



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Projektbericht

Ivan Demin

**SecondCast - Entwicklung einer Untersuchungsplattform zur
Durchführung von Second Screen Experimenten**

*Fakultät Technik und Informatik
Studiendepartment Informatik*

*Faculty of Engineering and Computer Science
Department of Computer Science*

Ivan Demin

**SecondCast - Entwicklung einer Untersuchungsplattform zur
Durchführung von Second Screen Experimenten**

Eingereicht am: 18. Mai 2015

Inhaltsverzeichnis

Verzeichnisse	iv
1. Einführung	1
2. SecondCast-Client	3
3. SecondCast-Server	5
3.1. Arbeitsweise von Dexter	5
3.2. Systemaufsetzung von Dexter	6
3.3. Evaluierung	7
4. Herausforderungen	10
5. Schlussbetrachtung	11
A. Datensätze	13

Verzeichnisse

Tabellenverzeichnis

3.1. Gesamtergebnis der Performance von Dexter anhand der untersuchten Dokumentationsbeschreibungen.	9
A.1. Detaillierte Tabelle von Ergebnissen aus der Performance-Untersuchung von Dexter mit Titel, Link zum Beschreibungstext, Recall-, Precision- und F1 score-Werten.	13

Abbildungsverzeichnis

1.1. Angestrebte Gesamtansicht als Komponentendiagramm von der Second Screen Untersuchungsplattform SecondCast.	1
2.1. Umgesetzte User Story für die Anzeige von Dokumentationen auf dem First Screen, sowie der Bereitstellung von Zusatzinformationen auf dem Second Screen.	4
3.1. Architektur von Dexter (Ceccarelli und Lucchese, 2015)	7
3.2. Dexter-Analyser für die Evaluierung der Messergebnisse anhand der Bildung von <i>Recall</i> , <i>Precision</i> und <i>F1 score</i>	8

Abkürzungsverzeichnis

CSV	Comma-separated values
GUI	Graphical user interface
HTTPS	HyperText Transfer Protocol Secure
JSON	JavaScript Object Notation
REST	Representational State Transfer
XML	Extensible Markup Language

1. Einführung

Ziel der Arbeit ist die Entwicklung der experimentellen Umgebung SecondCast zur Durchführung von Second Screen Untersuchungen. Die angestrebte Gesamtansicht vom SecondCast-System ist in Abbildung 1.1 abgebildet. Dabei besteht die Untersuchungsplattform aus den zwei Hauptkomponenten SecondCast-Client und SecondCast-Server (Farbe Orange). Client-seitig wird neben Social-Media-Diensten (Farbe Weiß) auf eine Mediathek von Dokumentationen (Farbe Gelb) und für das Abspielen der Videos auf dem First Screen der Streaming-Media-Adapter Chromecast (Farbe Grau) verwendet. Server-seitig erfolgt die Kommunikation über den SecondCast-Core (Farbe Pink). Dieser verwendet eine Kontext-Komponente (Farbe Türkis) zur Extraktion von Schlagwörtern aus Texten, welche für Informationsabfragen von Informationsdiensten (Farbe Grün) verwendet werden. Der SecondCast-Summerizer (Farbe Lila) soll die erhaltenen Texte für den Nutzer zusammenfassen.

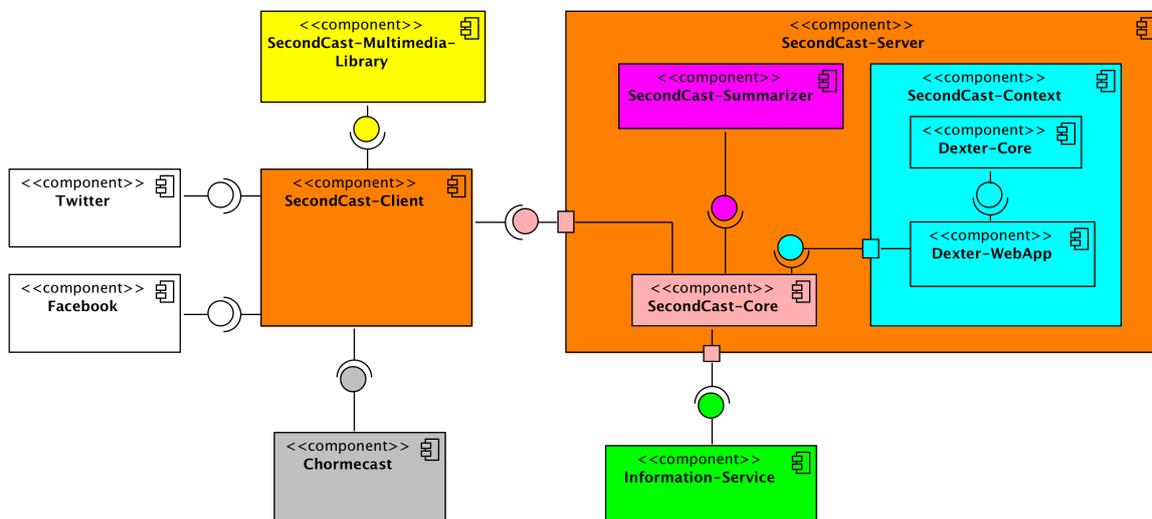


Abbildung 1.1.: Angestrebte Gesamtansicht als Komponentendiagramm von der Second Screen Untersuchungsplattform SecondCast.

Es gibt verschiedene Arten der Rollenverteilung zwischen den beteiligten Mediengeräten im Second Screen Bereich. Dabei kann der First Screen ein Tablet sein wo ein Video abgespielt wird und der Fernseher Informationen zu dem Geschehen liefern. In der Arbeit wird sich auf den meist genutzten Fall konzentriert nachdem der Second Screen ein Smartphone oder Tablet ist und der First Screen ein Fernseher darstellt (Nielsen, 2012). SecondCast fungiert dabei als Master und der Fernseher als Slave.

Die Projektarbeit soll auf Basis der Ergebnisse von Demin (2015) die Weiterentwicklung von SecondCast ermöglichen. Für die Umsetzung der Anwendung müssen Lösungen für die aufgeführten Problemstellungen erarbeitet werden. Unter anderem muss der SecondCast-Client als Android-Applikation nach dem erarbeiteten und evaluierten GUI-Konzept entwickelt werden. Dieser muss

1. Einführung

eine Anbindung an Chromecast für das Streaming der Medieninhalte aufweisen. Des Weiteren muss SecondCast-Client eine Funktionalität als Fernbedienung für die Steuerung der auf dem Fernseher abgespielten Videos ermöglichen. Für die Bereitstellung von Zusatzinformationen zu den angezeigten Dokumentationen muss eine Komponente für die Extraktion von Schlagwörtern aus den Dokumentationsbeschreibungen zur Generierung von Anfragen an Informationsdienste entwickelt werden. Um die Kommunikation zwischen SecondCast-Client und SecondCast-Server zu ermöglichen muss eine **REST**-Schnittstelle bereitgestellt werden.

2. SecondCast-Client

Die interaktive Second Screen Anwendung SecondCast-Client ¹ konnte als Android-Applikation anhand des erarbeiteten GUI-Konzeptes umgesetzt werden. Um den Probanden einen Zugriff auf die Applikation zu erleichtern und Updates der Anwendung zu bewerkstelligen, wurde SecondCast-Client in den Google Play Store geladen. In der ursprünglichen Version sollte als Medienserver die ARD-Mediathek² verwendet werden. Diese bietet aber keine HTTPS-Verbindungen an, für das Streaming von Medieninhalten mit Chromecast wird aber das Kommunikationsprotokoll gebraucht. Um die Sicherheitsrichtlinien von Google zu erfüllen musste ein eigener Medienserver mit Dokumentationen aufgesetzt werden. Im weiteren Verlauf wird auf den realisierten Hauptanwendungsfall eingegangen.

Die Abbildung 2.1 zeigt anhand von Ausschnitten aus der Applikation die User Story (COHN, 2010) für die Anzeige von Dokumentationen auf dem First Screen und die Bereitstellung der Zusatzinformationen auf dem Second Screen dar. Beim Aufruf vom SecondCast-Client wird der Nutzer mit einem Auswahlmenü begrüßt. Neben einer Infoansicht und einem eingebauten Tutorial, kann der Nutzer auf die Mediathek der ausgewählten Dokumentationen zugreifen.

Der Punkt A1 verdeutlicht das Klick-Ereignis, nachdem der Nutzer zur Liste mit den von SecondCast-Multimedia-Library angebotenen Dokumentationen gelangt. Falls ein Chromecast in der Nähe ist, leuchtet im Titelbereich rechts ein Media-Streamer-Knopf auf.

Der Übergang A2 zeigt das Aufspringen eines Dialogfensters mit der Liste aller sich in der Nähe befindenden Media-Streamer an. Nachdem der Nutzer sich in diesem Fall mit dem Chromecast-ChromeDB verbunden hat, gelangt er durch die Auswahl einer Dokumentation in die Fernbedingungsansicht vom SecondCast-Client. In dieser Ansicht hat der Benutzer die Möglichkeit neben den Steuerungselementen wie der Lautstärkeregelung das ausgewählte Video auf dem verbundenen First Screen abzuspielen.

Nach der Betätigung des Abspielknopfes gelangt der Nutzer im Übergang A3 in die Zusatzinformationsansicht. Hier werden dem Benutzer die aufbereiteten Informationen von dem SecondCast-Server präsentiert. Die Fernbedingungsansicht kann dabei jederzeit mit der Durchführung einer Swipe-Right-Geste ausgeklappt und eingeklappt werden.

Nachdem der Nutzer sich die Dokumentation zu Ende angeschaut hat, kann er durch die Betätigung des Media-Streamer-Knopfes, wie in A4 dargestellt, die Verbindung mit dem Chromecast beenden.

¹<https://play.google.com/store/apps/details?id=de.haw.ivandemin.cast>

²www.ardmediathek.de/tv

2. SecondCast-Client

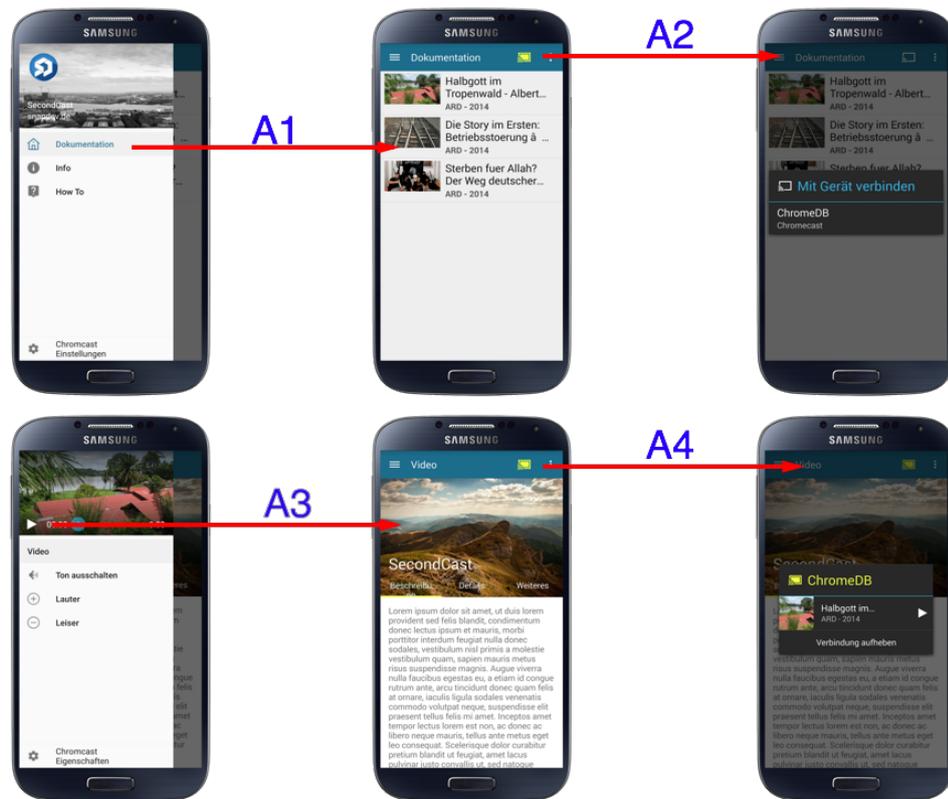


Abbildung 2.1.: Umgesetzte User Story für die Anzeige von Dokumentationen auf dem First Screen, sowie der Bereitstellung von Zusatzinformationen auf dem Second Screen.

3. SecondCast-Server

Für die Entwicklung der Kontext-Komponente wurde das Open-Source-Framework Dexter³ zur Entität-Verlinkung (Ceccarelli und Lucchese, 2013) verwendet. In diesem Kapitel wird die Arbeitsweise, sowie Verfahren welche hinter Dexter stehen, erläutert. Außerdem werden die einzelnen aufgesetzten Systemkomponenten vorgestellt. Des Weiteren wird auf die durchgeführte Evaluierung anhand des entwickelten Dexter-Analysers⁴ eingegangen.

3.1. Arbeitsweise von Dexter

Dexter arbeitet in drei Schritten: i) **Spotting**: Aus dem Dokument werden alle möglichen Kandidaten herausgesucht. ii) **Disambiguation**: Die Mehrdeutigkeit einzelner Entitäten wird aufgelöst. iii) **Ranking**: Die Liste der ausgesuchten Entitäten wird nach einem zugeordneten Wert sortiert. Im weiteren Verlauf wird auf die einzelnen Schritte näher eingegangen.

Spotting

Die Beschaffung aller möglichen Linkkandidaten basiert auf der Aufarbeitung der einzelnen Wörter aus dem Dokument. Für die Berechnung wird das Aufkommen der Entitäten im gebildeten Link-Graphen von Wikipedia gemessen. Falls zu einer Entität ein Wikipedia-Artikel existiert, wird es in die Ausgabeliste aufgenommen.

Disambiguation

Die Auflösung der Mehrdeutigkeit von Wörtern aus der Liste von erkannten Kandidaten basiert auf den Erkenntnissen wie dem Verfahren aus algorithmisch-anwendungsorientierten Perspektive zur automatischen Kreuzreferenzierung von Dokumenten mit Wikipedia-Artikeln nach Milne et al. (Milne und Witten, 2008) auf. Dieses beschreibt einen Weg mit Hilfe von maschinellen Lernens Zusatzinformationen anhand von Links zu Wikipedia-Artikeln bereitzustellen. Dazu erlernten die Autoren das Wikifizierungs-Verfahren⁵ von Wikipedia. Hierfür übernahmen die Entwickler von *Dexter* die beiden Feature: *Commonness* und *Relatedness*. Die *Commonness* beschreibt dabei die Häufigkeit, wie oft auf den Wikipedia-Artikel von Wikipedia im Vergleich zu den anderen möglichen Bedeutungen der Wörter verwiesen wird:

$$COMMONNESS(a, w) = \frac{|L_{a,w}|}{\sum_{w' \in W} |L_{a,w'}|} \quad (3.1)$$

Jedem gefundenen Linkkandidaten L wird ein Fixpunkt a zu einem bestimmten Wikipedia-Artikel w zugeordnet. Der Grundgedanke ist, dass Linkkandidaten mit Fixpunkten, welche zum selben Wikipedia-Artikel verweisen, passendere Repräsentationen sind, als Linkkandidaten die oft auf

³<https://github.com/dexter/dexter>

⁴<http://snapdev.de/chromeCast/dexter/dexter-analyser/>

⁵<http://de.wikipedia.org/wiki/Wikipedia:Wikifizieren>

andere Ziele verweisen. Dagegen beschreibt *Relatedness* die Verbundenheit der Linkkandidaten mit dem Artikel. Dazu werden die eingehenden und ausgehenden Links der beiden Artikel untersucht. Je höher der gemessene Wert ist, desto höher ist die semantische Gemeinsamkeit des Linkkandidaten zu dem untersuchten Artikel:

$$RELATEDNESS(a, b) = \frac{\log(\max(|A|, |B|)) - \log(|A \cap B|)}{\log(|W|) - \log(\min(|A|, |B|))} \quad (3.2)$$

Hier sind a und b zwei ausgesuchte Artikel, wobei A und B die Sets aller Artikel sind, die zu a und b verweisen. W ist das Set aller Artikel in Wikipedia. Die Verbundenheit der beiden Texte basiert auf dem durchschnittlichen Kontext der in Verbindung stehenden Artikel-Sets.

Ranking

Die gefundenen Entitäten werden nach verschiedenen Strategien sortiert. Unter anderem wird ein *Score* für die einzelnen Entitäten gebildet. Diese besagt, dass im Kontext U die Wikipedia-Artikel w die selbe Thematik wie andere Entitäten adressieren, wenn sie am meisten mit den eindeutigen Entitäten eine hohe Gemeinsamkeit an Verlinkungen aufweisen:

$$SCORE(c | U) = \frac{1}{\sum_u w_u} \sum_{u \in U} w_u * p^{MW}(u, c) \quad (3.3)$$

3.2. Systemaufsetzung von Dexter

Das Open-Source-Framework Dexter besteht aus mehreren Komponenten, welche in mehreren Arbeitsschritten aufgesetzt wurden. Der Gesamtausschnitt von Dexter kann der Abbildung 3.1 entnommen werden. Mit Hilfe der Komponente **JSON-Wikipedia** wurde der Wikipedia-Dump⁶ von 07.04.2015 mit rund 1,7 Millionen deutschsprachige Wikipedia-Artikeln vom **XML**-Format in **JSON**-Format umgewandelt. Danach wurde mit Hilfe von Dexter-Core der Speicherauszug in mehreren Teilschritten zu einem Entität-Link-Graphen (**Cudré-Mauroux u. a., 2009**) umgewandelt. Unter anderem wurde darunter mit Hilfe von Apache Lucene⁷ aus dem Wikipedia-Dump ein Artikel-Index generiert. Danach musste die Komponente Dexter-Webapp konfiguriert werden. Das Teilsystem besteht dabei aus einer Web-App, welche eine Swagger-Komponente⁸ für die Visualisierung der angebotenen REST-API anbietet. Sowohl die Web-App⁹ als auch die Visualisierte Form der REST-API¹⁰ wurden online aufrufbar gemacht.

⁶<https://dumps.wikimedia.org/dewiki/latest/>

⁷<https://lucene.apache.org/core/>

⁸<http://swagger.io/>

⁹<http://snapdev.de:8080/dexter-webapp/>

¹⁰<http://snapdev.de:8080/dexter-webapp/dev/>

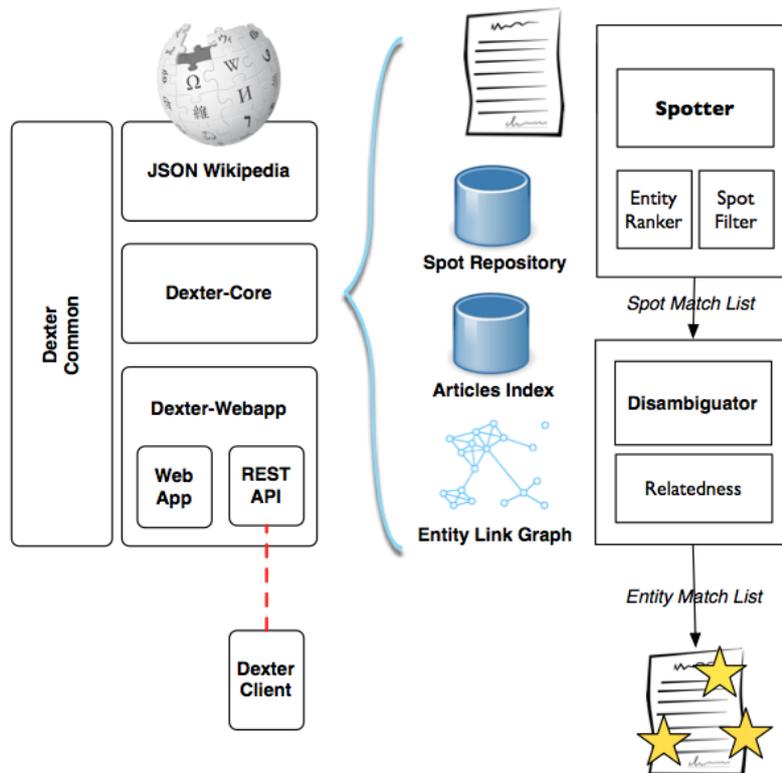


Abbildung 3.1.: Architektur von Dexter (Ceccarelli und Lucchese, 2015)

3.3. Evaluierung

Dieser Abschnitt befasst dich mit der Evaluierung vom aufgesetzten Dexter-System. Die Untersuchung der Performance von Dexter erfolgt anhand von Dexter-Analyser, welche Evaluierungsansicht veranschaulicht wird. Danach wird auf die durchgeführte Untersuchung eingegangen und eine Bewertung der Messergebnisse gemacht.

Dexter-Analyser

Um die Qualität der aufgesetzten Kontextkomponente zu messen wurde das Evaluierung-System Dexter-Analyser entwickelt. Die Abbildung 3.2 verdeutlicht den Aufbau der Evaluierungsansicht der Anwendung. Insgesamt ist die Ansicht in vier Abschnitte: F1, F2, F3 und F4 aufgeteilt. Der obere Abschnitt F1 beinhaltet die einzelnen Regler für die Filterung der ausgegebenen Datensätze. Darunter kann die minimale Linkwahrscheinlichkeit und Ergebnis angegeben werden.

Im Abschnitt F2 wird dem Nutzer die Möglichkeit geboten, einen Text für die Analyse einzugeben und an die REST-Schnittstelle des aufgesetzten Dexter-Systems zu versenden.

3. SecondCast-Server

Im nachfolgenden Teilbereich F3 kann der Anwender die Wörter eingeben, welche er für das Grundverständnis des Textes als wichtig erachtet. Anhand der eingegeben Wörter und der Ergebnisse von Dexter werden die Werte für *Recall*, *Precision* und *F1 score* (W. Berry und Kogan, 2010) gebildet.

Die konkrete Ausgabe der erhobenen Datensätze von Dexter werden im Abschnitt F4 in einer Tabelle aufgliedert. Neben der Möglichkeit die Ergebnisse im **CSV**-Format zu exportieren wird die Funktionalität der Sortierung nach den Ausgabeparametern angeboten.

The screenshot displays the Dexter-Analyser interface with four main sections:

- F1 (Parameters):** Includes sliders for 'Min-conf: 0.8', 'Min-linkprobability: 0.6', and 'Min-score: 0.6', and a 'Number: 50' input field.
- F2 (Text):** Contains a text area with a German article snippet about the 1983 political deal between East and West Germany. A 'Send' button is at the bottom right.
- F3 (Terms):** Shows a list of extracted terms: 'Bayerischer_Ministerpräsident', 'Staatsbankrott', and 'DDR'. Below the list, it displays 'Recall: 0.750', 'Precision: 0.600', and 'F-measure: 0.667'.
- F4 (Table):** Features an 'Export to CSV' button and a table with the following data:

Wikiname	Linkprobability	Commonness	Score
Alexander_Schalck-Golodkowski	1.000	1.000	1.000
Bayerischer_Ministerpräsident	1.000	1.000	1.000
Niederländische_Sprache	0.758	0.746	0.746
Staatsbankrott	1.000	1.000	1.000
Überschuldung	1.000	1.000	1.000

At the bottom right of the table, there are pagination controls for 10, 25, 50, and 100 items.

Abbildung 3.2.: Dexter-Analyser für die Evaluierung der Messergebnisse anhand der Bildung von *Recall*, *Precision* und *F1 score*.

Untersuchung

Für die Messung der Qualität von Ausgabewerten wurde eine Untersuchung durchgeführt. Als Testdaten wurden die 40 ausgesuchten Dokumentationsbeschreibungen aus dem Projektbericht (Demin, 2015) verwendet. Alle implementierten Filter wurden auf 80 % gestellt. Das Gesamtergebnis der Untersuchung kann aus der Tabelle 3.1 entnommen werden. Die einzelnen Teilergebnisse für jede Dokumentationsbeschreibung werden in der Tabelle A.1 aufgeführt.

Tabelle 3.1.: Gesamtergebnis der Performance von Dexter anhand der untersuchten Dokumentationsbeschreibungen.

<i>Recall:</i>	39,37
<i>Precision:</i>	22,09
<i>F1 score:</i>	26,15

Bewertung

Die erreichten Werte sind ausbaufähig. Dexter hat Recall-Werte von knapp unter 40 Prozent erreicht. Die Werte für Precision und F1 score sind unter 30 Prozent. Wenn die detaillierte Liste zur Bewertung von Teilergebnissen hinzugezogen wird, ist es auffällig, dass bei insgesamt zehn Dokumentationsbeschreibungen eine Nullrunde erreicht wird. Es ist bei der Untersuchung aufgefallen, dass Dexter die Wörter nicht richtig umbricht. Damit werden oft Ortslandschaften gefunden und der ganze Text in Bezug zu diesen Entitäten genommen. Es stellte sich heraus, dass der Spotter nicht richtig arbeitet, da die Komponente Dexter-Core noch einen NER (Named Entity Recognition) für die Englische Sprache verwendet. Damit werden den Entitäten falsche Klassen zugeordnet. Des Weiteren sind Antwortzeiten bei größeren Texten von über 500 Wörtern zu lang.

4. Herausforderungen

Die Arbeiten an der Untersuchungsplattform SecondCast schreiten weiter voran. Dessen Umsetzung wirft neue Herausforderungen, welche in der Master-Thesis gelöst werden sollen. Auf die weiteren Problemstellungen wird in diesem Abschnitt eingegangen.

SecondCast-Client

Die Komponente wurde so entwickelt, dass sie von einem Server Datensätze erhalten kann. Dazu muss SecondCast-Client an den SecondCast-Core angebunden werden, welcher die Zusatzinformationen zu dem Geschehen liefert. Die gelieferten Datensätze werden dann aufbereitet und in der Zusatzinformationsansicht angezeigt. Außerdem soll das Nutzererlebnis anhand einer Verfeinerung das GUI-Konzept durch Umsetzung des Nutzerfeedbacks verbessert werden. Um den Benutzern den Einstieg in die Anwendung zu vereinfachen, soll das eingebaute Tutorial erweitert werden. Des Weiteren sollen die Social-Media-Dienste Facebook und Twitter angebunden werden.

SecondCast-Context

Die Evaluierung von Dexter als SecondCast-Context zeigte nicht die gewünschten Ergebnisse. Um die Performance der Komponente zu erhöhen soll der von dem Spotter verwendete englische NER durch einem deutschen NER¹¹ ausgetauscht werden. Die Evaluierung der Performance soll anhand der rausgesuchten Dokumentationsbeschreibungen mit besonderem Augenmerk auf die Texte, welche zu einer Nullrunde geführt haben, gemessen werden.

SecondCast-Core

Aufgrund der nicht ausreichenden Performance von SecondCast-Context, sowohl bei den Antwortzeiten bei der Extraktion von Schlagwörtern von längeren Texten, als auch die Qualität der erhaltenen Datensätze muss eine zusätzliche Komponente hinzugezogen werden. SecondCast-Core soll für die Bereitstellung der Datensätze für SecondCast-Client die evaluierten Informationen in einer Datenbank vorhalten. Erst wenn keine Informationen zur Dokumentationsbeschreibung gefunden werden können, soll eine Anfrage an den SecondCast-Context gestellt werden. Des Weiteren soll die Zwischeninstanz aus den erhaltenen Entitäten von SecondCast-Context Informationsabfragen generieren um Zusatzinformationen von semistrukturierten Archiven zu erhalten. Im ersten Schritt soll eine Anbindung an die Informationsquelle Wikipedia entwickelt werden.

¹¹http://www.nlpado.de/~sebastian/software/ner_german.shtml

5. Schlussbetrachtung

Zusammenfassung

Aufbauend auf den Projektergebnissen aus [Demin \(2015\)](#) wurden weitere Infrastrukturmaßnahmen zur Entwicklung einer Experimentumgebung zur Durchführung von Second Screen Untersuchungen geschaffen. Unter anderem wurde SecondCast-Client als Android-Applikation mit Umsetzung vom erarbeiteten GUI-Konzept mit Anbindung an den Streaming-Media-Adapter Chromecast geschaffen. Die Anwendung wurde im Google Play Store platziert. Da die ARD-Mediathek nicht die Sicherheitsrichtlinien von Google für das Streamen von Videos über Chromecast aufgrund fehlender HTTPS-Funktionalität erfüllt hat, musste ein Medienserver SecondCast-Multimedia-Library mit Dokumentationen aufgesetzt werden. Server-seitig wurde für die Schlagwortextraktion aus Dokumentationsbeschreibungen die Komponente SecondCast-Context basierend auf Dexter eingerichtet. Hierfür wurde ein Entität-Link-Graph aus 1,7 Millionen deutschsprachigen Wikipedia-Artikeln generiert. Um die Performance von SecondCast-Context zu messen wurde das Evaluierungssystem Dexter-Analyser entwickelt. Mit Hilfe der Evaluierungsansicht und den rausgesuchten 40 Dokumentationsbeschreibungen aus der vorherigen Projektarbeit wurde eine Untersuchung mit der Messung von Recall, Precision und F1 score durchgeführt. Die gemessenen Werte zeigten, dass die Qualität der Komponente nicht optimal ist. Als Grund für das schlechte Abschneiden wurde der Spotter ausfindig gemacht. Da der NER noch auf Englisch ausgerichtet ist, muss er mit einem deutschen NER ausgetauscht werden. Des Weiteren wurden weitere Herausforderungen festgestellt, welche in der nachfolgenden Arbeit ausgebessert werden sollen.

Ausblick

Die Entwicklung von SecondCast schreitet weiter voran. Die Projektergebnisse ermöglichen es, einen weiteren großen Schritt in Richtung einer fertigen Experimentumgebung für Second Screen Anwendungen zu machen. Die Platzierung der Applikation im Google Play Store ermöglicht nun Nutzern an der Entwicklung von SecondCast zu profitieren und indirekt oder direkt an der Qualitätssicherung teilzunehmen. Trotz des großen Projektfortschritts muss noch einiger Aufwand in die Entwicklung von SecondCast betrieben werden. Der Bewältigung dieser Herausforderungen wird zuversichtlich entgegen geschaut.

Literaturverzeichnis

- [Ceccarelli und Lucchese 2013] CECCARELLI, Diego ; LUCCHESE, Claudio: Dexter: an open source framework for entity linking. In: *Proceedings of the sixth ...* (2013), S. 2–4. – URL <http://dl.acm.org/citation.cfm?id=2513212>. ISBN 9781450324137
- [Ceccarelli und Lucchese 2015] CECCARELLI, Diego ; LUCCHESE, Claudio: *Dexter, an Open Source Framework for Entity Linking*. 2015. – URL <http://dexter.isti.cnr.it/>. – Zugriffsdatum: 2015-04-07
- [COHN 2010] COHN, Mike: *User Stories: für die agile Software-Entwicklung mit Scrum, XP u.a. mitp Professional: Amazon.de: Mike Cohn: Bücher*. Muenchen, 2010. – ISBN 978-3-826-65898-3
- [Cudré-Mauroux u. a. 2009] CUDRÉ-MAUROUX, Philippe ; HAGHANI, Parisa ; JOST, Michael ; ABERER, Karl ; DE MEER, Hermann: idMesh. In: *Proceedings of the 18th international conference on World wide web - WWW '09*. New York, New York, USA : ACM Press, April 2009, S. 591. – URL <http://dl.acm.org/citation.cfm?id=1526709.1526789>. – ISBN 9781605584874
- [Demin 2015] DEMIN, Ivan: Projektbericht / HAW-Hamburg. Hamburg, 2015. – Forschungsbericht. – 1–11 S. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2015-proj/demin.pdf>
- [Milne und Witten 2008] MILNE, David ; WITTEN, Ian H.: Learning to link with wikipedia. In: *Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08*. New York, New York, USA : ACM Press, Oktober 2008, S. 509. – URL <http://dl.acm.org/citation.cfm?id=1458082.1458150>. – ISBN 9781595939913
- [Nielsen 2012] NIELSEN: *State of the Media - The Cross-Platform Report*. 2012
- [W. Berry und Kogan 2010] W. BERRY, Michael ; KOGAN, Jacob: *Text Mining: Applications and Theory*. New York, New York, USA : John Wiley & Sons, 2010. – 136 S. – ISBN 047068965X

A. Datensätze

Tabelle A.1.: Detaillierte Tabelle von Ergebnissen aus der Performance-Untersuchung von Dexter mit Titel, Link zum Beschreibungstext, Recall-, Precision- und F1 score-Werten.

Titel	Beschreibungstext	Recall	Precision	F1 score
Geschichte im Ersten: Der Milliardendeal - Strauß und die DDR	Beschreibungstext 1	0,75	0,6	0,667
Die Story im Ersten: Tod vor Lampedusa	Beschreibungstext 2	0,833	0,5	0,625
Im Zug von Chicago nach Kalifornien	Beschreibungstext 3	0,667	0,4	0,5
Abenteuer Diagnose	Beschreibungstext 4	0,5	0,333	0,4
Zug in die Freiheit	Beschreibungstext 5	0	0	0
Macht.Mensch.Schröder	Beschreibungstext 6	0,5	0,25	0,333
Kesslers Expedition	Beschreibungstext 7	0,667	0,153	0,25
Reportage im Ersten: Hebamme am Limit - Dauereinsatz für neues Leben	Beschreibungstext 8	0	0	0
Geschichte im Ersten: Mythos Dresden - Der lange Schatten einer Bombennacht	Beschreibungstext 9	0,667	0,182	0,286
7 Tage ... Auschwitz - Ein musikalisches Experiment	Beschreibungstext 10	0,333	0,2	0,25
Dokumentarfilm im Ersten: Night will fall - Hitchcocks Lehrfilm für die Deutschen	Beschreibungstext 11	0,667	0,4	0,5
Reportage im Ersten: Mit dem Moped durch Vietnam	Beschreibungstext 12	0	0	0
Putins Volk - Wohin wollen die Russen?	Beschreibungstext 13	0,5	0,091	0,154
Flüchtlinge - Aufnehmen oder abschieben?	Beschreibungstext 14	0	0	0
Alles anders: Die Handball-WM in Katar	Beschreibungstext 15	0,6	0,3	0,4
Die Story im Ersten - Schlachtfeld Internet Wenn das Netz zur Waffe wird	Beschreibungstext 16	0	0	0
Paris trauert - live aus der französischen Hauptstadt	Beschreibungstext 17	0,6	1	0,75
Traumstädte - Rom	Beschreibungstext 18	0,333	0,143	0,2

A. Datensätze

Die Story im Ersten: Polizei, Gewalt und Videos - Wenn Einsätze aus dem Ruder laufen	Beschreibungstext 19	0,333	0,143	0,2
Die Story im Ersten: Die geheime Macht von Google- Ein Film von Ulrich Stein	Beschreibungstext 20	0,5	0,2	0,286
Reportage im Ersten: China extrem - Mit dem Rad an Tibets Grenzen	Beschreibungstext 21	0,5	0,111	0,182
Die Story im Ersten: ChinAmerika - Duell der Supermächte	Beschreibungstext 22	0	0	0
Mission Rosetta - Auf der Suche nach dem Ursprung des Lebens	Beschreibungstext 23	0	0	0
Reportage im Ersten: Der Wilde Westen zu Fuß - Abenteuer auf dem Pacific Crest Trail	Beschreibungstext 24	0,5	0,333	0,4
Die Story im Ersten: Ranga Yogeshwar in Fukushima - Unterwegs in der größten Nuklearbaustelle der Welt	Beschreibungstext 25	0	0	0
Die Story im Ersten: Kalifat des Schreckens - Bedrohung durch den IS-Terror	Beschreibungstext 26	0,5	0,333	0,4
Der Friedensweg - Auf Spurensuche am Sentiero della Pace	Beschreibungstext 27	0,333	0,167	0,222
Die Story im Ersten: Das geplünderte Erbe - Terrorfinanzierung durch deutsche Auktionshäuser	Beschreibungstext 28	0,5	0,5	0,5
Reportage im Ersten: Einsam im Atlantik - Die Färöer Inseln	Beschreibungstext 29	0,833	0,556	0,667
Die Story im Ersten: Tod vor Lampedusa - Europas Sündenfall	Beschreibungstext 30	0,333	0,125	0,182
Die Story im Ersten: Der Mietreport - Wenn Wohnen unbezahlbar wird	Beschreibungstext 31	0,8	0,308	0,444
Die Story im Ersten: Hausbesuch vom Amtsgericht - Die neuen Siedler von Mannheim	Beschreibungstext 32	0,75	0,375	0,5

A. Datensätze

Die Story im Ersten: Der Kinderreport - Nachwuchssorgen im Wohlstandsland	Beschreibungstext 33	0,5	0,167	0,25
Sterben für Allah? Der Weg deutscher Gotteskrieger nach Syrien	Beschreibungstext 34	0	0	0
Die Story im Ersten - Mission unter falscher Flagge Radikale Christen in Deutschland	Beschreibungstext 35	0	0	0
Die Story im Ersten: Der Arbeitsmarktreport Das Märchen vom Fachkräftemangel	Beschreibungstext 36	0	0	0
Kicker, Zocker und Rebellen Die WM-Helden von '74	Beschreibungstext 37	0,667	0,182	0,286
Die Story im Ersten: Betriebsstörung - Macht die Bahn noch mobil?	Beschreibungstext 38	0,333	0,111	0,167
Europawahl 2014 Der Europa-Check	Beschreibungstext 39	0,5	0,5	0,5
Halbgott im Tropenwald - Albert Schweitzers Lambarene	Beschreibungstext 40	0,75	0,429	0,545