MARIANNE ENGLERT PREIS

newcomer-forum im vfm – Neues aus den Hochschulen

Marianne-Englert-Preis-Jury

Mitglieder der Jury:
Dr. Ute Essegern
Sächsische Zeitung Dresden, Dokumentation & Leserdialog

Frank Dürr
WDR Funkhaus Düsseldorf/Dokumentation und Archive

Vanessa Freudrich
Schweizer Radio und Fernsehen SRF, D+A Dokumentation und Archive

Michael Vielhaber
Österreichischer Rundfunk ORF Multimediales Archiv

Der Verein für Medieninformation und Mediendokumentation (vfm) zeichnete zum fünften Mal zukunftsweisende Nachwuchsarbeiten von Studierenden oder Absolventen aus den Bereichen Information, Dokumentation, Archiv und Bibliothek aus. Der Preis wurde am 25. April 2017 in Mainz während der Frühjahrstagung der Medienarchivare verliehen. In diesem Jahr gingen die mit jeweils 500 Euro dotierten Preise an ein Team von fünf Studierenden des Hasso-Plattner-Instituts Potsdam, an Catharina Boss (Technische Hochschule Köln/infoNetwork GmbH) sowie an Dr. Julia Lorke (Imperial College London).

In Vertretung der Gesamtgruppe präsentierten Moritz Finke und Julian Risch unter dem Titel „match me if you can“ die semantische Aufbereitung von Fußball-Daten. Die jungen Medienwissenschafter zeigten, dass Fakten aus 60 Jahre Fußballgeschichte mit über 500 Mannschaften und 40.000 Spielern der Champions League sowie der 1. und 2. Bundesliga, auch mit einfachen Mitteln innovativ verknüpft und visualisiert werden können.

Michael Vielhaber, Dr. Julia Lorke, Moritz Finke, Julian Risch, Catharina Boss und Mario Müller

Catharina Boss wurde für eine Arbeit ausgezeichnet, die sich mit kuratierten Twitterlisten beschäftigt. Social Media Content gewinnt für die mediale Berichterstattung zunehmend an Bedeutung – und stellt Journalisten wie Mediendokumentare vor die Herausforderung, relevante Inhalte aus vertrauenswürdigen Quellen zügig auffindbar zu machen. Um Struktur in die Datenflut sozialer Netzwerke zu bringen, gibt es eine Vielzahl an Werkzeugen und Methoden, die Boss verglichen und bewertet hat.

Dr. Julia Lorke beschäftigte sich ebenfalls mit einem Thema aus dem Bereich Social Media. Sie untersuchte das Potential sozialer Netzwerke zur Interaktion zwischen Radiomachern und Hörern. Ob Radioprogramme Facebook, Twitter & Co bereits für einen echten Austausch zwischen Produzenten und Publikum nutzen, oder lediglich als kostengünstiges Marketinginstrument hat Lorke am Beispiel ausgewählter Wissenschaftssendungen geprüft.
Die bisherigen Preisträger finden sich hier: www.vfm-online.de/weblog/newcomerforum/preistraeger/

Wir möchten noch intensiver aktuelle Forschungen veröffentlichen, die sich mit Fragestellungen zur Informationsgesellschaft befassen und ihren Blick insbesondere auf mediendokumentarische oder kommunikationswissenschaftlich-technische Themen lenken. Bitte geben Sie dies an Ihre Fachbereiche weiter, an Graduierende und Absolventen. Weitere Auskünfte erteilt das Redaktionskollegium der info 7 (Kontakt: redaktion@info7.de).

Gleichzeitig möchten wir bereits jetzt dazu ermuntern, Abschlussarbeiten für den Marianne-Englert-Preis 2018 einzureichen. Der Bewerbungsschluss endet am 31. Januar 2018. (Kontakt: newcomer@vfm-online.de). •

„Match Me If You Can“ – Sammeln und semantisches Aufbereiten von Fußballdaten*

Moritz Finke und Julian Risch

ABSTRACT

Interviews, Spielstatistiken oder Videoaufzeichnungen sind für Fußballfans zwar zahlreich im Internet verfügbar, aber auf viele verschiedene Websites verstreut. „Semantic Media Mining“ verknüpft nun Fußballdaten aus unterschiedlichen Quellen, bereitet sie semantisch auf und führt sie auf einer einzigen Website zusammen. Dadurch dokumentieren und visualisieren wir mehr als 50 Jahre Fußballgeschichte mit über 500 Mannschaften und 40.000 Spielern der Champions League, sowie der 1. und 2. Bundesliga.

FUSSBALLDATEN IM INTERNET

Zu einem populären Themenbereich wie Fußball steht eine Vielzahl verschiedenartiger Datenquellen zur Verfügung. Außer den offiziellen Websites der Ligen und Vereine existieren Internetauftritte von Sportmagazinen wie Kicker mit redaktionell aufbereiteten, aktuellen Inhalten. Twitter und YouTube aus dem Social Web oder Wikipedia als Enzyklopädie bieten nutzergenerierte Daten als weitere Möglichkeit sich zu informieren.

Semantisch verknüpfte Informationen im Internet (Semantic Web) bieten aufgrund des großen verfügbaren Datenumfangs viel Potential für Anwendungen unterschiedlichster Art. Alle oben genannten Plattformen stellen ihre Daten jedoch nur semi- oder unstrukturiert bereit und beleuchten dabei lediglich einzelne Aspekte. Texte, Statistiken, Bilder und Videos sind zudem auf viele verschiedene Websites verstreut. Dadurch muss sich ein Nutzer, seinen Bedürfnissen entsprechend, auf mehreren Seiten einen Überblick verschaffen.

Interessante Statistiken ergeben sich häufig aber erst aus der Kombination der Daten. Zusammenhänge, die sich auf einen großen Zeitraum oder auf Daten aus mehreren Quellen beziehen, bleiben dem Nutzer größtenteils verborgen. Die Frage nach einem Video vom Fußballspiel mit den meisten vergebenen Karten einer Saison oder die Frage nach der Mannschaft, welche am meisten vom Wechsel von der 2- auf die 3-Punkte-Regel profitiert hat, lässt sich beispielsweise nur sehr aufwendig und über Umwege beantworten.

Unsere Arbeit beschäftigt sich mit der Lösung des beschriebenen Problems, indem sie folgenden Ansatz mit Hilfe von Konzepten des Semantic Web verfolgt: Für jede einzelne Quelle werden die Daten aus ihrem bisherigen Kontext extrahiert. Anschließend werden die Daten vereinheitlicht und in einem Gesamtdatenbestand verknüpft. Dieser Datenbestand wird für vielfältige analytische Anfragen und Visualisierungen aufbereitet.

Als Voraussetzung für die Zusammenführung der Daten ergibt sich das Verknüpfen der einzelnen Informationen mit einer umfangreichen und gut strukturierten Datenquelle aus der Linked Open Data Cloud, wie beispielsweise DBpedia oder Freebase. Basierend auf dieser strukturierten Datenquelle können die gesammelten Daten in neuen Zusammenhängen interpretiert und übersichtlich dargestellt werden. Zusätzlich besteht die Möglichkeit den Datenbestand mit Informationen anzureichern, die nicht in einem direkten Zusammenhang zur betrachteten Fußball-Thematik stehen. Beispielswiese lassen historische Aufzeichnungen des Deutschen Wetterdienstes Rückschlüsse auf das Wetter und insbesondere den Zustand des Fußballfeldes bei einer bestimmten Begegnung zu.

DATENMODELL

Standardisierte Informationsbeschreibungsmittel wie das Ressource Description Framework (RDF) und die Web Ontology Language (OWL) bilden eine der technischen Grundlagen des Semantic Web. RDF ist ein vom Wolrd Wide Web Consortium (W3C) herausgegebener Standard zur Beschreibung von Metadaten im Internet[1]. Dabei ordnet sich dieses Framework in

Autoren:
Moritz Exner, Moritz Finke, Julian Risch, Timo Wagner, Tim Zimmermann
Hasso-Plattner-Institut an der Universität Potsdam/IT-Systems Engineering
Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam
+49 331 5509 272
julian.risch@hpi.de

*Im Rahmen eines Seminars „Semantic Media Mining“ haben wir, ein fünfköpfiges Studententeam, uns mit dem Sammeln und Aufbereiten von vielfältigen Fußballdaten beschäftigt. Drei von uns setzen derzeit ihr Studium am Hasso-Plattner-Institut fort, die anderen beiden haben nach ihrem Bachelorabschluss bereits angefangen zu arbeiten. Die Extraktion und das semantische Verknüpfen von Informationen aus heterogenen Quellen ist ein fortdauerndes Forschungsgebiet unter anderem am Lehrstuhl „Internet-Technologien und -Systeme“ von HPI-Direktor Prof. Dr. Christoph Meinel.

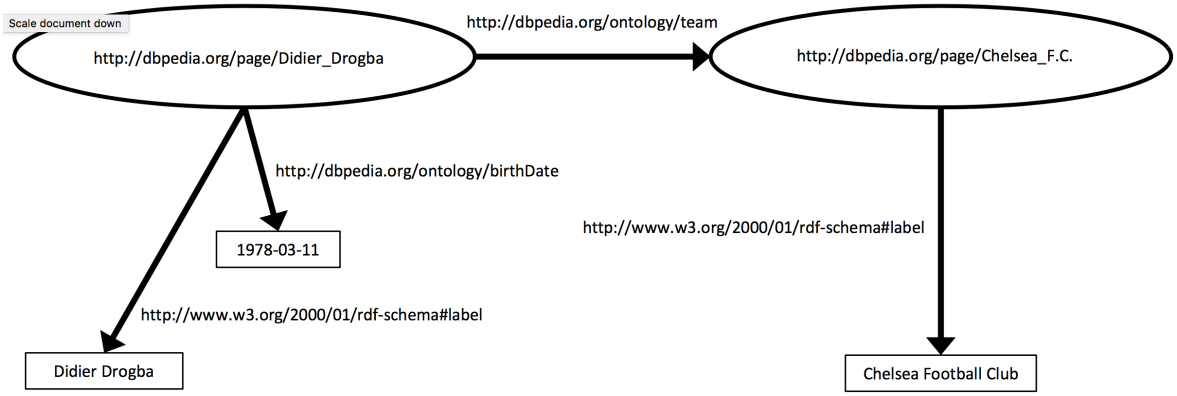

Abbildung 1: RDF Graph mit beispielhaften Fußballdaten

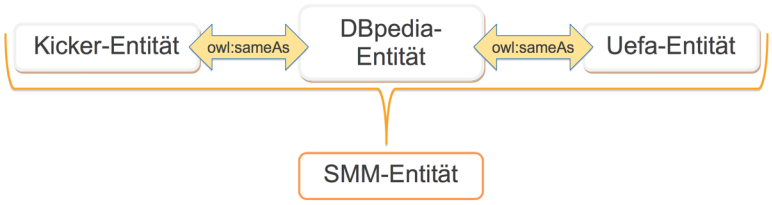

Abbildung 2: Vereinfachte Darstellung der owl:sameAs

das Semantic Web ein, indem es eine Möglichkeit bietet, verteilte, dezentrale Daten zu strukturieren und so zum Beispiel für Web-Anwendungen nutzbar zu machen. Hierbei werden Aussagen in Tripeln der Form <Subjekt, Prädikat, Objekt> getroffen, um beliebige Entitäten als sogenannte Ressourcen zu beschreiben. Das Objekt ist ein einfacher Datenwert (Literal) oder eine Ressource, welche selbst wieder als Subjekt in Erscheinung treten kann. So entstehen Strukturen, die als Graph interpretierbar sind und ein zusammenhängendes System beschreiben (siehe Abbildung 1). Es werden Uniform Resource Identifiers (URIs) nach bestimmten Konventionen verwendet, um eine eindeutige Identifizierung der Ressourcen zu gewährleisten. Ein RDF Schema (RDFS) verwendet RDF zur klassenbasierten Beschreibung eines bestimmten Datenraums. In diesem Zusammenhang wird ein Vokabular für das Beschreiben von Klassen mit Hilfe von RDF definiert.

Die Probleme, die bei der Zusammenführung von heterogenen Daten entstehen, sind detailliert von Bleiholder und Naumann beschrieben (Bleiholder, Jens; Naumann, Felix: Data fusion. In: ACM Comput. Surv. 41 (2009), Januar, Nr. 1, S. 1:1–1:41. – ISSN 0360–0300) und bilden die Grundlage für das Verständnis dieser Arbeit. Zu unterscheiden sind vor allem zwei Probleme: (i) Datenkonflikte, wobei zwei Quellen für dasselbe Attribut eines realen Objektes unterschiedliche Werte vorsehen (Fusion) und (ii) das Problem, festzustellen welche Datensätze unterschiedlicher Quellen dasselbe reale Objekt beschreiben (Matching).

DATENVERWALTUNG

Ein wesentlicher Bestandteil der Arbeit ist das Sammeln und Verknüpfen von Daten aus heterogenen Datenquellen. Dies setzt eine Strategie zur Datenverwaltung voraus. Unsere Arbeit verfolt einen RDF-basierten Ansatz, um die verwendeten Fußballdaten semantisch aufzubereiten und zu verwalten. Wir verwenden bei unserem Projekt das Datenbankmanagementsystem Virtuoso, weil es auf unsere RDF-basierte Anwendung zugeschnitten ist. Es bietet mit einem Quad-Store die Möglichkeit, RDF- Graphen zu benennen und unterstützt mit SPARQL eine mächtige Anfragesprache. Darüber hinaus wartet es aufgrund einer besseren Skalierbarkeit mit Vorteilen im Vergleich zu ähnlichen DBMS-Lösungen auf. Besonders bei großen Datenmengen lassen sich Geschwindigkeitsvorteile feststellen[2].

DATENBASIS

Die Datenbasis der Arbeit ist der englischsprachigen DBpedia[3] entnommen. DBpedia ist ein communitybasiertes Projekt mit dem Ziel, strukturierte Informationen aus Wikipedia zu extrahieren, um sie in einer Virtuoso-Datenbank zur Verfügung zu stellen. Diese Quelle eignet sich für unser Projekt besonders gut, da sie für in Europa aktive Fußballspieler und Vereine Informationen in bereits semantisch aufbereiteter Form anbietet. Daten aus anderen Quellen bilden wir auf die DBpedia als Basis ab. Bei diesem Prozess, der Teil unseres sogenannten Matching-Verfahrens ist, wird einer DBpedia-Entität mittels owl:sameAs-Beziehung eine Entität aus anderen Datenquellen (beispielsweise Kicker Magazin oder UEFA) zugeordnet und umgekehrt (siehe Abbildung 2).

Werden auf diese Weise mehrere Zuordnungen zur selben Entität vorgenommen, so erfolgt die Zuordnung der beteiligten Entitäten durch Bildung der transitiven Hülle. Diese Transitivität wird mit der Same-As-Traversal-Funktion in Virtuoso durch Voranstellen von 'DEFINE input:same-as' in einer SPARQL Anfrage realisiert.

[1] http://www.rdfabout.com/intro/?section=2

[2] http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V6/#comparison

[3] http://dbpedia.org/About

Link	D	EN	ES	IT	EM	WM	F_ D	F_ EM	F_ WM	CL	DFB
de.uefa.com/	-	-	-	-	x	-	-	x	-	x	-
www.kicker.de	x	x	x	x	x	x	x	x	x	x	x
www.youtube.com/user/ SkySportHD	x	-	-	-	-	-	-	-	-	x	x
www.footytube.com/	x	x	x	x	-	-	-	-	-	x	x
de.soccerway.com/	x	x	x	x	x	x	x	x	x	x	x
pesstatsdatabase.com	x	x	x	x	-	-	-	-	-	-	-
twitter.com/Kicker_live	x	-	-	-	x	x	-	-	-	x	x
www.fifa.com/	-	-	-	-	-	x	-	-	x	-	-
www.goalsarena.org/	x	x	x	x	x	x	-	-	-	x	x
www.openligadb.de/	x	x	x	x	x	x	-	-	x	x	x
www.soccerstats.com/	x	x	x	x	x	x	-	-	-	x	x
www.worldfootball.net/	x	x	x	x	x	x	x	x	x	x	x

Link	API	Lizenz	Umfang
de.uefa.com/	-	private Nutzung erlaubt	CL komplett, national nur aktuelle Saison
www.kicker.de/	-	private Nutzung erlaubt	D komplett, andere ab Ende 1990er-Jahre
www.youtube.com/user/ SkySportHD	x	Videoeinbindung erlaubt	Videos zur aktuellen Saison
www.footytube.com/	-	Videoeinbindung erlaubt	teilweise auch Videos vergangener Saisons
de.soccerway.com/	-	private Nutzung erlaubt	BL komplett, CL ab 2000
pesstatsdatabase.com	-	frei	nur Spielerdaten, keine Ergebnisse etc.
twitter.com/Kicker_live	*	frei	Live-Ergebnisse
www.fifa.com/	-	private Nutzung erlaubt	komplett
www.goalsarena.org/	-	private Nutzung erlaubt	ab 1999
www.openligadb.de/	x	frei	ab 2008, aber unvollständig
www.soccerstats.com/	-	private Nutzung erlaubt	ab Mitte 2000
www.worldfootball.net/	-	private Nutzung erlaubt	Frauenfußball ab Mitte der 1990er-Jahre

Tabelle 1: Verschiedene Datenquellen und der Umfang der zur Verfügung stehenden Daten. Beziehung.

Neben DBpedia verwenden wir auch andere Quellen. Eine Übersicht der ursprünglichen Kandidaten, von denen die ersten vier ausgewählt wurden, ist in Tabelle 1 dargestellt.

Die Quellen werden mit niedriger Bandbreite und Pausen zwischen den einzelnen Seitenaufrufen indiziert, um die Last auf der Quelle möglichst gering zu halten. So erhaltene Daten werden in das RDF-Format transformiert, indem HTML-Inhalte extrahiert und mit zusätzlichen Informationen aus anderen Dateien dieser Quelle angereichert werden. So können beispielsweise Fußballspieler mit passenden Informationen wie vollem Namen, Geburtsdatum und Bildern erweitert werden.

Da die Struktur der Daten für verschiedene Zeiträume unterschiedlich sein kann, werden zur Inhaltsextraktion möglichst allgemeine reguläre Ausdrücke verwendet. Ebenso kann sich die Zeichenkodierung der Seiten ändern, was die Umwandlung in einen einzigen Zeichensatz notwendig macht.

Zusätzlich zur Zuordnung offensichtlicher Fakten bezüglich einer Entität, können auch erste semantische Verknüpfungen in der Transformation stattfinden. Beispielsweise war bei einem Spiel des FC Chelsea am 19.05.2012 Didier Drogba in der Startaufstellung. Diese Information wird bei unserem Projekt unter anderem auf ein Tripel mit <Didier Drogba, club-2012, FC Chelsea> gebildet. Von der Startaufstellung einer Mannschaft zu einem bestimmten Datum extrahiert unser Ansatz also auch die Zugehörigkeit der einzelnen Spieler zur Mannschaft. Dies erleichtert später die Berechnung einer Spielerlaufbahn.

Des Weiteren können Überschriften oder wiederkehrende Passagen, beispielsweise Stimmen zum Spiel, aus Spielberichten extrahiert und gespeichert werden. Insbesondere Überschriften fassen meist spielentscheidende Aktionen oder ganze Spiele kurz und prägnant zusammen.

Eine essentielle Aufgabe für die Zusammenführung und semantische Aufbereitung der gesammelten Daten ist das sogenannte Entity-Matching. Diese Herausforderung lässt sich als formales Problem spezifizieren. Gegeben seien zwei Datenquellen S_a und S_b und die zugehörigen Entitätsmengen $A \in S_a$ sowie $B \in S_b$. Dann sollen die Entitäten aus A und B, welche jeweils dasselbe reale Objekt beschreiben, einander zugeordnet werden. Das Matching setzt alle Entitäten in eine Äquivalenzrelation. Die damit verbundenen Transitivitäts- und Symmetrieeigenschaften sind Voraussetzungen dafür, dass das Problem auf das Abbilden in einen zentralen Basisdatensatz zurückgeführt werden kann.

Da die von uns verwendeten Quellen im Wesentlichen duplikatfrei sind, ist es nicht nötig, Entity-Matching innerhalb einer Datenquelle zu betreiben. Wir gehen davon aus, dass zwei beliebige Entitäten innerhalb einer Quelle stets unterschiedliche reale Objekte beschreiben. Die Schwierigkeit besteht vielmehr darin, Entitäten aus heterogenen Datenquellen eindeutig und vor allem korrekt im Sinne der oben definierten Vorgaben einander zuzuordnen.

Für eine eindeutige Zuordnung mit Hilfe der vorliegenden Attribute (beispielsweise Spielername) spielen unter Umständen Kriterien zur Einschränkung des Ergebnisraums eine Rolle. Für den Fall, dass kein Geburtsdatum eingetragen ist, kann kein eindeutiges Matching garantiert werden. Beispielsweise existiert in unserem Basisdatensatz ein Eintrag zu einem Spieler, der den vollen Namen des Spielers und sein Geburtsdatum enthält. Wenn bei einem Fußballspiel ein Spieler mit diesem Namen erwähnt wird so kann zusätzlich zur Namensübereinstimmung überprüft werden, ob der Spieler zum Zeitpunkt des gegebenen Spiels in einem spielfähigen Alter war oder ob er zu diesem Zeitpunkt noch gar nicht geboren war.

Die Qualität der zusammenzuführenden Datenquellen ist ein wesentlicher Faktor für ein erfolgreiches Matching. Speziell die Hierarchie und Beschaffenheit der DBpedia-Datensätze sind im Zusammenhang dieser Arbeit interessant. Im Folgenden wird dargestellt, wie diese Eigenschaften eingesetzt werden, um typischen Schwierigkeiten beim Matching im Anwendungsbereich Fußball entgegenzuwirken.

Das Kernproblem stellen unterschiedliche Bezeichnungen für die gleiche, reale Entität dar. So ist zum Beispiel in manchen Datenquellen von ”FC Chelsea“ und in anderen von ”The Blues“ die Rede, wenn der Champions League-Sieger der Saison 2011/12 gemeint ist. Abweichungen in der Benennung können aber nicht nur durch Spitznamen, sondern auch durch unterschiedliche Schreibweisen entstehen.

In der Ontologie von DBpedia existieren viele unterschiedliche Entitätstypen, die einem Objekt einen oder mehrere Namen zuordnen. Um Namenskonflikte aufzulösen, verwenden wir Attribute wie ”rdfs:label“, ”dbprop:name“ und ”dbo:wikiPageRedirects“, wobei jeder dieser Entitätstypen gleichberechtigt in das Matching einbezogen wird. Damit findet im ersten Schritt eine Art Ontology-Matching zwischen den namensgebenden Attributen statt.

Darüber hinaus gibt es in der DBpedia unterschiedliche Möglichkeiten, Entitäten in einen speziellen Kontext einzuordnen. In unserer Arbeit optimieren wir das Matching mit Rücksicht auf die zuvor genannten Ansätze. Dazu werden die namensgebenden Attribute aus der DBpedia mit einem regulären Ausdruck abgeglichen, der auf dem Namen des jeweiligen Objekts basiert. Wird beispielsweise der FC Chelsea in den ”dbo:wikiPageRedirects“ auch als ”The Blues“ geführt, so kann die Mannschaft auch über diesen Namen zugeordnet werden. Zusätzlich wird der betrachtete Entitätsbereich beispielsweise durch Listen oder Kategorien wie ”dbo:SoccerClub“ und ”dbo:SoccerPlayer“ eingeschränkt. Der anschließend verwendete reguläre Ausdruck darf zu Gunsten einer erhöhten Matching-Rate durchlässiger werden, ohne dass dabei die Gefahr von falsch oder mehrfach zugeordneten Entitäten erhöht wird.

MATCHING

Beispiel des Mannschafts-Matching

Für das Team-Matching werden zuerst sämtliche Nicht-ASCII-Zeichen innerhalb des jeweiligen Mannschaftsnamens durch einen Punkt ersetzt. Dadurch entsteht ein regulärer Ausdruck, der für Anfragen an DBpedia verwendet wird. Diese Anfrage besteht neben dem Abgleich der Namen auch aus einer Spezifizierung des Entitätstyps, nämlich dbo:SoccerClub, und der Überprüfung, ob es sich auch wirklich um einen Fußballverein handelt. Dies ist notwendig, da neben Fußballmannschaften fälschlicherweise auch viele andere Sportvereine in dieser Kategorie in der DBpedia vertreten sind. Wir überprüfen, welcher Liga ein Verein zugeordnet ist. Konkret muss die Liga ”Template:Infobox football league“ verwendet werden.

Zusätzlich können hier auch noch spezielle Textkürzel implizit ausgeschlossen werden. Sinnvoll ist dies vor allem zum Ausschluss von Reserve- oder falsch kategorisierten Mannschaften einer anderen Sportart.

Findet diese Anfrage genau einen passenden Verein, so wird dieser in die Datenbank geladen und mittels sameAs mit dem angefragten Objekt beidseitig verknüpft.

Bei mehreren Treffern werden Teilwörter mit drei oder weniger Buchstaben eliminiert und durch .* beliebige Wörter zwischen den resultierenden Teilwörtern erlaubt. Dies hat den Grund, dass die verwendeten Mannschaftskürzel wie FC oder VfB an unterschiedlichen Plätzen innerhalb des Namens stehen können.

Bei keiner oder mehreren gefundenen Mannschaften wird eine zusätzliche Anfrage gestellt. Diese ist weitestgehend analog zur ersten Anfrage, allerdings wird auf eine Überprüfung des verwendeten Ligentemplates verzichtet. Gibt sie mehrere Teamnamen zurück, so wird der kürzeste Name als Treffer betrachtet. Dieser Treffer wird dann geladen und mit sameAs bidirektional verknüpft. Die Betrachtung von Teams mit dem kürzesten Namen hat sich als notwendig erwiesen, da es einige Vereine gibt, die den selben Namen plus eine Staats- oder Stadtbe-

Abbildung 3: Screenshot der entstandenen Website nach der Zusammenführung verschiedener Datenquellen.

schreibung haben. Ein Beispiel hierfür ist der ”Chelsea F.C.“ . Neben diesem Verein kann ebenfalls ”Berekum Chelsea F.C.“ gefunden werden. Dies hat vor allem bei der Champions League Bedeutung, wohingegen bei den deutschen Ligen über Kategorien wie ”Category:Football clubs in Germany “ oder ”GermanFootballCluBs“ in YAGO der Wertebereich eingeschränkt werden kann.

In selten Fällen, in denen für eine Mannschaft kein Matching durchgeführt werden kann, findet eine manuelle Zuordnung statt. Dies ist allerdings nur der Fall, wenn der Teamname stark von den entsprechenden Namen in der DBpedia abweicht. Betroffen sind unter anderem Mannschaften, die in den letzten Jahrzehnten neugegründet oder mit anderen Vereinen zusammengeschlossen wurden. Ein Beispiel hierfür ist der Club ”Vorwärts Leipzig“, der später ”Vorwärts Berlin“, bis 2012 ”Frankfurter FC Victoria“ hieß und sich jetzt mit dem ”MSV Eintracht Frankfurt“ zum ”1. FC Frankfurt E/V“ zusammengeschlossen hat.

Beispiel des Spieler-Matching

Als Beispiel für das Spieler-Matching wird ein Spieler namens Sergij Dkhtjar mit dem Ge- burtsdatum 26.08.1975 verwendet. Die Anfrage nach Namen der Fußballspieler mit diesem Geburtsdatum ergibt folgendes Ergebnis (Gruppiert in Entitäten):

- ’Âîñêàíÿí, Àðàì Ãàéêîâè÷’, ’Aram Voskanyan’
- ’Akide’, ’Mercy Akide-Udoh’, ’Mercy Akide’, ’Mercy Akide Udoh’
- ’Momar Njie’
- ’Sergei Dichtjar’, ’Sergej Dichtiar’, ’Sergei Dikhtyar’, ’Sergej Dikhtjar’, ’Sergey Dikhtyar’, ’Serhij Dychtjar’, ’Serhij Dychtjar’, ’Serhij Dichtiar’, ’Serhij Dychtjar’, ’Serhiy Dikhtiar’
- ’Timur Yanyali’

Anschließend wird durch Substitution der Buchstaben ”aeiouyjncsz “ mit einer beliebigen Zeichenkette (.*) ein regulärer Ausdruck erstellt. Zusätzlich werden beliebige Wörter vor, zwischen und nach den Namensteilen erlaubt. Da dieser Spielername keine Mittelnamen oder doppelten Buchstaben enthält, müssen keine weiteren Vorkehrungen getroffen werden. Das Resultat sieht also folgendermaßen aus: Sergij Dkhtjar -> .*S.*rg.*D.*kht.*r.* Nun wird dieser reguläre Ausdruck auf die Treffer der Anfrage an DBpedia angewendet und diese nach ihrer Levenshtein-Distanz zum ursprünglichen Namen sortiert:

- ’Sergej Dikhtjar’ (Levenshtein-Distanz 0)
- ’Sergei Dikhtyar’ (Levenshtein-Distanz 2)
- ’Sergey Dikhtyar’ (Levenshtein-Distanz 2)

Das gesuchte Objekt wird auf die Entität des Eintrags mit der kleinsten Distanz abgebildet.

WEBSITE

Der Fokus der Webseite liegt auf der Verknüpfung der Daten mit multimedialen Inhalten wie Videos und Diagrammen. Auf den Seiten zu einem Spiel werden beispielsweise Informationen über vorherige Begegnungen der beiden Mannschaften angezeigt und ausgewertet, sodass sich Nutzer leicht einen Überblick verschaffen können. Die Website ist beispielhaft in Abbildung 3 dargestellt. Darüber hinaus wird das jeweilige Spiel in seinen zeitlichen Zusammenhang eingeordnet, um die zum Spieltag aktuelle Tabelle und andere Partien anzuzeigen. Zusätzlich bilden multimediale Inhalte, wie Videos, einen wesentlichen Bestandteil der dargestellten Inhalte. Die Kombination dieser Elemente fehlt auf bestehenden Seiten und ist ausschließlich in unserem System umgesetzt. Bestehende Seiten beschränken sich meist

auf Teilaspekte, sodass Nutzer auf Informationen mehrerer Seiten zurückgreifen müssen, um sich einen Gesamtüberblick zu verschaffen. Unsere Arbeit informiert auf einer einzigen Seite und möglichst facettenreich über die betrachteten Entitäten, um neue, individuelle Perspektiven auf die Daten zu eröffnen. Darüber hinaus besteht für einen Nutzer die Möglichkeit, in Diagrammen die für ihn persönlich interessanten Statistiken einzublenden. Auf den Seiten zu einer kompletten Saison ist dies in besonderem Maße möglich. Hier dienen die Tabellen von jedem Spieltag als Datenquelle für ein Diagramm. Besonders zeitliche Veränderungen über den Saisonverlauf lassen sich so graphisch beobachten. Auf existierenden Fußball-Websites fehlt dieses Maß an Individualität und Benutzerfreundlichkeit oder ist weniger ausgeprägt.

Beispielhafte Ergebnisse

Durch die historische Vollständigkeit der Daten können wir Statistiken abrufen, die ansonsten nur sehr schwer zu generieren sind. So lässt sich beispielsweise ermitteln, dass in der Zeit zwischen der ersten Bundesliga-Saison 1963/64 und der Saison 2011/12 Matthias Scherz die meisten Tore nach der Einwechslung erzielte (insgesamt 19), Jürgen Kreyer mit 6,23 Karten pro Spiel der strengste Schiedsrichter war und Borussia Dortmund in der Saison 1995/1996 die meisten Tore in der ersten Halbzeit erzielte (insgesamt 32). Die besondere Schwierigkeit liegt darin, dass die Bezeichnungen sehr verschieden sein können. Unsere Software muss zum Beispiel erkennen, dass Christiano Ronaldo häufig mit CR7 abgekürzt wird und sich die Informationen auf ein und dieselbe Person beziehen.

Fritz-Walter-Wetter

Fritz Walter war ein deutscher Fußballspieler, der unter anderem bei der WM 1954 als Kapitän der Nationalmannschaft antrat. Er bevorzugte regnerisches Wetter bei Fußballspielen und war insbesondere auf nassem Boden seinen Gegnern überlegen. Fritz-Walter-Wetter bezeichnet demzufolge regnerisches Wetter in Anlehnung an das Finale der WM 1954 . Unser System kann durch die Einbindung von Wetterdaten des Deutschen Wetterdienstes auch folgende Frage beantworten: Welches war das torreichste Spiel, bei dem es geregnet hat? Durch eine weitere Verknüpfung der Datenquellen liefert unser System darüberhinaus gleich ein Video des Spiels.

ZUSAMMENFASSUNG UND AUSBLICK

Wir zeigen, wie viele unterschiedliche mediale Inhalte zusammengeführt und auf einer Website zusammen mit passenden Tweets und Wetterdaten kombiniert zur Verfügung gestellt werden können. Der Schwerpunkt liegt darauf, Informationen aus verschiedenen Quellen über ein und dasselbe Objekt semantisch aufzubereiten und zu verknüpfen (Matching). Ein Objekt kann dabei beispielsweise eine Fußballmannschaft oder ein Fußballspieler sein. Unser Ansatz vereinheitlicht unterschiedliche Bezeichnungen für dasselbe Objekt durch die Betrachtung von Name und Geburtsdatum bei Fußballspielern als Schlüsselattribute. Wir stellen Verfahren vor, die diesen Ansatz unter Verwendung der Levenshtein-Distanz und Algorithmen zur Normalisierung der Objektbezeichner realisieren. Damit lassen sich die verteilten Informationen zusammenführen, als Gesamtheit analysieren und übersichtlich präsentieren, um interessante und detaillierte Aussagen über die letzten 50 Jahre Fußballgeschichte treffen zu können. Unser Datenbestand umfasst 575 Mannschaften, 21.000 Spiele und 40.000 Spieler aus Champions League, sowie 1. und 2. Bundesliga. Diese Daten sind durch insgesamt mehr als 190.000 Entitäten und 3,5 Millionen Tripel repräsentiert.

In Zukunft könnten Geodaten (zusätzlich zu den von uns betrachteten Daten) im Zusammenhang mit Mannschaften und Spielern erfasst und analysiert werden. Anhand von Geodaten kann ermittelt werden, in welchen Städten oder Ländern bestimmte Spieler oder Teams besonders erfolgreich sind. Der gleiche Ansatz ließe sich auch mit den Wetterdaten verfolgen. Welche Mannschaften schneiden bei Regen am besten ab oder welcher Spieler ist auch bei niedrigen Temperaturen in Topform? Neben zusätzlichen Inhalten ließen sich durch die Berücksichtigung

weiterer Wettbewerbe, wie der Frauen-Bundesliga oder Weltmeisterschaften, nationale oder geschlechtsspezifische Unterschiede auswerten. Besonders beim Frauenfußball besteht die Möglichkeit, viele Statistiken erstmalig zu erstellen, da der Informationsumfang dort bisher vergleichsweise gering ist.

Die DBpedia-Anfragen könnten mit Hilfe von YAGO oder weiteren Kategorien noch um Spieler ergänzt werden, die bisher durch falsche Kategorisierung innerhalb der DBpedia nicht berücksichtigt wurden. Lokalisierte Varianten der DBpedia könnten darüber hinaus zur Gewinnung eines breiteren Spektrums historischer Daten verwendet werden.

Hinsichtlich der Website könnte der Aspekt von Live-Informationen stärker berücksichtigt werden. Hierfür ließen sich beispielsweise für bereits angekündigte Spiele die entsprechenden Entitäten anlegen, um diese dann mit Echtzeitdaten aus Twitter oder Live-Tickern zu füllen. Zusätzlich könnten Ansätze zur Analyse der Konnotation von Tweets benutzt und verbessert werden, sodass es zum Beispiel ermöglicht wird, einem Spieler, Team oder Spiel Beliebtheitswerte zuzuordnen. Abschließend stellen wir fest, dass mit unserer Arbeit zum Thema Fußball eine solide Basis im Hinblick auf die genannten Erweiterungsmöglichkeiten geschaffen wurde, die wichtige Konzepte des Semantic Web aufgreift und dieses mit einem weiteren Beitrag bereichert. Leider erlauben es viele der von uns genutzten Quellen aus urheberrechtlichen Gründen nicht, dass die so gewonnenen Informationen auch veröffentlicht werden. •

LITERATUR

[BN09] Bleiholder, Jens ; Naumann, Felix: Data fusion. In: ACM Comput. Surv. 41 (2009), Januar, Nr. 1, S. 1:1–1:41. – ISSN 0360–0300

[BLHL01] Berners-Lee, Tim ; Hendler, James ; Lassila, Ora: The Semantic Web. In: Scientific American 284 (2001), Mai, Nr. 5, S. 34–43

[LS11] Lanagan, James ; Smeaton, Alan F.: Using Twitter to Detect and Tag Important Events in Live Sports. In: Artificial Intelligence (2011), S. 542– 545

[OED12] Oorschot, Guido van ; Erp, Marieke van ; Dijkshoorn, Chris: Automatic Extraction of Soccer Game Events from Twitter. In: Erp, Marieke van (Hrsg.) ; Hollink, Laura (Hrsg.) ; Hage, Willem R. (Hrsg.) ; Troncy, Raphaël (Hrsg.) ; Shamma, David A. (Hrsg.): Proceedings of the Workhop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2012) Bd. 902. Boston, USA : CEUR, 11 2012, S. 21–30

[PS09] Patman, F. ; Shaefer, L.: Is Soundex Good Enough for You? On the Hidden Risks of Soundex-Based Name Searching. 41 (2009), Januar, Nr. 1

[QLW+10] Qian, Xueming ; Liu, Guizhong ; Wang, Huan ; Li, Zhi ; Wang, Zhe: Soccer video event detection by fusing middle level visual semantics of an event clip. In: Proceedings of the Advances in multimedia information processing, and 11th Pacific Rim conference on Multimedia: Part II. Berlin, Heidelberg : Springer-Verlag, 2010 (PCM'10), S. 439–451

ZZWV11] Zhao, Siqi ; Zhong, Lin ; Wickramasuriya, Jehan ; Vasudevan, Venu: Analyzing twitter for social tv: Sentiment extraction for sports. In: Proceedings of the 2nd International Workshop on Future of Television, 2011

*Vortragsmanuskript (gehalten auf der Frühjahrstagung des vfm am 25. April 2017)

Konzeption und Aufbereitung kuratierter Twitter-Listen als Recherchewerkzeug

Catharina Boss

Catharina Boss
Technische Hochschule Köln
Institut für Informationswissenschaft
catharina.boss@infonetwork.de

Das Projekt „Kuratierte Twitter-Listen als Recherchewerkzeug“ wurde im Sommersemester 2016 an der Technischen Hochschule Köln im berufsbegleitenden Masterstudiengang Bibliotheks- und Informationswissenschaften (MALIS) durchgeführt und von Prof. Dr. Petra Werner und Miriam Schmitz betreut.

*Vortragsmanuskript (gehalten auf der Frühjahrstagung des vfm am 25. April 2017)

Meist reicht ein Blick in die Nachrichten und es wird schnell klar – Inhalte aus sozialen Medien haben einen festen Platz in der Berichterstattung eingenommen. Redakteure binden Fotos von Facebook oder Instagram in Beiträge ein, zeigen Videos von YouTube oder Snapchat und zitieren Tweets. So genannte „Netzreaktionen“ und User Generated Content sind seit Langem fester Bestandteil journalistischer Formate. Soziale Netzwerke dienen aber nicht nur der Anreicherung von Beiträgen mit Bildmaterial. Worüber das Netz diskutiert, was gerade „viral“ ist und sich überdurchschnittlich schnell in den Communities des World Wide Web verbreitet, dient auch als Trendbarometer und Fundus für die Themenfindung.

Aber nicht alles ist bekanntlich Gold, was glänzt. Denn während soziale Netzwerke einerseits wie eine schier unerschöpfliche Quelle quotenträchtiger Stoffe anmuten, so bergen sie doch andererseits einige Hindernisse und Stolpersteine auf dem Weg zum vermeintlich strahlenden Content. Da wäre zum einen die schiere Masse an Posts und Tweets, an Snaps und Live-Videos. So offenbart allein die Nutzerstatistik der Plattform Twitter aus dem Jahr 2016 eine Bilanz von etwa 500 Millionen Tweets – täglich. Zum anderen müssen authentische Inhalte von Spam und Fakes getrennt werden. Journalisten – und alle, die sie bei ihrer Tätigkeit unterstützen – stehen also vor der nicht zu unterschätzenden Herausforderung, soziale Medien effizient zu beobachten, in ihnen zu recherchieren sowie interessante und gleichzeitig echte Inhalte zu selektieren.

An die Kuration (lat. curare = sorgen, sich kümmern) von Webinhalten werden bestimmte Erwartungen gestellt. Sie erfordert Expertenwissen auf den abgedeckten Gebieten, ein gutes Informationsmanagement, eine enge Vernetzung mit Quellen und Kanälen sowie die Fähigkeit, den Überblick zu behalten. Für Journalisten und Mediendokumentare ist die Kuration von Inhalten eine besondere Herausforderung, denn die Themenvielfalt in der medialen Berichterstattung ist groß. Ressorts decken unterschiedliche Themen ab, die ereignisgetriebene Berichterstattung dominiert die Schlagzeilen. Nachrichtenagenturen und Medienhäuser entwickeln deshalb Strategien zur Kuration sozialer Netzwerke. Das Kuratieren erfordert insbesondere zu Beginn oftmals einen hohen Arbeits- und Zeitaufwand, bringt aber System und somit Überblick in die Recherche. Zum Monitoring eignen sich beispielsweise institutionelle und private Fachwebsites, Verlagsangebote und thematische Gruppen. Einen besonders leichten Einstieg in das Social Media Monitoring stellen thematische Listen dar. Sie bringen Struktur in die Suche, indem sie den Radius der Betrachtung fokussiert eingrenzen. Aufbereitet mittels einer Softwareapplikation können sie ein nützliches Instrument zur Recherche in sozialen Netzwerken sein. In diesem Beitrag wird die Nutzung am Beispiel der Plattform Twitter vorgestellt.

HOW TO: TWITTER-LISTEN ALS RECHERCHEWERKZEUG

Twitter ist eines der bekanntesten sozialen Netzwerke und besticht durch die Grundfunktion, Textmitteilungen mit der begrenzten Länge von 140 Zeichen zu veröffentlichen. Videos, Fotos und Links können ebenfalls eingebettet werden. Auf der Timeline, der Startseite eines Accounts, laufen in Echtzeit alle Nachrichten ein, die von abonnierten Nutzern gepostet werden. Je nach Zahl der abonnierten User kann die Timeline mengenmäßig und inhaltlich schnell unübersichtlich werden. Um die einlaufenden Informationen systematisch zu ordnen, können Listen angelegt werden. Twitter definiert diese Listen als „eine benutzerdefinierte Gruppe von Twitter-Accounts“. Öffnet man eine solche Liste, werden in der Timeline ausschließlich Tweets der enthaltenen User ange-

zeigt. Listen können privat oder öffentlich sein, d.h. sie können anderen Nutzern zur Verfügung gestellt werden, oder nicht. So kann neben dem Erstellen eigener Listen auch den öffentlichen Listen anderer User gefolgt werden. Wer Netzwerke wie Facebook und Instagram mit Listen recherchierbar machen möchte, kann auf Webapplikationen wie Crowdtangle zurückgreifen.

Beim Erstellen von Listen sollten zwei Dinge berücksichtigt werden: der Bedarf und Kriterien, nach denen dieser gedeckt wird. Bei Content Services, dem Archiv von infoNetwork, wird u.a. mit Sachkategorien und ergänzenden Schlagwörtern, die insbesondere ereignisgetriebene Themen beschreiben, gearbeitet. Für das Medienarchiv eines Fernsehsenders könnten folgende Kategorien von Listen definiert werden:

- Listen basierend auf Sachgruppen/-kategorien
- Listen basierend auf ereignisgetriebenen Schlagwörtern
- Allgemeine Listen

Eine Liste, die sich an einer Sachkategorie orientiert, ist „Politik Deutschland“. Eine ähnliche Liste, die aber einem ereignisgetriebenen Schlagwort folgt, ist „Bundestagswahl 2017“. Über diese beiden Typen hinaus bieten sich allgemeine Listen an, z.B. „Deutsche Medien“ oder „Top 50 Journalisten“.

Leider gewährt bisher keine der für das Monitoring von sozialen Medien geeigneten Applikationen das Erstellen von „Unterlisten“ oder Hierarchien, sodass schon beim Anlegen abgewogen werden muss, ob eine übergreifende Liste für eine Kategorie, oder separate Listen für Subkategorien eines Themenkomplexes benötigt werden, wie folgendes Beispiel zeigt: Der Liste „Deutscher Fußball“ wären Teams und Spieler deutscher Ligen hinzuzufügen, wodurch sie sehr viele Accounts enthielte, das Anlegen separater Listen für Ligen und Teams bedeutet aber gleichzeitig mehrere Listen, die recherchiert werden müssten. Beide Optionen neigen unter Umständen zu Unübersichtlichkeit.

Um einen gleichmäßigen Qualitätsstandard zu gewährleisten, lohnt es sich Kriterien zu definieren, nach denen Accounts einer Liste hinzugefügt werden. Hier eine Übersicht möglicher Merkmale:

- Relevanz
- Aktualität/Aktivität
- Reichweite
- Authentizität

Für das Medienarchiv eines Fernsehsenders spielt insbesondere der Faktor Relevanz eine Rolle. Ein Beispiel: infoNetwork produziert für die Mediengruppe RTL, d.h. bei der Erstellung einer Liste „Deutsche Stars und Sternchen“ spielen vor allem solche Prominente eine Rolle, die mit Programm und Berichterstattung der Sendergruppe verknüpft sind – Schauspieler der „Lindenstraße“ oder Kandidatinnen von „Germany‘s Next Topmodel“ sind weniger relevant als Anwärter auf die Dschungelkrone von „Ich bin ein Star – Holt mich hier raus“.

Wenn die Relevanz eines Accounts gegeben ist, sollte im nächsten Schritt seine Aktualität geprüft werden. Eine Person, die selten oder gar nicht postet, büßt ihren Wert für die Beobachtung ein. Mancher Prominenter pflegt keinen Twitter-Account, ist aber auf Instagram aktiv. Möchte man ihn oder sie trotzdem beobachten, sollte die Recherchemethode angepasst werden und z.B. auf das bereits genannte Crowdtangle ausgewichen werden.

Die Reichweite eines Accounts – d.h. die Anzahl der vorhandenen Follower bzw. Personen, die den Nutzer abonnieren – ist ebenfalls ein Auswahlkriterium. Sie sagt zwar grundsätzlich nichts über die Qualität der Tweets aus, lässt aber einen Rückschluss über den gesellschaftlichen Einfluss eines Users zu. Je mehr Abonnenten eine Person hat, desto höher ist ihre Popularität und dementsprechend ihre „Berichterstattungswürdigkeit“.

Aufschluss über die Echtheit eines Accounts kann das Merkmal „verifizierter Account“ bieten, erkennbar an einem blauen Haken neben dem Titel. Twitter stuft Accounts, die von öffentlichem Interesse sind, als verifiziert ein, wenn die Authentizität bei einem Antrag auf Verifikation nachgewiesen wird. Dabei werden u.a. Kontaktdaten, persönliche Angaben und Biografie einer Person oder eines Unternehmens geprüft. Bei Personen verlangt Twitter konkrete Belege über die Relevanz für die Branche, in der sie aktiv sind und behält sich vor, zur Identifikation die Vorlage von Ausweisdokumenten zu fordern. Auf diese Weise sollen authentische Accounts von gefälschten unterschieden werden können. Dies bedeutet aber keinesfalls, dass jeder Account, der nicht verifiziert ist, ein Fake ist. Möchte man einen solchen Account in eine Liste aufnehmen, bietet es sich an, die Person oder Institution näher zu überprüfen, beispielsweise, indem Auftritte in anderen Netzwerken gecheckt oder auf einer ggf. angegebenen Website nachgesehen wird. Auch kann es nützlich sein, die Follower eines Accounts zu begutachten, oder zu verfolgen, ob und in welchen Listen er sich wiederfindet.

Ein Tipp: Natürlich sind selbsterstellte, selbst kuratierte Listen perfekt auf den eigenen Bedarf zugeschnitten, aber oftmals fehlt die Zeit, sich ausreichend mit der Pflege von Listen zu beschäftigen. Es lohnt sich daher immer auch ein Blick über den eigenen Tellerrand hinaus, denn viele Twitter-Nutzer und insbesondere Medienhäuser pflegen kuratierte Listen, die öffentlich zugänglich sind. Diese können als Inspiration für eigene Listen dienen – oder auch abonniert werden. Mit Tools wie Listcopy oder

TweepDiff ist z.B. das Kopieren fremder Listen in den eigenen Account bzw. das Abgleichen von Accounts in Listen auf Doppelungen hin möglich.

TOOLS ZUR AUFBEREITUNG

Mit einem Set an kuratierten Listen ist bereits viel erreicht. In der von Twitter zur Verfügung gestellten Timeline kann jedoch nicht effektiv recherchiert werden, da sie keine nennenswerten Such- oder Filteroptionen bietet. Glücklicherweise bietet der Markt zu diesem Zweck diverse kostenfreie und –pflichtige Web-Anwendungen an, die sich in Aufbau und Funktion oftmals ähnlich sind. Drei Bedingungen könnten bei der Entscheidung für ein Tool zum Tragen kommen: Preis-/Leistungsverhältnis, Funktionsumfang (insb. Such- und Filteroptionen) und Usability.

Bewährt und beliebt ist die kostenfreie App TweetDeck, die mittlerweile von Twitter übernommen und als Browseranwendung angeboten wird. Der Schwerpunkt liegt auf der Erstellung und Sortierung von Kolumnen, auch Streams genannt, die neben selbsterstellten Listen u.a. auch einzelne Schlagwörter, User, Trends und Hashtags abbilden können. TweetDeck bringt eine Vielzahl von Funktionen mit:

- Erstellung und Bearbeitung von Twitter-Listen
- Sortierung der Kolumnen per Drag & Drop
- Suche in Listen mittels Such-Operatoren und Filtern
- Detailansicht mit Einblendung von Kommentaren
- Share-Funktion per E-Mail
- Alert per optischem Signal oder Ton oder Desktop-Benachrichtigung
- Personalisierung durch Anpassung von Design, Spalten- und Schriftgröße

Als besonders positiv zu bewerten sind die Funktionen und Befehle, die direkt in den Kolumnen ausgeführt werden können. Für den schnellen Überblick können Tweets überflogen werden; schürt ein bestimmter Tweet aber Interesse, kann eine Detailansicht aufgerufen werden, die angehängte Fotos vergrößert oder Kommentare zum Tweet anzeigt. Auch die Filteroptionen erweisen sich als nützlich. So können Inhalte differenziert (nur Text, Text und Foto/Link), Tweets nach Quelle (verifizierter User, alle User) und Art (Tweet, Retweet) unterschieden und Tweets mit Suchbegriffen eingegrenzt werden. Innerhalb der Suchschlitze ist die Anwendung von diversen Operatoren – z. B. Booleschen Operatoren, aber auch Geodaten für eine lokationsbasierte Suche – möglich. Ein Nachteil entsteht indes bei der Nutzung von TweetDeck durch mehrere Nutzer, da keine unterschiedlichen Ansichten gespeichert werden, sondern TweetDeck immer wieder neu konfiguriert werden muss. Dies ist insbesondere bei optischen Einstellungen ärgerlich. Abhilfe würde die integrierte Team-Funktion schaffen, mit der mehrere Personen einen Account verwalten können. Allerdings benötigt dazu jeder Nutzer einen eigenen Twitter-Account.

Weitere Webanwendungen, die sich auf Monitoring und Recherche sozialer Netzwerke spezialisiert haben, sind Hootsuite und CrowdTangle. Während Hootsuite optisch TweetDeck sehr ähnlich ist und das Anlegen mehrerer Kolumnen zulässt, bietet CrowdTangle die aus sozialen Netzwerken bekannte weniger übersichtliche Timeline. Beide Programme bestechen durch die Möglichkeit, mehrere soziale Netzwerke beobachten zu können. Des Weiteren legen beide Tools einen Schwerpunkt auf die Verwendung zu Marketingzwecken, indem sie Analysefunktionen zur Verfügung stellen. CrowdTangle besticht zudem durch ein Scoring-System („Weights“), mit dem überdurchschnittlich gut bewertete Posts nach selbst zu vergebendem Fokus herausgefiltert werden.

EINSATZ IM REDAKTIONELLEN ALLTAG

Das Monitoring sozialer Medien sollte, wenn von Mediendokumentaren durchgeführt, so redaktionsnah wie möglich angesiedelt sein. Da Instrumente wie Twitter-Listen die redaktionelle Arbeit unterstützen sollten, ist es sinnvoll, sie an einer Schnittstelle einzugliedern, die sich mit Recherche und Materialbereitstellung beschäftigt und so in direktem Kontakt zur Redaktion steht. Viele Medienhäuser haben Planungsredaktionen, die in der Themen- und Bildrecherche unterstützt werden können.

Beim Einsatz kommt es auf eine fallbezogene Verwendung von Recherchemethoden und –werkzeugen an. Für das Aufspüren aktueller Trends und viraler Hits eignen sich Anwendungen mit auf Social Media zugeschnittenen Auswertungsfunktionen, wie CrowdTangle, besser als TweetDeck. Im Fall von Breaking News, z.B. in Zusammenhang mit Terroranschlägen o.ä., bieten sich geodatenbasierte Anwendungen wie Ban.jo an. Twitter-Listen in Kombination mit TweetDeck eignen vor allem als breitgefächertes Informationsmittel, denn das Herausfiltern unverbrauchter Themen benötigt ohne zielgerichtete Suche einen höheren Zeitaufwand. •

SCREENSHOTS

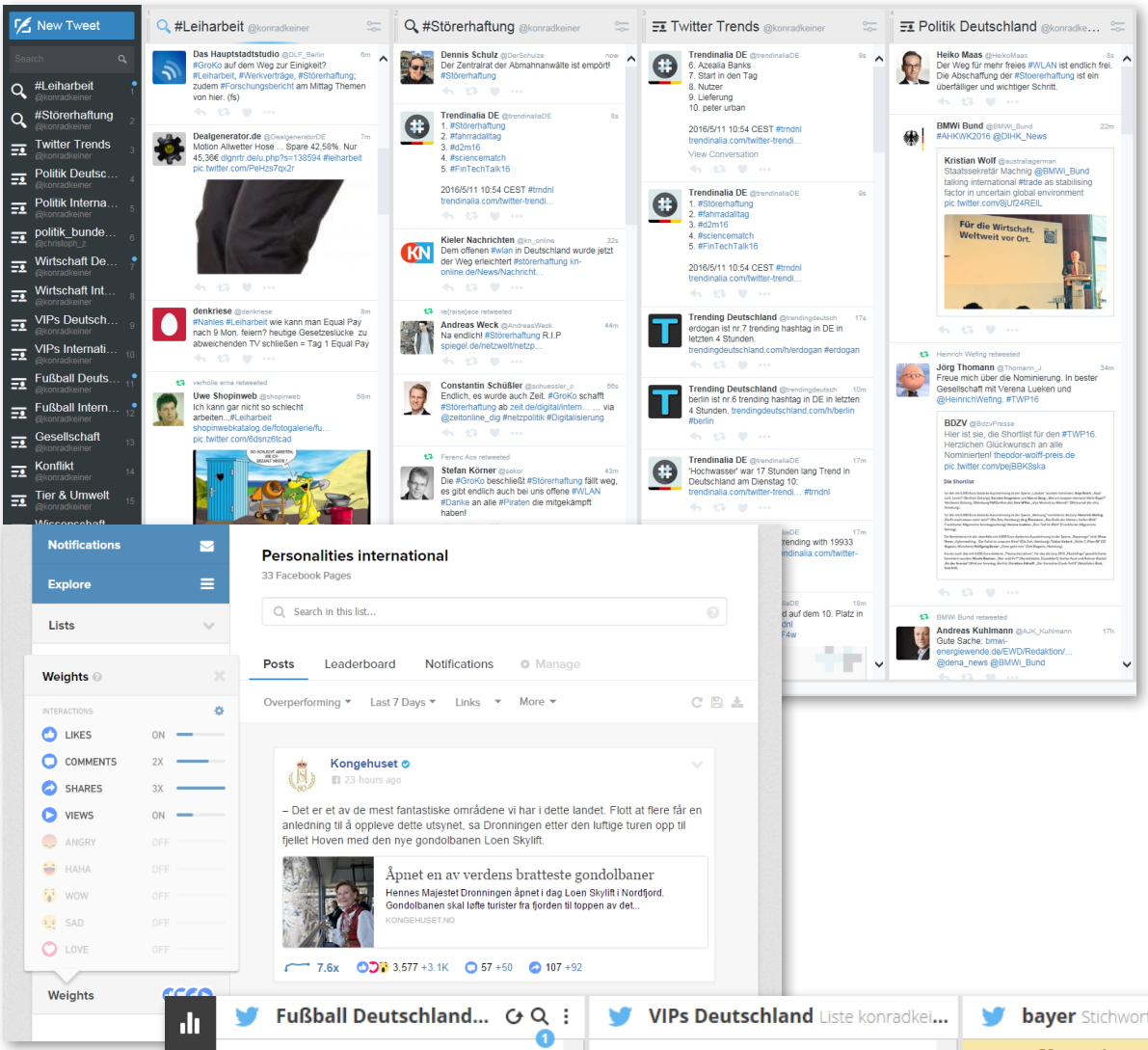

Abbildung 1: Nutzeroberfläche der Anwendung „TweetDeck“ (Screenshot).

Abbildung 2: Nutzeroberfläche der Webanwendung „CrowdTangle“ (Screenshot).

Abbildung 3: Nutzeroberfläche der Webanwendung „Hootsuite“ (Screenshot).

Von »Old Media« zum interaktiven Radio?

Julia Lorke

Dr. Julia Lorke
Imperial College London
MSc Science Communication
South Kensington Campus
London SW7 2AZ
+49 2821 8067 39730
julialorke@gmail.com

*Vortragsmanuskript (gehalten auf der Frühjahrstagung des vfm am 25. April 2017)

Man startet gemeinsam in den Tag, sobald der Radiowecker klingelt, man liest gemeinsam Zeitung und fährt gemeinsam zur Arbeit. Spätestens zur Rush Hour trifft man sich im Auto wieder, kocht gemeinsam, geht zu Bett und ist am nächsten Morgen, wenn der Wecker klingelt, wieder vereint. Die Rede ist hier nicht von zwischenmenschlichen Beziehungen sondern es geht um die Beziehung zwischen dem Radio und seinen Hörern. Radio wird nicht umsonst als „the intimate medium" bezeichnet, denn Radio begleitet uns in Alltagssituationen, sogar solche, die wir sonst nur mit unserem Partner oder unserer Familie verbringen. Aber selbst wenn wir ganz alleine Radio - oder natürlich Podcasts – hören werden wir meist direkt angesprochen von den Moderatoren. Zudem sind Radiosendungen häufig live, oder zumindest als „as-live"-Format produziert, um genau diesen live-Charakter zu bewahren. Es scheint also häufig so, als würde man als Hörer tatsächlich Zeit mit den Moderatoren der Radiosendung verbringen. Allerdings scheint die Kommunikation hier eher unidirektional zu sein, und das ist eine eher wenig attraktive Kommunikationsform in zwischenmenschlichen Beziehungen. Dank des technologischen Fortschritts sind jedoch auch im Bereich von Radio und seinem Publikum vielfältige Kommunikationsformen möglich. Nach Tiziano Bonini lässt sich die historische Entwicklung der Beziehung des Mediums Radio zu seinem Publikum in vier Phasen einteilen (siehe Tabelle 1).

Phasen	Merkmale
1) 1920-1945	- Propagandainstrument - Medium für Bildungs- und Erziehungszwecke - Publikum ist unsichtbar für Radiomacher und vice versa - Leserbriefe => top-down/ one -to-many broadcast
2) 1945-1994	- Transistorradio erhöht Mobilität - Freie Radiosender, Piratensender - Call-ins
3) 1994-2004	- Citizen journalism - Email, Textnachrichten - Podcasttechnologie
4) 2004-heute	- Social network sites (social media) => Kommunikationsformen: horizontal; one-to-one, one-to-many, many-to-one und many-to-many

Diese Phasen zeigen deutlich, wie technologische Entwicklungen dazu beigetragen haben, dass neue Kommunikationsformen fürs Radio möglich wurden. Durch Smartphones und Social Media kann das Publikum nun nicht mehr nur passiver Konsument sein, sondern aktiv zur Programmgestaltung beitragen. Mehr Partizipation aufseiten des Publikums, hat aber auch Konsequenzen für die Tätigkeit von Produzenten und Redakteuren, so gewinnt bei interaktiven Formaten das Kuratieren an Bedeutung. Zudem ermöglichen Plattformen wie Facebook, Twitter & Co die direkte Kommunikation zwischen Radiomachern und Hörern außerhalb der Sendezeit.

Wie so oft ist also in der Theorie vieles möglich, aber finden diese Möglichkeiten wirklich Eingang in unseren Alltag? Sind wir wirklich schon in Phase 4 angekommen? Sind die Tage des einsamen Radiomachers, der seiner anonymen, passiven Hörerschaft Nachrichten ins Ohr spricht, wirklich gezählt? Wird das Potential von Social Media zur Interaktion zwischen Radiomachern und Hörern für einen echten Dialog mit ihrem Publikum oder lediglich als kostengünstiges Marketinginstrument genutzt?

Diesen Fragen bin ich in meiner Abschlussarbeit im Rahmen meines „Science communication"- Masterstudiengangs am Imperial College London nachgegangen und zwar an einem konkreten Beispiel: Radiosendungen rund um Technologie und Wissenschaft. Aus der Perspektive der Wissenschaftskom-

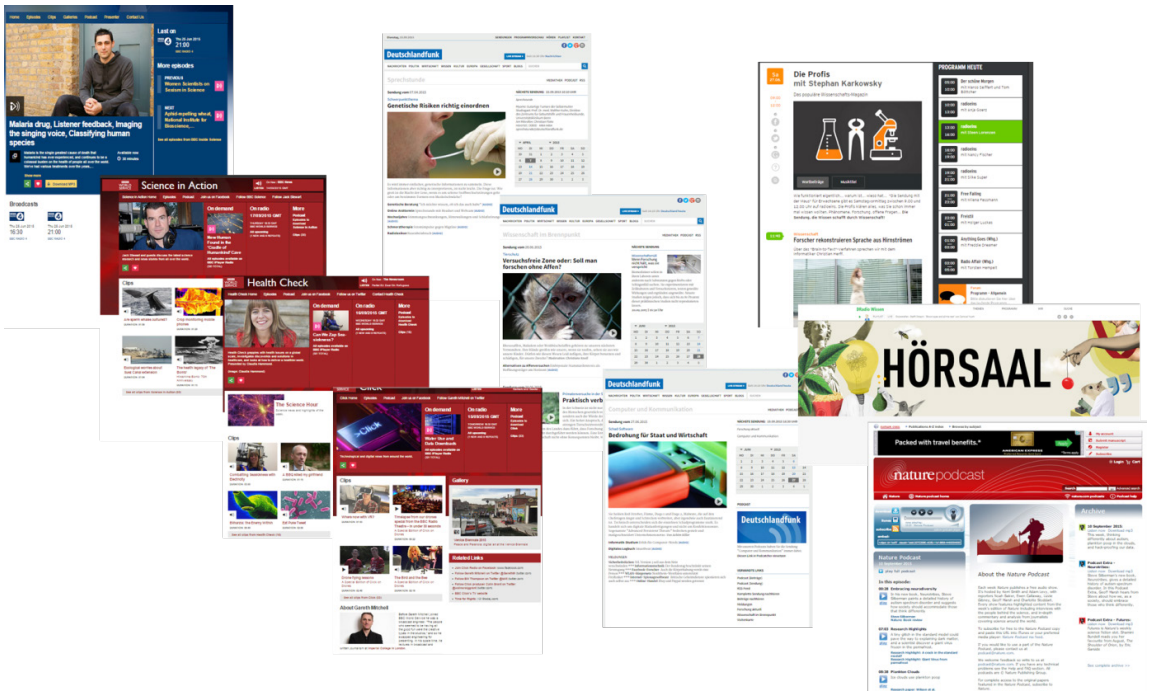

Abbildung 1: Screenshots der analysierten Websites

munikation tangiert diese Fragestellung einen zentralen akademischen Diskurs. So hat in der Theorie der Wissenschaftskommunikation ein Paradigmenwechsel vom Defizit-Modell zum Dialog-Modell stattgefunden. Die Frage ist nun, inwieweit sich jedoch traditionelle Formate wie Radioprogramme zu Naturwissenschaften und Technik dem Dialog zwischen Experten und Gesellschaft öffnen. Gerade durch den Fokus auf ein bestimmtes Themengebiet macht hier eine Analyse auf Programmebene mehr Sinn als auf Senderebene.

VERGLEICHENDE WEBSITENANALYSE

Ein erster Anlaufpunkt für Hörer, um mit einem Radioprogramm online zu interagieren, ist die Website der entsprechenden Sendung. Im Rahmen der Studie habe ich im Zeitraum Juli-August 2015 Websites der folgenden neun Radioprogramme und eines Podcasts analysiert.

BBC Radio 4:
- Inside Science

BBC World Service:
- Click
- Science in Action
- Health Check

Deutschlandfunk
- Sprechstunde
- Wissenschaft im Brennpunkt
- Computer und Kommunikation

RBBradioeins
- Die Profis

DRadio Wissen
- Hörsaal

Nature Podcast

Neben Unterschieden in Bezug auf die Dauer der Verfügbarkeit von Streaming- und Downloadangeboten sowie der Bandbreite und Regelmäßigkeit von Multimediainhalten, zeigten sich auch unterschiedliche Social-Media-Strategien. Während Programmwebsites von Deutschlandfunk einheitlich auf die Social Media Accounts des Senders linkte, wirkte die Verlinkung von Social Media Accounts auf den Programmwebsites der BBC eher willkürlich (Tabelle 2).

Generell linken aber alle Programmwebsites, außer die des *Nature Podcasts*, zu Social Media Accounts. Bis auf eine Ausnahme sind die Accounts jedoch nicht programmspezifisch, sondern Accounts des Senders, der Moderatoren oder allgemeinere Facebook-Newsseiten. *Click* ist die einzige Sendung, die zu einer programmspezifischen Präsenz linkt. Bei der *Digital Planet Listeners* Gruppe handelt es sich um eine Facebook-Gruppe, die von Hörern initiiert wurde und zu der der Moderator der Sendung erst nachträglich Administratorrechte bekommen hat. Diese Sonderstellung von *Click* gab Anlass dazu, die Interaktionen zwischen dem *Click*-Team und den *Click*-Hörern genauer zu untersuchen. Vergleichend dazu wählte ich Inside Science aus, da es in dieser Sendung zweimal im Jahr eine Sonderausgabe gibt, in der Experten Fragen der Hörer beantworten. Diese

Tabelle 1: Die historische Entwicklung der Beziehung zwischen Radio und seinem Publikum (verändert nach Bonini, T., 2014. *The new role of radio and its public in the age of social network sites*. First Monday, Volume 19, Number 6)

Tabelle 2: Links zu Social Media Accounts auf den Programmwebsites

Programm	Facebook	Twitter
Health Check	BBC World Service (Page)	BBC World Service
Click (vormals Digital Planet)	Digital Planet Listeners (Group)	Presenter, Producer & Co-Host
Inside Science	-	BBC Radio 4
Science in Action	BBC Science News (Page)	BBC Science

Facebook-Seiten/Gruppen	
BBC Science News	Digital Planet Listeners
• 234283 Follower • 77 Posts • 363 Kommentare & 3440 likes • 323 geteilte Beiträge • 1 Kommentar von BBC Science News • 23 Posts über die STEM- Radioprogramme	• 9236 Mitglieder • 59 Posts (18 Mitglieder) • 414 Kommentare & 643 likes • Gareth Mitchell: 4 Posts & 11 Kommentare • 2 Posts Audience Engagement

Tabelle 3: Ergebnisse der Analyse der Hashtags der Radiosendungen

Hashtags	
#BBCInsideScience	#BBCClickRadio
124 (106+ 18 Antworten) 15 vom Moderator 11 von anderen BBC-Mitarbeitern 0 vom Producer 5 von Gästen	44 (40 + 4 Antworten) 9 vom Moderator 8 vom Co-Moderator 6 vom Producer 4 von Gästen
60% Promotion 27% Behind the scenes 13% Antworten 0% Audience Engagement	9% Promotion 74% Behind the scenes 17% Antworten 0% Audience Engagement

Tabelle 4: Ergebnisse des Vergleichs der BBC Science News Facebook-Page und der Digital Planet Listeners Facebook-Gruppe

fand auch im Untersuchungszeitraum von April bis Juni 2015 statt. Die Hörer wurden in den Sendungen zwar aufgefordert, über Email und Twitter in Kontakt zu treten, allerdings wurden nur in 2 von 12 regulären Sendungen Kommentare der Zuschauer in die Sendung aufgenommen. *Click* hingegen fordert Hörer auf, über die Facebook-Gruppe oder Twitter in Kontakt zu treten. Hörerbeiträge wurden im Untersuchungszeitraum zwar nicht in die eigentliche Sendung aufgenommen, aber in die Podcast-Extras und das in 11 von 12 Episoden. Im Rahmen des Podcasts wurden auch Livestreams über Periscope angekündigt. Interaktionen mit den Hörern scheint also bei Inside Science eher eine Ausnahme im Rahmen der eigentlichen Sendung zu sein, während es bei *Click* ein fester Bestandteil des Podcasts ist.

CASE STUDIES

Eine Analyse der Tweets mit den Hashtags der Programme #BBCInsideScience und #BBCClickRadio über den Zeitraum von zwölf Wochen, zeigt deutliche Unterschiede. #BBCInsideScience ist deutlich aktiver als #BBCClickRadio (Tabelle 3). Der Anteil der Tweets vom *Click*-Team ist allerdings höher als der vom *Inside Science*-Moderator. Eine qualitative Analyse zeigt, dass der Schwerpunkt der Tweets bei *Inside Science* auf der Promotion der Sendung liegt, während das *Click*-Team Twitter eher zu nutzen scheint, um Einblicke hinter die Kulissen der Sendung zu geben.

Alternativ zur Interaktion über Hashtags, ermöglicht Twitter auch den direkten Kontakt mit den Moderatoren der Sendung. Der Vergleich der Aktivitäten der Moderatoren auf Twitter am Tag der ersten Ausstrahlung der wöchentlichen Sendung und dem Tag danach zeigt, dass Adam Rutherford (Inside Science) deutlich aktiver ist als Gareth Mitchell (Click). Diese Suche nach @AdamRutherford ergab 366 Tweets in zwei Tagen, davon waren 60 Tweets vom Moderator selbst gepostet. Im Vergleich dazu ergab die Suche nach @GarethM nur 22 Tweets, davon sind allerdings 10 vom Moderator selbst. 32% der Tweets von Adam Rutherford beziehen sich auf die Radiosendung, bei Gareth Mitchell sind es 70%. Twitter wird also von beiden Moderatoren genutzt, um über die Radiosendung zu kommunizieren. Twitter hat demnach Potential als Kommunikationsplattform für Moderatoren. Dabei ist einerseits eine hohe Aktivität der Moderatoren auf Twitter aus Hörersicht sicherlich erwünscht, allerdings stellt sich die Frage, ob die Aktivität wie im Fall von Adam Rutherford nicht auch zu hoch und zu wenig auf die Sendung bezogen sein könnte, um Hörer zum Folgen und zur Interaktion auf Twitter anzuregen. Die fehlenden Audience Engagement-Aufrufe und der geringe Anteil von Antworten zeigen vielleicht an, dass Twitter

nicht die beste Plattform für horizontale Kommunikation innerhalb einer Community ist.

Auf den BBC Websites wird, etwa im Fall von *Science in Action*, auf die Facebook-Page *BBC Science News* verlinkt und im Fall von *Click* auf die *Digital Planet Listeners* (DPL)-Facebook-Gruppe. Nicht überraschend ist hier, dass die News-Page deutlich mehr Follower hat als die DPL-Facebook-Gruppe, dementsprechend höher ist auch die Anzahl der Likes (Tabelle 4). Die Anzahl der Posts im Untersuchungszeitraum war hingegen auf einem ähnlichen Level, mit 77 Posts in *BBC Science News* und 59 Posts von 18 verschiedenen Mitgliedern der DPL-Gruppe. Die Anzahl der Kommentare war mit 414 sogar höher in der DPL-Gruppe als mit 363 auf der News-Page. Der Moderator ist in der DPL-Gruppe aktives Mitglied mit vier Posts und, im Sinne der horizontalen Kommunikation noch wichtiger, elf Kommentaren. Bei zwei der Posts handelt es sich um klare Aufforderungen zum Audience Engagement. Diesen kamen die Hörer in Form von Kommentaren nach, welche dann wiederrum vom Moderator im Podcast aufgegriffen wurden.

In der DPL-Facebook-Gruppe finden also tatsächlich Interaktionen zwischen Moderatoren und Hörern sowie Hörern untereinander statt. Hörer können kontinuierlich aktiv zur Programmgestaltung beitragen. Im Untersuchungszeitraum galt dies nur für die Podcast-Extras, mittlerweile sind aber auch Kommentare im Rahmen des Radioprogramms erwähnt worden. Ein eindrucksvolles Beispiel dafür, dass hier tatsächlich Kommunikation im Sinne von Bonini's vierter Phase stattfindet, ist ein von einer Hörerin verfasster Audiobeitrag, der als Folge eines Aufrufs zum Audience Engagement vom Moderator in der Facebook-Gruppe entstanden ist, im Radioprogramm gesendet und dann in der Facebook-Gruppe von Hörern und der Hörerin selbst diskutiert wurde.

Als Ergebnis der Studie lässt sich also festhalten, dass interaktives Radio im Sinne von Boninis vierter Phase möglich ist, auch für Naturwissenschafts- und Techniksendungen. Ein solcher Ansatz kann von Hörern initiiert werden, braucht aber die kontinuierliche Unterstützung durch die Radiomacher. Die geringe Anzahl programmspezifischer Social Media-Accounts zeigt, dass das Potential hier noch bei Weitem nicht ausgeschöpft ist.

Die Auseinandersetzung mit dem Thema führte für mich zu weiteren noch offenen Fragen. So bin ich mir gar nicht mehr so ganz sicher, was eigentlich ein Radioprogramm ist – die ausgestrahlte Sendung, der Podcast oder beides und alle Konversationen rund um die Sendung in den sozialen Medien?

Wessen Aufgabe ist die Interaktion mit den Hörern? Sollte das vertraglich geregelt und vergütet werden? Oder soll die Ausgestaltung der Interaktion, wie im Fall von *Click*, von der Motivation und Einstellungen der Radiomacher abhängen:

> *"To me there is no point in radio unless people are interacting and being involved. It enhances the programme. Radio should be a conversational, interactive medium. Luckily the social media has come along and allows us to do that."*
> *Gareth Mitchell, 2015*

Gerade bei öffentlichen Radiosendern stellt sich die Frage, wie man den Zugang zur Interaktionsplattform sicherstellt. Will man sich von Drittanbietern abhängig machen oder gibt es dazu eigentlich schon keine Alternative mehr?

Auch rechtliche Regelungen sind ein interessanter Aspekt. Wer hat die Rechte an Tweets und Facebook-Posts? Besteht ein Konflikt bei öffentlich-rechtlicher Finanzierung von Inhalten welche dann Werbeeinnahmen für Social-Media-Plattformen erzielen? Wie verhält es sich mit den Rechten an user-generierten Inhalten, die Eingang ins Radioprogramm finden?

Wie sind Monitoring und Archivierung geregelt? Und was sollte archiviert werden - Posts vom Senderaccount, von persönlichen Accounts der Radiomacher und von Hörern? Eine Dokumentation der Interaktionen mit Medien würde sicher großes Potential bieten etwa für die Medienrezeptionsforschung.

Im Rahmen der Diskussion im Anschluss an den Vortrag wurde auf die Schaffung von Stellen für Community Manager beim Deutschlandradio hingewiesen und darüber spekuliert, wie wohl die nächste Phase der Beziehung zwischen Radio und seinem Publikum aussehen könnte. •