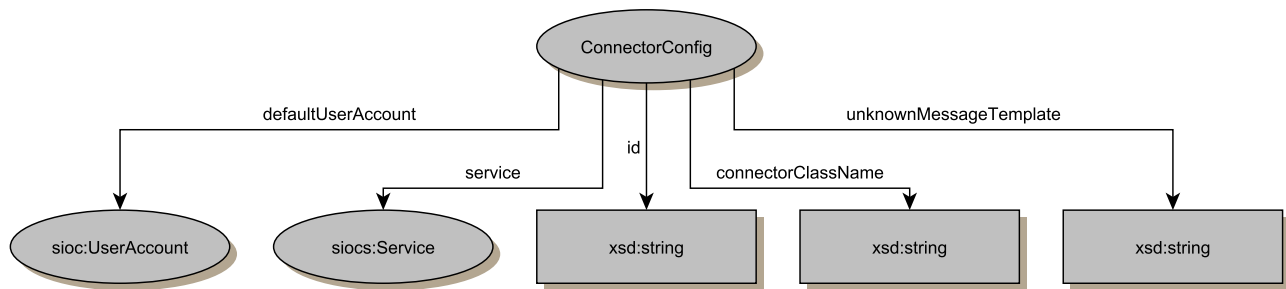


Abbildung 4.1.: Connector Config Ontology

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)

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

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 4.1.2 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.

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 4.2 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.

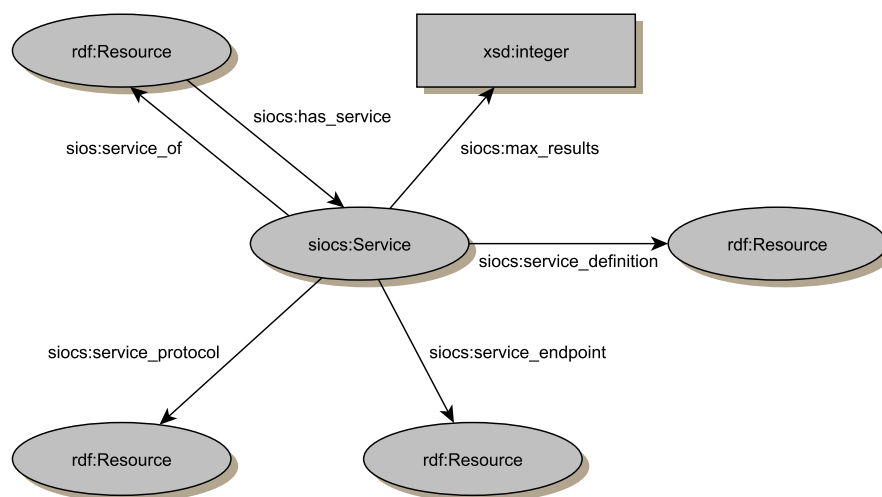


Abbildung 4.2.: SIOC Services Module

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.

Autorisierung und Authentifizierung

Die Wahl für SIOC als Datenformat zu hat sich nach den ersten Tests als eine sehr gute Entscheidung herausgestellt. Mittels SIOC ließen die wichtigsten Daten aller untersuchten Plattformen in einer einheitlichen Form speichern. Die ersten Probleme traten erst auf, als es daran ging Daten

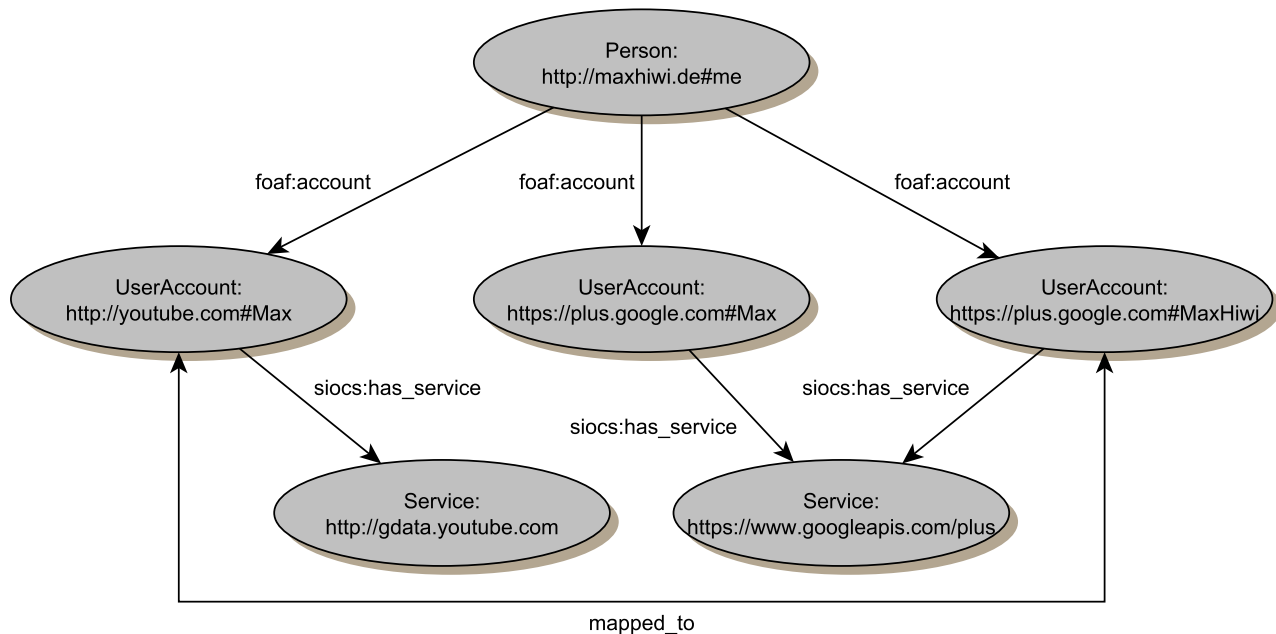


Abbildung 4.3.: Zusammenhang von Person, UserAccount und Service. Die inversen Eigenschaften `siocs:account_of` und `siocs:service_of` wurden zu einer besseren Übersicht weggelassen

für die Autorisierung (Feststellen ob jemand eine Handlung ausführen darf) beziehungsweise Authentifizierung (Feststellen ob jemand der ist, den er vorgibt zu sein) an den verschiedenen Programmierschnittstellen einen geeigneten Ort zu finden. Dazu musste aber erst festgestellt werden, welche verschiedenen Daten alle vorliegen können.

Username/Passwort ist wohl eine der ersten und häufigsten Mechanismen, um den Zugriff sensibler Daten vor Dritten zu schützen. Das in Abschnitt 5.1.1 beschriebene LMS Moodle, setzt zum Beispiel den Username und Password eines angemeldeten Benutzers zu Authentifizierung ein.

OAuth ¹ 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 eingesetzt². 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* [15][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[15][S. 9] nötig. Da dieser Accesstoken in der Regel nur eine bestimmte Zeit gültig ist, wird je nach Implementierung des Standards noch ein Refresh token mitgeliefert. Mit diesem Refresh token ist das Programm in der Lage ohne

¹ <http://oauth.net/>

² http://en.wikipedia.org/wiki/OAuth#List_of_OAuth_service_providers

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 5.1.4 beschriebene Google Youtube API hierzu ein gutes Beispiel.

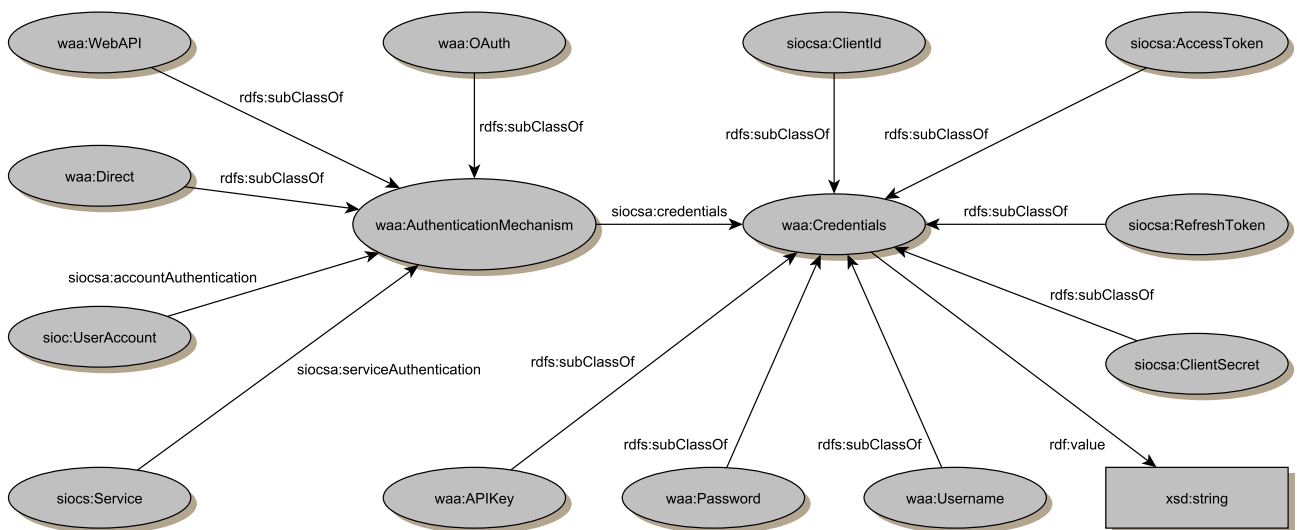


Abbildung 4.4.: SIOC Services Authentication Ontology

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 das 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* [17] mit Zugriffssteuerungsliste. Eine Ausnahme stellt die *Authentication Ontology*³ des *OmniVoke*⁴ Frameworks dar. Die Art der Authentifizierung wird darin durch die Klasse *AuthenticationMechanism* modelliert. Unterklassen für die wichtigsten Mechanismen wie OAuth, WebAPIs und Username/Password (dort *Direct* genannt) sind vorhanden. Jedem *AuthenticationMechanism* Objekt können dann sogenannte *Credentials* (engl. für Anmeldedaten) angehängt werden.

Das einzige Manko an dieser Ontologie war das Fehlen von *Credentials* für OAuth in der Version 2.0. Im einzelnen waren dies Klassen für `client_id`, `client_secret` sowie für Access- und

³ <http://omnivoke.kmi.open.ac.uk/authentication/>

⁴ <http://omnivoke.kmi.open.ac.uk/framework/>

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 Schlüssen 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 *sio:sa:*) und die übernommenen Teile der Authentication Ontology wurden danach im *SIOC Services Authentication Module* zusammengefasst. Graphisch ist sie in Abbildung 4.4 und im Anhang A.2 als OWL Schema zu sehen.

Autorisierung

Da für viele Menschen im Internet ihre Privatsphäre und nicht wollen, dass ohne ihr Wissen Informationen von ihnen weitergegeben werden, ist es wichtig von den Benutzern die Erlaubnis zu holen Beiträge weiter zu leiten. Ein verbreitetes Mittel für eine solche Zugriffssteuerung sind Access Control Lists (ACL) (engl. für Zugriffssteuerungsliste). Die anderen nur einen begrenzten Zugriff auf eine feste Ressource gewähren. Für den Einsatz in dieser Arbeit wurde die Basic Access Control Ontologie [17, 1] ausgewählt (Im Folgenden nur als ACL bezeichnet). Der Hauptgrund lag darin, dass sie auf FOAF aufbaut und sich so einfach integrieren ließ.

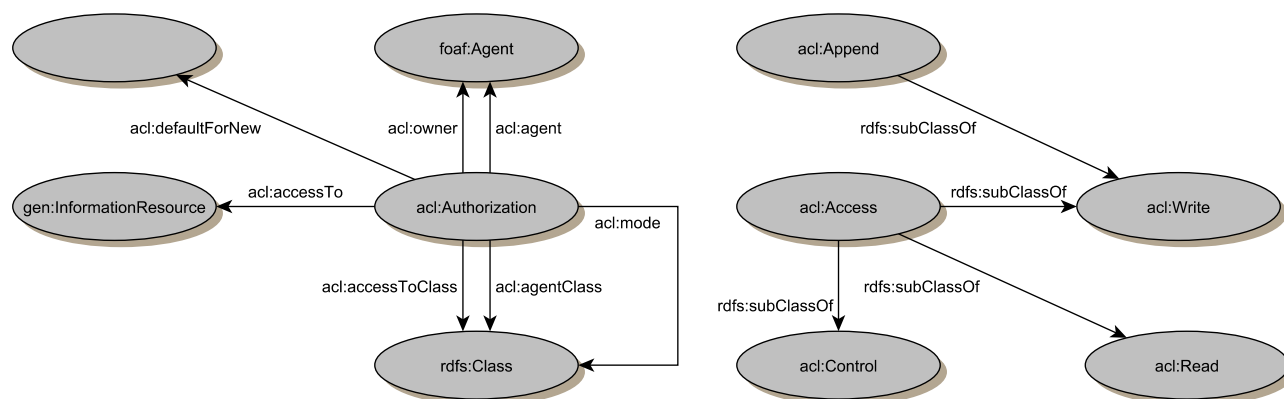


Abbildung 4.5.: Basic Access Control Ontologie

Die ACL stellt die Klasse `acl:Access` für die Abbildung eines Zugriffsrechts auf eine Ressource zur Verfügung. 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 Inhalten. Diese sind für diese Arbeit aber nicht von Bedeutung.

Die eigentliche Regel für die ACL besteht aus der Klasse `acl:Authorization`. Der Besitzer dieser Regel wird durch die Eigenschaft `acl:owner` festgelegt. Der Besitzer ist im Fall von SOCC ein Objekt der Klasse `Person`, wie in Abschnitt 4.1.2 „Benutzerdaten“ beschrieben, da diese eine Unterklasse von der ACL geforderten `foaf:Agent` ist. Mittels `acl:agent` kann signalisiert werden,

dass diese Regel nur für eine bestimmte Person/Agenten gilt. Das selbe gilt für `acl:agentClass`, wobei hier eine bestimmte Klasse gemeint ist. Soll zum Beispiel ein öffentlicher Zugriff definiert werden, wird `foaf:Agent` eingesetzt [1, „Public Access“]. Innerhalb von SOCC werden nur Regel angewendet, die einen solchen öffentlichen Zugriff erlauben. Die eigentlichen Rechte werden über die Eigenschaft `acl:mode` festgelegt. Erlaubt ist die Angabe jeder beliebigen Klasse, SOCC testet aber nur auf die Klassen `acl:Read` und `acl:Write` zum Lesen und Schreiben von Beiträgen. Auf welche Ressource sich ein Regel letztendlich 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. Für einen Zugriff auf alle Beiträge prüft SOCC ob die Eigenschaft `acl:accessToClass` auf die Klasse `sioc:Post` verweist. So müsste nicht jede einzelne Seite angegeben werden.

Das Listing 4.1 zeigt ein Beispiel wie ein ACL Regel aussehen könnte. Der Besitzer dieser Regel wird durch die URI `http://example.org#john` beschrieben (Zeile 6). Diese Person erlaubt nun öffentlichen Zugriff (Zeile 7) all all seine Beiträge (Zeile 8). Dieser Zugriff kann sowohl lesend als auch schreibend erfolgen (Zeile 9).

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, acl:Write .

```

4.1.3 Aufbau der SOCC

4.1.4 Design eines Connectors

SOCC Context

Der Kontext eines Connectors beschreibt die Umgebung innerhalb dem er seine Arbeit verrichtet. Im aktuellen Status erhält der Connector zum einen Zugriff auf ein außenstehendes Model (TripleStore), das wichtige Daten für den Betrieb und enthält oder abgelegt werden können. Eine Referenz auf dieses Model erhält der Connector über den Aufruf der Funktion `getModel()`. Durch die Methode `getAccessControl()` kann der Connector über die im nächsten Abschnitt beschriebene AccessControl Schnittstelle auf die Information für die Zugriffssteuerung für das Lesen und Schreiben von Beiträgen.

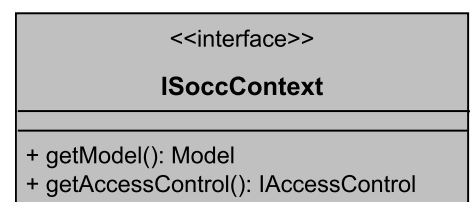


Abbildung 4.8.: SOCC Context

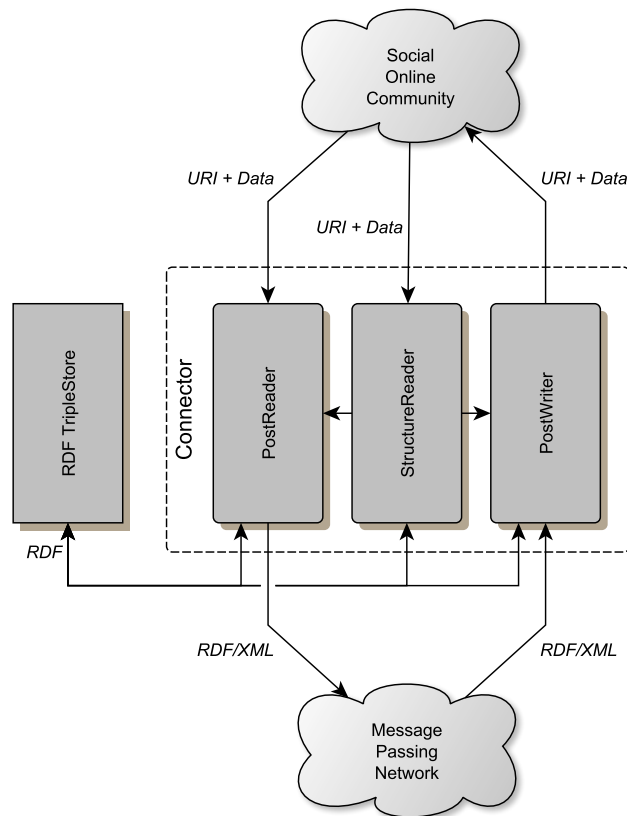


Abbildung 4.6.: Übersicht der Komponenten der SOCC

AccessControl

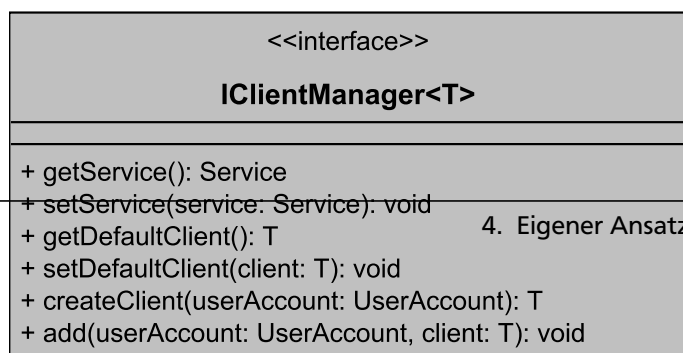
Die AccessControl Schnittstelle ist sehr einfach gehalten und dient für den Zugriff auf die in Abschnitt 4.1.2 „Autorisierung“ beschriebene ACL. Die Methode `checkAccessTo(...)` prüft, ob der Zugriff auf eine Ressource mit allen übergebenen Zugriffsmodi erlaubt ist. Die andere Methode `checkAccessToClass` ist zur Überprüfung, die die Rechte für den Zugriff auf eine komplette Klasse von Ressourcen.



Abbildung 4.9.: AccessControl

ClientManager

Der Zugriff auf eine API innerhalb eines Programms erfolgt in der Regel über eine sogenannte Clientobjekt (kurz Client). Dieser Client erlaubt es mit den Anmeldedaten oder den Accesstoken für ein Benutzer-



4. Eigener Ansatz

konto 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 ClientManager. Der interne Aufbau eines Client ist dabei stark von der verwendeten API abhängig und arbeitet mit dem Connector zusammen für den er geschrieben wurde. Für alle benutzerunabhängigen Daten erhält der ClientManager ein wie Abschnitt 4.1.2 beschriebenes Serviceobjekt. Ein neuer Client kann dann durch den Aufruf der Methode `createClient(...)` erstellt werden. Als Parameter wird der Methode ein Benutzerkonto in Form eines SIOC UserAccounts übergeben. Sind alle erforderlichen Autorisierung- und Authentifizierungsinformation aus Abschnitt 4.1.2 vorhanden, wird ein neuer Client erzeugt und zurück gegeben. Dieser Client wird aber dadurch nicht automatisch vom ClientManager verwaltet. Hierzu muss der im vorherigen erzeugte Client durch den Aufruf von `add(userAccount: UserAccount, client: T)` dauerhaft mit den angegebenen UserAccount verknüpft und intern gespeichert. Wichtig ist hierbei, 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)` Holen von Clients, sowie `contains(userAccount: UserAccount)` für Tests ob ein Client zu einem UserAccount existiert. 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 4.1.2 angesprochenen Defaultuser. Dieser Defaultclient genannte Client kann über die Methode `setDefaultClient(client: T)` gesetzt und durch `getDefaultClient()` jederzeit wieder abgerufen werden.

StructureReader

Um auf Informationen über die Struktur von Foren, sozialen Netzwerken und so weiter im SIOC Format zugreifen zu können, implementiert jeder Connector dazu einen StructureReader. Die Struktur lässt sich, wie im Abschnitt 4.1 vorgestellt, durch die SIOC Klassen *Site* und *Container* (und Unterklassen davon) beschrieben. Um auf diese Struktur zugreifen zu können, enthält die StructureReader Schnittstelle mehrere Methoden (Siehe Abbildung 4.11).

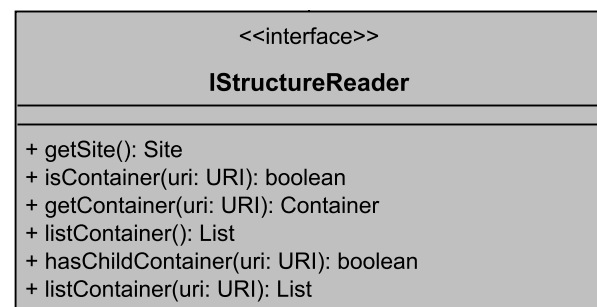


Abbildung 4.11.: StructureReader

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

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

`getContainer(URI)` ist dazu gedacht die Information eines einzelnen Containers erhalten der sich hinter eine URI verbirgt.

`listContainer(...)` sind Methoden welche für den die Auflisten aller Container einer Seite zur Verfügung stehen. Die Methode ohne Parameter listet alle Container auf der ersten Ebene auf. Dies könnten zum Beispiel alle Kurse auf einen Canvas LMS Seite oder alle Gruppen auf Facebook sein. Die zweite Methode mit URI Parameter gibt eine Liste aller Container, welche den Container hinter der übergebenen URI als Elternteil haben, zurück. Als Beispiel wären alle Themen innerhalb eines Forums zu nennen.

`hasChildContainer(uri: URI)` überprüft ob der Container hinter einer URI überhaupt 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 auftritt.

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 dieses Beitrags das Lesen dafür erlaubt hat. Falls nicht, wird der Beitrag aus der Ergebnisliste gelöscht oder ein Fehler ausgegeben. Die Funktionsweise der einzelnen Methoden ist wie folgt:

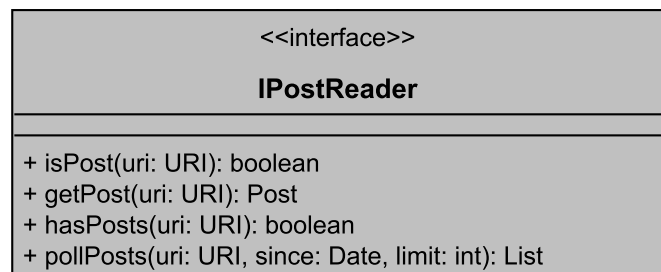


Abbildung 4.12.: PostReader

`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 SIOC Post Objekt zurück beziehungsweise einen Fehler, falls der Beitrag nicht mit diesem Connector gelesen werden kann.

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

`pollPosts(uri: URI, since: Date, limit: int)` ist eine Methode mit der alle Beiträge hinter eine URI liest und anhand der übergebenen Kriterien filtert. Insgesamt erhält diese Methode drei Parameter. Der Erste ist eine URI die den Ort angibt von der alle Beiträge gelesen werden können. 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 folgten zu lesen. Der letzte Parameter gibt eine obere Schranke an, wie viele Beiträge maximal pro Aufruf dieser Methode gelesen werden dürfen.

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 einer SOC geschrieben werden, wird die Methode `writePost(URI, String, Syntax)` mit dem Zielort als URI, dem Beitrag als serialisiertes RDF Objekt und dem verwendeten Serialisierungsformates aufgerufen. Begonnen wird damit, dass als erstes nach einem UserAccount für den Service des aktuellen Connectors des Beitragautors gesucht. Im Idealfall befindet sich für den UserAccount des Beitragsautors ein Link zu seiner FOAF Person und von ein weiterer Link zum UserAccount für den aktuellen Service. Mit diesem UserAccount kann dann vom ClientManager ein Clientobjekt für die verwendete API angefordert werden. Sollte die Suche negativ verlaufen, steht der Defaultclient zur Verfügung. Mit diesem Client, ob mit der des Autors oder dem Defaultclient, wird im letzten Schritt der Beitrag im von der API verwendeten Format in die SOC geschrieben.

4.2 SOCC-Camel

4.2.1 SoccComponent

4.2.2 SoccPostPollingConsumer

4.2.3 SoccPostProducer

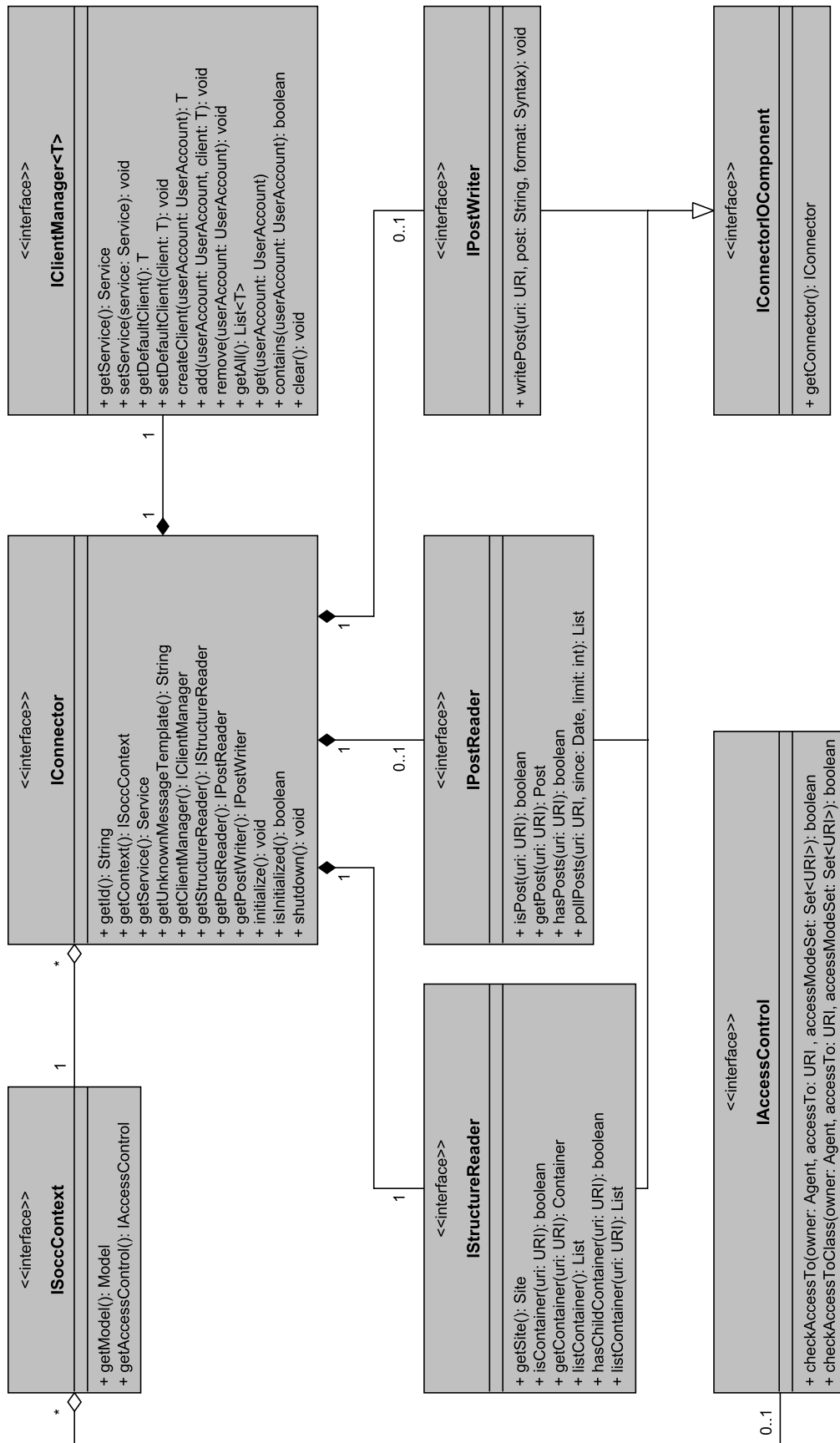


Abbildung 4.7.: UML Klassendiagramm eines Connectors

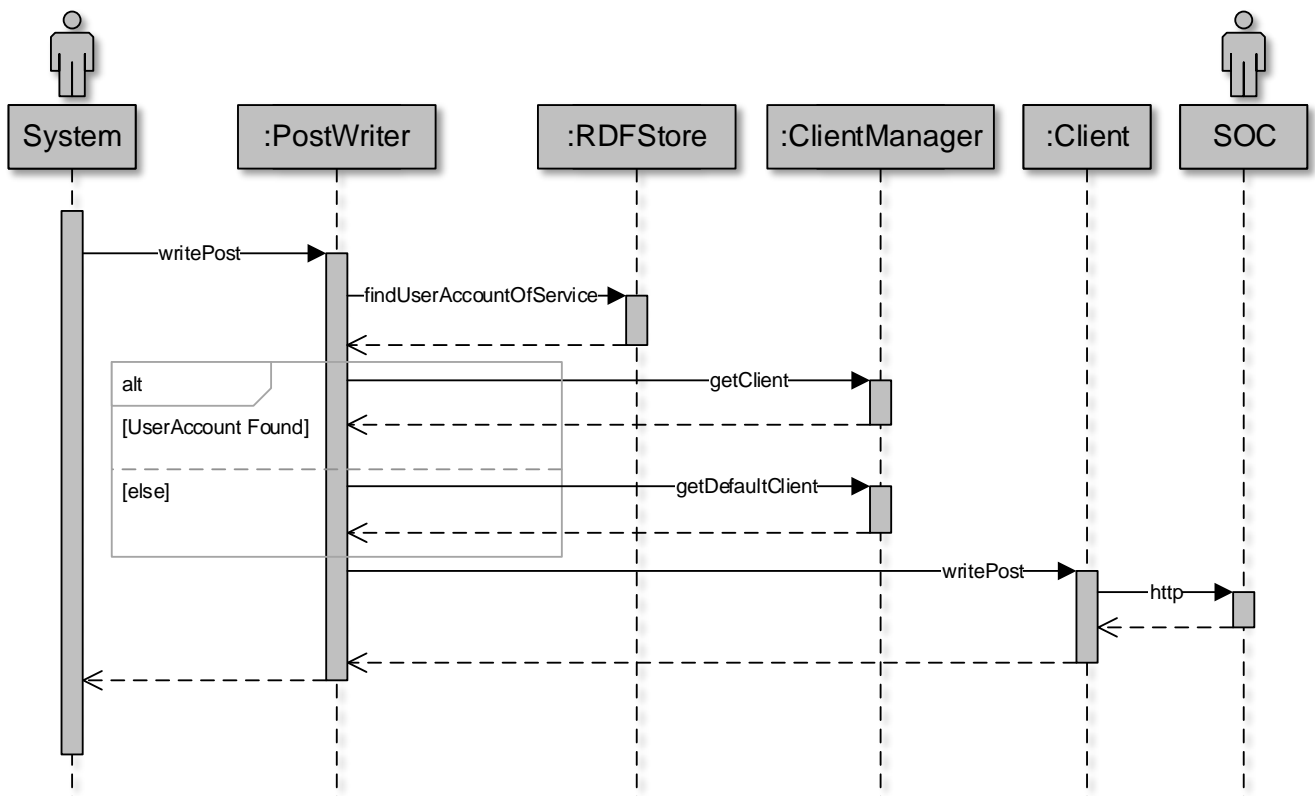


Abbildung 4.13.: UML Sequenzdiagramm eines PostWriters

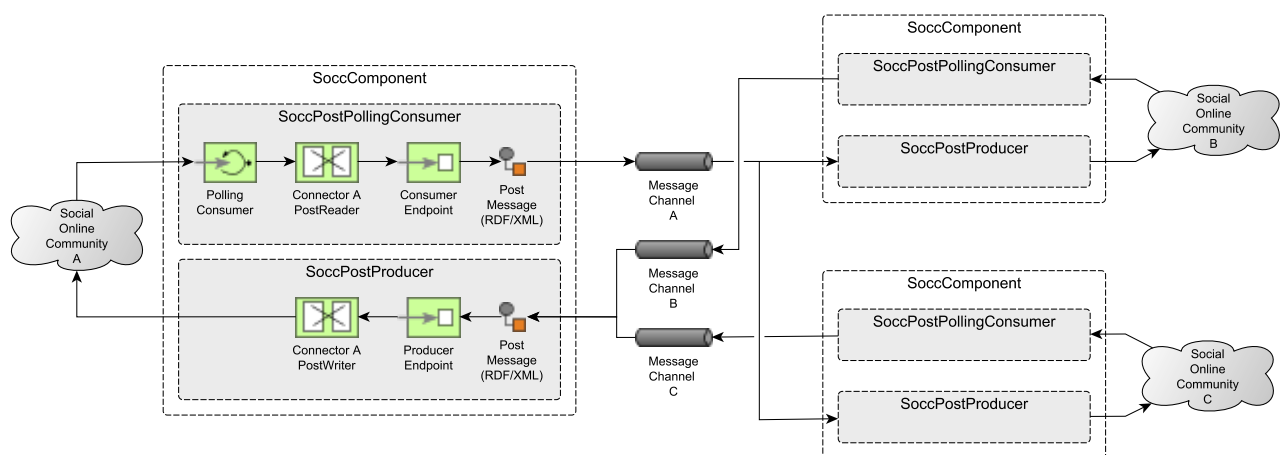


Abbildung 4.14.: Übersicht des Apache Camel Moduls Socc-Camel



5 Implementierung

5.1 Implementierung der Connectoren

Das muss rein:

- Mapping nach SIOC
- Zugriff über die API
- Probleme bei der Implementierung

5.1.1 Moodle

- Eingebaute REST Schnittstelle, aber kein Lesen von Beiträgen
- Webservice Plugin MoodleWS (REST oder SOAP)
 - <https://github.com/patrickpollet/moodlews>
 - ClientAPI existieren von selber Autor
 - REST defekt, kein schreiben von Beiträgen möglich
 - SOAP funktioniert mehr oder weniger
 - Verschluckt Fehlermeldungen
 - kein lesen einzelner Posts/Threads/Foren
 - SOAP ClientAPI neu generieren, weil vorhandene nicht mit 2.4 funktioniert.
 - Username/Password + Session Token/Id
 - “Use an auto generated wsdl” -> No
 - schreiben von neuen Beitrag direkt in thread nur als Antwort auf ersten Beitrag möglich
 - Rückgabe aller Beiträge in einem Objekt

SIOC Mapping

API

Herausforderungen

5.1.2 Facebook

- REST API + JSON
- keine offizielle Java API für Desktop -> Web + Mobile only
- GraphAPI, Facebook Query Language
- OAuth 2.x
 - kein Refresh token
 - Token Haltbarkeit 2h (2 Monate, wenn extended)
 - token nur über webbrowser
- RestFB alternative Java API für die REST Schnittstelle der GraphAPI
- Typ der zurückgelieferten Daten nicht anhand der URI erkennbar, häufig erst durch Angabe von *metadata=1*
- beim herunterladen einzelner Posts nicht immer erkennbar wo sie geschrieben wurden

SIOC Mapping

API

Herausforderungen

5.1.3 Google+

- Einfach REST API + JSON
- OAuth
 - Refresh token (token laufen quasi nie ab)
 - holen von token ohne webbrowser möglich
- Objekte aufgebaut aus Actor (wer machte was), Verb (wie machte er es), Object (was machte er) + Metadata
- verschieden Sprachen + Plattformen
- lesen nur von öffentlichen Beiträgen
- kein Schreiben von Beiträgen

SIOC Mapping

API

Herausforderungen

5.1.4 Youtube

- Aktueller Umbau der API (ähnlich google+) v3
 - keine lesen von kommentaren
 - kein schreiben
- alte GData Feed API v2 basiert auf RSS + Youtube Erweiterung
- Mapping teilweise durch basis auf RSS einfach, manchmal auch nicht
- Wichtigen Metadaten nur implizit vorhanden (comment id in uri aber nicht in datenformat)

SIOC Mapping

API

Herausforderungen

5.1.5 Canvas

- relativ neues LMS
- super Bedienung
- super REST API
- keine Java API
- rudimentäre Eigenentwicklung einer Java API, Funktionsweise ähnlich G+
- viel API Funktionen wohl nicht extern nutzbar (UserProfil lesen, vll. Falsche Berechtigung -> test nötig)

SIOC Mapping

API

Herausforderungen



6 Evaluation



7 Abschlussbetrachtung

7.1 Fazit

7.2 Ausblick



Literaturverzeichnis

- [1] WebAccessControl. <http://www.w3.org/wiki/WebAccessControl>, Zugriff: 2013-09-12.
- [2] David Beckett and Tim Berners-Lee. Turtle - Terse RDF Triple Language. <http://www.w3.org/TeamSubmission/turtle/>, Zugriff: 2013-09-13, 2011.
- [3] Tim Berners-Lee. Socially Aware Cloud Storage. <http://www.w3.org/DesignIssues/CloudStorage.html>, Zugriff: 2013-08-26, 2011.
- [4] 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.
- [5] 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.
- [6] Dan Brickley and Libby Miller. FOAF Vocabulary Specification 0.98. <http://xmlns.com/foaf/spec/>, Zugriff: 2013-09-13, 2010.
- [7] 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.
- [8] Jo Davies and Martin Graff. Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, 36(4):657–663, 2005.
- [9] Digital Enterprise Research Institute. SIOC - Semantically-Interlinked Online Communities. <http://sioc-project.org/>, Zugriff: 2013-08-15.
- [10] Stephen Downes. E-learning 2.0. *eLearn Magazine*, 2005.
- [11] Facebook. Key Facts. <http://newsroom.fb.com/Key-Facts>, Zugriff: 2013-09-11, 2013.
- [12] 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.
- [13] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. PhD thesis, University of California, 2000.
- [14] 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.
- [15] D. Hardt. The OAuth 2.0 Authorization Framework. <http://tools.ietf.org/html/rfc6749>, Zugriff: 2013-08-30, 2012.

-
- [16] 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.
- [17] 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.
- [18] 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.
- [19] Graham Klyne and Jermey J. Carrol. Resource Description Framework (RDF): Concepts and Abstract Syntax. <http://www.w3.org/TR/rdf-concepts/>, Zugriff: 2013-08-19, 2004.
- [20] Frank Manola and Eric Miller. RDF Primer. <http://www.w3.org/TR/rdf-primer/>, Zugriff: 2013-08-19, 2004.
- [21] 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.
- [22] Nilo Mitra and Yves Lafon. SOAP Version 1.2. <http://www.w3.org/TR/2007/REC-soap12-part0-20070427/>, Zugriff: 2013-09-09, 2007.
- [23] 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.
- [24] Felix Schwenzel and Sascha Lobo. Reclaim Social. <http://reclaim.fm/>, Zugriff: 2013-08-14.
- [25] 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.
- [26] 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.
- [27] Youtube. Statistik. <http://www.youtube.com/yt/press/de/statistics.html>, Zugriff: 2013-09-10.

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`