Distributed Discussions in Online Social Networks

Masterarbeit

Florian Müller
Betreuer: Prof. Dr. Max Mühlhäuser
Verantwortlicher Mitarbeiter: Dipl.-Inform. Kai Höver
Darmstadt, 26. August 2013

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Informatik
Telekooperation
Prof. Dr. Max Mühlhüuser

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, den 26. August 2013

_____

(Florian Müller)

Zusammenfassung

Inhalt...

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Liste der noch zu erledigenden Punkte

1 Einleitung

- Motivation/Problemstellunng
- Anforderungen: was soll entwickelt werden
- Übersicht über alle Kapitel

Anfang

Jedem ist es heutzutage möglich eigene Inhalte ohne großen Aufwand ins Internet zu stellen und anderen an seinen Wissen teilhaben zu lassen. Dadurch was es noch nie so einfach an Wissen Dritter zu kommen wie heute und dieses Wissen in seinen eigenen Lernprozess mit einfließen zu lassen.

Soziale Netzwerke erleben seit den letzten Jahren einen großen Boom. Sie ermöglichen mit seinen Freunden in Kontakt zu bleiben über zeitliche und räumliche Hürden hinweg. Diese sozialen Netzwerke erlauben es auch ohne physische Präsenz sich untereinander zu organisieren. Qiyun Wang et. al. [8] zeigten, dass zum Beispiel Facebook[1] sich hervorragend für den Einsatz als Lernplattform beziehungsweise Learning Management System (LMS) eignet. Gerade die Organisation der Lerngruppe als auch die Benachrichtigung über Ereignisse funktionierte reibungslos. Jedoch uneingeschränkt konnte Facebook als LMS nicht empfohlen werden. Bemängelt wurden unter anderem die aufwändige Integration von Lernmaterialien „tutor noticed that it was quite troublesome to add teaching materials“[8, S. 435]. Aber auch da es nicht möglich war Diskussionen in einzelne Themen zu unterteilen sondern alle Beiträge nur chronologisch angeordnet sind ist negativ aufgefallen.

[1] `http://www.facebook.com`

2 Grundlagen

2.1 Semantic Web

2.1.1 Resource Description Framework

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 [5, 6] 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“[5].

Das Datenmodell von RDF ist sehr einfach aufgebaut um es effizient verarbeiten zu können. Die Grundlage bilden simple Tripel aus Subjekt, Prädikat und Objekt, welche an die natürliche Sprache angelehnt als Sätze[4] bezeichnet werden. Mehrere solcher Tripel bilden einen RDF Graphen. Das Prädikat beschreibt hierbei eine Beziehung zwischen Subjekt und Objekt und wird daher oft als Eigenschaft beschrieben. In der natürlichen Sprache kann man dies zum Beispiel so ausdrücken; Das Subjekt hat eine Eigenschaft mit den Wert Objekt. Graphisch wird dies durch eine gerichtete Verbindung von Subjekt und Objekt mit der Beschriftung der Prädikats dargestellt (siehe Abbildung 2.1).

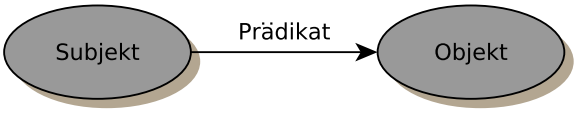

Abbildung 2.1.: Graphische Darstellung eines RDF Tripels

Für Subjekt, Prädikat und Objekt besteht die Möglichkeit sogenannte Uniform Resource Identifiers (URI), Literale oder leere Knoten als Werte einzusetzen.

URIs sind eindeutige Bezeichner die eine beliebige reale oder abstrakte Ressource und werden wie in RFC 2396[1] beschrieben formatiert, wobei relative URIs nach [5] nicht verwendete werden sollen.

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"^^xsd:integer` erweitert werden. Bei Literalen ist aber auch darauf zu

[1] `http://www.isi.edu/in-notes/rfc2396.txt`

achten, dass `"Objekt` und `"Objekt"@de` beschreiben zwar beide den gleichen Wert werden aber von RDF nicht als gleich angesehen. Zwei Literale können nur gleich sein, wenn sie die selbe Sprache beziehungsweise den selben Datentyp besitzen.

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 aber nicht unbedingt eine eigene URI nötig ist und sind nur innerhalb eines Graphen eindeutig

Beispiel

.

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 verwendeten.

RDF/XML, N3

RDF/XML

N3 ist die Kurzform für Notation 3 und wurde von Tim Berners-Lee als Sprache für RDF entwickelt. Tripel in N3 werden dabei wie Sätze in meisten natürlichen Sprachen geschrieben. Erst das Subjekt, dann das Prädikat und am Ende das Objekt gefolgt von einen Punkt, wie in Listing 2.1 zu sehen ist.

Listing 2.1: Einfaches N3 Beispiel

```
1 <http://example.de/florian> <http://example.org/#name> "Florian" .
2 <http://example.de/florian> <http://example.org/#age> "28" .
```

Die URI `http://example.de/florian` beschreibt hierbei eine Ressource welche einen Namen `Florian` und ein Alter `28` besitzt. Es ist darauf zu achten, dass alle URI immer zwischen spitzen Klammern stehen. Da nun einzelne Prädikate recht häufig innerhalb eines Graphen auftauchen können, kann es einfach sein diese abgekürzt schreiben zu können. Hierzu ist es möglich sogenannte Präfixe (auch Namensräume genannt) am Beginn des Dokumentes zu definieren und einen so Schreibarbeit abzunehmen.

Listing 2.2: Präfixe

```
1 @prefix person: <http://example.org/#> .
2 <http://example.de/florian> person:name "Florian" .
3 <http://example.de/florian> person:age "28" .
```

Am Anfang von Listing 2.2 wird durch Einleiten mittels des Schlüsselwortes `@prefix` ein neuer Präfix `person:` für die URI `http://example.org/#` festgelegt (man beachte wieder den Punkt am Ende der Zeile). Dieser Präfix kann nun überall innerhalb des Dokumentes verwendet werden, wobei die spitzen Klammern der vorherigen URI weggelassen werden können.

Leere Knoten + kurz Turtle

2.1.2 Abfragesprache SPARQL

2.1.3 Ontologien

3 Verwandte Arbeiten

- Verwendete Ansätze, Methoden und/oder Modelle (Sprachen, Entwurfsmethoden,Datenmodelle, Analysemethoden, Formalismen)

3.1 Semantic Interlinked Online Communities

Semantic Interlinked Online Communities[1] (SIOC, ausgesprochen „schock“) ist ein Projekt, welches von Uldis Bojārs und John Breslin begonnen wurde um unterschiedliche, webbasierte Diskussionslattformen(Blog, Forum, Mailinglist,. . .) untereinander verbinden zu können [2]. Der Kern von SIOC besteht aus einer Ontologie, welche den Inhalt und die Struktur diese 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 designt. 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 Friend of a Friend[2] (FOAF) und für einige Inhaltliche- und Metadaten (Titel, Inhalt, Erstelldatum, . . .) Dublin Core Terms[3] eingesetzt.

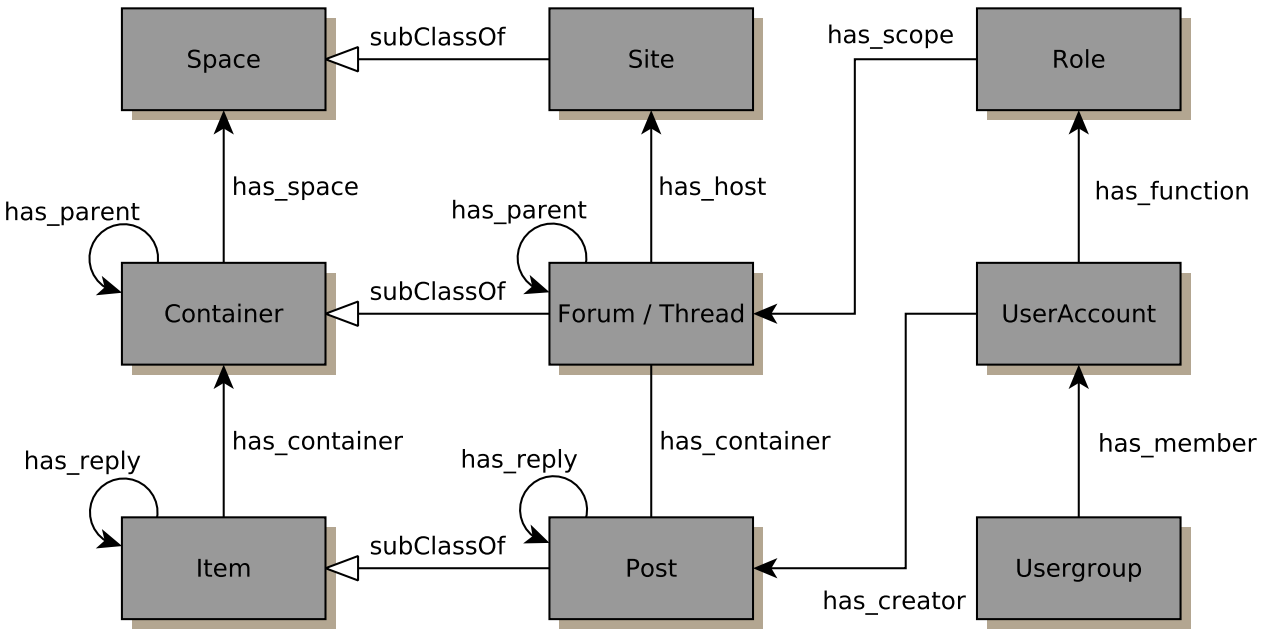

Abbildung 3.1.: Aufbau von SIOC (modifiziert) - Originalquelle: [3]

[1] http://sioc-project.org/
[2] http://www.foaf-project.org/
[3] http://dublincore.org/documents/dcmi-terms/

3.2 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 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:publica[4] 2013 ihr gestartetes Projekt „Reclaim Social“ [7] vorgestellt.

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[5]“ und der dafür vorhandenen Erweiterung „FeedWordPress“[6]. Diese Kombination ermöglichst alle Internetseiten, welche einen RSS Feed[7] anbieten, in die Datenbank von WordPress zu spiegeln. Das Problem hierbei besteht darin, dass einige sehr beliebte Internetseiten solche RSS Feeds nicht anbieten (`https://facebook.com`, `https://plus.google.com`) oder eingestellt haben (`https://twitter.com`). Für einige solcher Seiten wurden „proxy-scripte“[7, Tecg 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“[7, How Does It Work]

[4] `http://re-publica.de/`
[5] `http://wordpress.org/`
[6] `http://feedwordpress.radgeek.com/`
[7] `http://www.rssboard.org/rss-specification`

4 Anforderungsanalyse

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.Das vollständige Ablaufdiagramm befindet sich im Anhang A.1.

4.1 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 4.1).

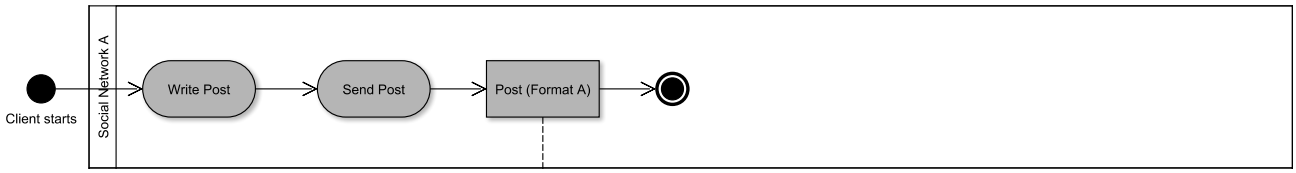

Abbildung 4.1.: Benutzer erstellt einen Beitrag im sozialen Netzwerk A.

4.2 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 Netzwerks 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 4.1 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 4.1.: 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 4.2 verdeutlicht den Ablauf unter Verwendung des eben beschriebenen Zwischenformats.

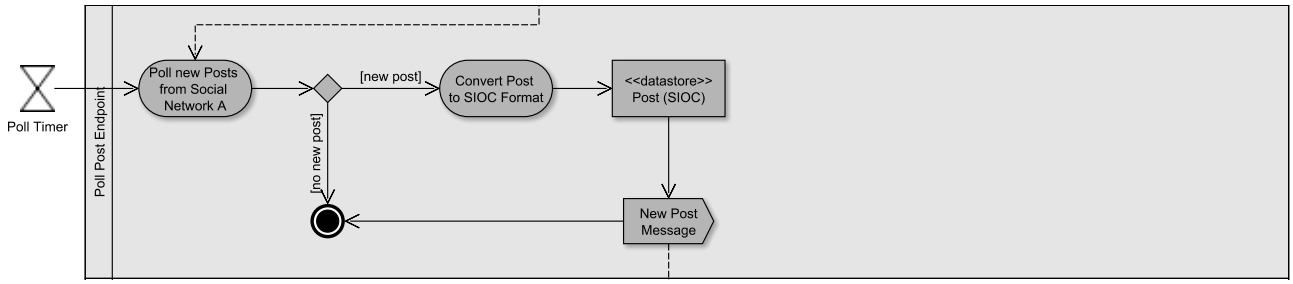

Abbildung 4.2.: Lesen des erstellten Beitrags und konvertieren in das Zwischenformat.

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 einen 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.

4.3 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 4.3 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 4.3 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.

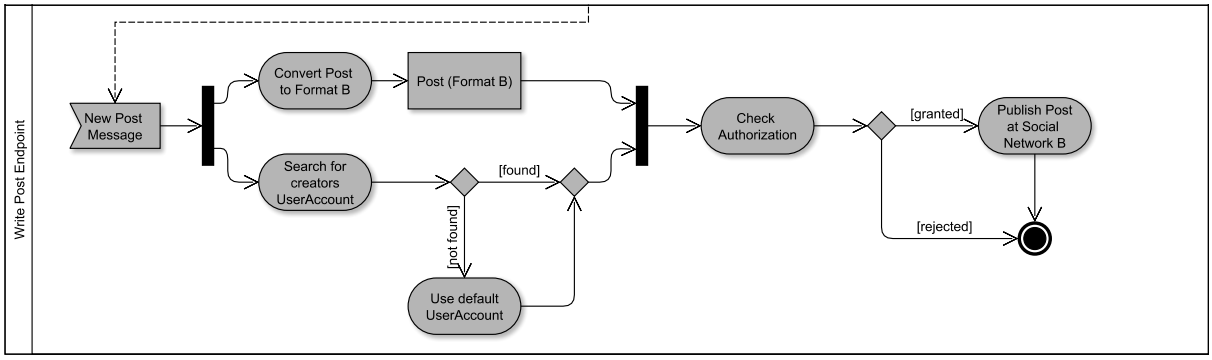

Abbildung 4.3.: Konvertierten des Beitrags in das Format B und schreiben in das soziale Netzwerk B

Gleichzeitig mit den Sammeln von Daten kommt immer auch das Thema zum Schutz der Privatsphäre auf. Wie soll damit umgegangen werden, wen ein Benutzer nicht möchte dass seine Beiträge gesammelt werden oder automatisch weiter geleiten werden sollen? Hierzu wäre es sinnvoll die eben beschriebenen Abläufe so zu erweitern, dass der Benutzer festlegen kann bestimme Quellen oder einzelne Teile für das automatische Lesen und Schreiben zu blockieren, wie es von Uldis Bojars et al. in [1] 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 abzubrechen.

4.4 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 einen sozialen Netzwerk in das System einlesen und diese in ein geeignetes Zwischenformat konvertieren können.
- Eine weitere Komponente nimmt Beiträte 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.

5 Design

- Die sollen die Anforderungen aus Kapitel 4 letztendlich in ein funktionstüchtiges Programm umgesetzt werden.

5.1 Wahl des Zwischenformats

- Warum kein eigenes Format
- Warum nicht RSS
- Warum nicht Dublincore + X
- Warum SIOC + FOAF + X

5.1.1 RSS

5.1.2 Dublin Core

5.1.3 SIOC und FOAF

5.2 Social Online Community Connectors

Einleitung über den Sinn und Zweck eines Connectors in SOCC

5.2.1 Aufbau

5.2.2 SIOC Service Auth Ontology

5.2.3 SOCC ConnectorCFG Ontology

5.2.4 W3 Web ACL Ontology

5.3 Routen der Einträge

- Eigener Ansatz mit JMS
- Apache Camel

6 Implementierung

- Probleme und deren Lösung während der Umsetzung
- Beispiel: Zugriff auf Moodle über Webservice

6.1 Verwalten SIOC Daten

- Sesame, Jena -> RDF2Go
- In Memory, XML/RDF File, Tripplestore

6.2 Zugriff auf Online Netzwerke und Abbildung in SIOC

6.2.1 Moodle

6.2.2 Canvas LMS

6.2.3 Facebook

6.2.4 Google Plus

6.2.5 Youtupe

6.3 Implementieren der Connectoren

-

7 Fazit

Literaturverzeichnis

[1] 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.

[2] 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.

[3] Digital Enterprise Research Institute. SIOC - Semantically-Interlinked Online Communities. \url{http://sioc-project.org/},[accessed: 2013-08-15].

[4] Hans-Werner Heinzen. Primer: Getting into RDF \& Semantic Web using N3 Deutsche Übersetzung. \url{http://www.bitloeffel.de/DOC/2003/N3-Primer-20030415-de.html}, accessed 2013-08-19.

[5] Graham Klyne and Jermey J. Carrol. Resource Description Framework (RDF): Concepts and Abstract Syntax. \url{http://www.w3.org/TR/rdf-concepts/}, accessed 2013-08-19, 2004.

[6] Frank Manola and Eric Miller. RDF Primer. \url{http://www.w3.org/TR/rdf-primer/}, accessed 2013-08-19, 2004.

[7] Felix Schwenzel and Sascha Lobo. Reclaim Social. \url{http://reclaim.fm/},[accessed: 2013-08-14].

[8] 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.

A Anhang

A.1 Anforderungsanalyse Ablaufdiagramm