

Alexander Hinneburg / Christian Oberländer

# Getting the Story from Big Data: Interaktive visuelle Inhaltsanalyse für die Sozialwissenschaften mit dem Topic- Explorer am Beispiel Fukushima

**Abstract** Immer mehr Menschen wollen öffentlich gehört werden und ihre Meinung einem breiten Publikum mitteilen. Dieser Trend lässt sich u. a. an der steigenden Beteiligung an sozialen Netzwerken (SNS) ablesen, auf denen folglich große Textmengen in digitaler Form entstehen. Sozialwissenschaftler/innen sind an den Inhalten und Prozessen gesellschaftlicher Interaktion und Meinungsbildung interessiert. Diese auf den SNS entstehenden riesigen Datenmengen (Big Data) in Textform können sie mit herkömmlichen Methoden jedoch bisher kaum auswerten. Als Lösung bieten sich Topic Models an. Sie werden in der Informatik schon länger angewendet und ermöglichen die explorierende Analyse großer Textsammlungen, die persönliche Äußerungen enthalten, indem sie eine Analyse nach häufig verwendeten Themen (*topics*) ermöglichen. Um Sozialwissenschaftlern Zugang zu diesen neuen Forschungsmöglichkeiten zu eröffnen, wurde in Zusammenarbeit von Informatik und Japanologie das Werkzeug TopicExplorer entwickelt. Als Pilotprojekt wurde ein Japan- und sozialwissenschaftlich aktuelles Thema gewählt, nämlich ein Ausschnitt aus der japanischen Internet-Debatte über die Atomkatastrophe von Fukushima und ihre Folgen. Anhand japanischer Blogs über radioaktiv verseuchtes Rindfleisch infolge der Atomkatastrophe von Fukushima im Jahr 2011 wurde demonstriert, wie mit dem TopicExplorer die Wahrnehmung eines gesellschaftlichen Ereignisses im Internet widergespiegelt, sinnvoll strukturiert und für eine vertiefte Analyse aufbereitet werden kann.

## 1. Topic Models für die Sozialwissenschaften

Das Kernproblem bei der Analyse großer Textmengen mit vielen Autoren ist, dass diese Texte zwar aufgrund ihrer Vielfalt interessant für Geistes- und Sozialwissenschaftler sind, aber ihre vollumfängliche Verarbeitung durch Menschen nicht durchführbar ist, weil diese zu zeitaufwendig bzw. nicht zumutbar wäre.

Außerdem wäre das Durcharbeiten aller persönlichen Äußerungen uninteressant, weil viele von ihnen entweder sich nur graduell vom Mainstream unterscheiden oder, im anderen Extrem, abstruse Meinungen vertreten. Dagegen macht explorierendes Text-Mining mit probabilistischen Topic Models große Textmassen spannend für Geistes- und Sozialwissenschaftler, weil mit diesen Methoden ohne manuelle Vorverarbeitung und Annotation, allein durch die Analyse der Wortverteilungen in den Dokumenten eine gegebene Textsammlung verdichtet und zu oft unerwarteten Themen strukturiert werden kann (Blei 2012). Die ermittelten Themen bieten Geistes- und Sozialwissenschaftlern einen fokussierten Zugang zu den informationshaltigen und somit tatsächlich interessanten Texten, die zwischen Mainstream und absurden Einzelpositionen verortet sind. Erst mit der Hilfe von Topic Models können sie sich also ein differenziertes Bild von den Inhalten großer Textsammlungen machen (Evangelopoulos und Visinescu 2012).

Trotz des großen Potenzials von Topic Models besteht eine signifikante Lücke bei ihrer Anwendung. Fernando Pereira, Forschungsdirektor von Google, stellt fest: "While they [topic models] are intriguing, we haven't yet gotten to the point that we can say, 'Yes, this is a practical tool.'" (Anthes 2010). Diese Lücke entsteht dadurch, dass Informatik-Instrumente fehlen, die den Transfer von probabilistischen Aussagen über Wortverteilungen in Dokumenten hin zu inhaltlichen Argumenten – in unserem Hallenser Projekt in geistes- und sozialwissenschaftlichen Kontexten – erlauben (Chang et al. 2009). Denn Topic Models berechnen die latenten Themen nicht nach semantischen Gesichtspunkten, sondern folgen allein informationstheoretischen Kriterien. Es bedarf also geeigneter Auswertungswerkzeuge, die dem Anwender helfen, inhaltliche Informationen und Schlussfolgerungen aus den Berechnungsergebnissen der Topic Models zu ziehen (Blei 2012). Dies wurde auch im Positionspapier (Ramage et al. 2009) des Mimir Projects bestätigt, in dem zwei Barrieren charakterisiert wurden, die in der Zusammenarbeit zwischen Informatikern und Sozialwissenschaftlern bei der Anwendung von Topic Models häufig auftraten: (a) technische Zugänglichkeit der Ergebnisse des Themenmodells für Sozialwissenschaftler und (b) Vertrauen der Sozialwissenschaftler in die Ergebnisse des Themenmodells.

Bei der Entwicklung von Auswertungswerkzeugen für Sozialwissenschaftler muss beachtet werden, dass die Anwender normalerweise über kein Hintergrundwissen zu Text-Mining und probabilistischen Modellen verfügen. Das heißt, die Anwender benötigen Auswertungswerkzeuge, um sich ohne detaillierte Kenntnisse der zugrunde liegenden mathematischen Theorie aus den Ausgangsdaten (der Dokumentensammlung) und einem daraus berechneten Themenmodell eine Plausibilitätsstruktur zur Interpretation der Themen und somit zum Verständnis des Inhaltes der Dokumentensammlung zu erschließen. Dabei dürfen ihnen auch keine falschen Schlüsse suggeriert werden. Deshalb

ist ein weiterer wichtiger Punkt, dass die Anwender/innen über die inhaltliche Auswertung hinaus auch bei der kritischen Überprüfung ihrer Interpretationen durch das Werkzeug unterstützt werden. Beispielsweise hat sich die Möglichkeit, die berechneten Themen zu den relevanten Originaldokumenten zurückzuverfolgen, als besonders hilfreich für die Anwender herausgestellt, denn dies gibt ihnen die Möglichkeit, sowohl ihre inhaltliche Interpretation der probabilistisch berechneten Themen zu überprüfen als auch aussagekräftige Textstellen aufzufinden und gezielt für ihre wissenschaftliche Argumentation zu nutzen.

Die Interaktivität des Auswertungsprozesses stellt dabei hohe Anforderungen an die Effizienz der zu entwickelnden Software-Komponenten. Neben inhaltlichen bestehen also auch hohe technische Anforderungen. Da sowohl die Dokumentsammlungen als auch die daraus berechneten Topic Models sehr große Datenmengen umfassen, müssen außerdem effiziente Such- und Auswertungsalgorithmen entwickelt werden. Diese Algorithmen sollen mit effektiven Visualisierungstechniken für Topic Models (Gohr et al. 2010, Hinneburg et al. 2012, Gohr et al. 2013) zu interaktiven, Web-basierten Benutzeroberflächen verknüpft werden.

## 2. Der TopicExplorer

### 2.1 Entwicklung der Analyseplattform

Das Analysesystem TopicExplorer (<http://topicexplorer.informatik.uni-halle.de>) besteht aus zwei Teilen: der erste Teil ist eine schlanke Software-Plattform, die nicht nur die entwickelten Auswertungswerkzeuge trägt, sondern die es auch ermöglicht, eine Vielzahl von neuen Funktionalitäten zur Exploration von Dokumentensammlungen umzusetzen und in das System zu integrieren, ohne es grundsätzlich im Kern verändern zu müssen. Der zweite Teil baut darauf auf und bildet das eigentliche TopicExplorer-System, das die speziellen Analysefunktionen zur Exploration großer Textsammlungen beinhaltet.

Der TopicExplorer stellt sich für den Anwender als eine interaktiv nutzbare Web-Applikation zur Analyse großer Dokumentensammlungen dar. Die Vorverarbeitung der Texte, die Themenberechnung sowie das Verknüpfen der berechneten Themen mit den Dokumenten sind aufwendige Prozesse, die je nach Größe der Dokumentensammlung zwischen Stunden und Tagen in Anspruch nehmen. Deshalb wurde das TopicExplorer-System in drei Bereiche aufgeteilt: Vorberechnung, Server und Client. Zum Bereich Vorberechnung gehören alle Software-Komponenten mit lang laufenden Aufgaben: Wort- und Satzanalyse der Dokumente, die Themenanalyse der aufbereiteten Dokumente, das Verknüpfen der berechneten Themen mit den Texten und ihren Meta-Daten. Das Ergebnis der

Vorbereitung ist eine Datenbank, die von den Software-Komponenten des Servers mit schnell laufenden Anfragen ausgelesen werden kann, um interaktiv den Client mit den benötigten Daten zu versorgen.

Die Nutzung von Software-Entwurfsmustern erlaubt es zu planen, für welche zukünftige Funktionalität die Software flexibel bleiben soll. Für die Bereiche Vorbereitung und Server-Backend wurde das Software-Entwurfsmuster Filter Pipelines in Kombination mit dem Kommando-Muster verwendet. So bleibt die Software-Architektur flexibel und kann immer wieder neue Verarbeitungsschritte in die Datenaufbereitung und Analyse einzufügen, die während der Entwicklung auftauchen. Dieses kombinierte Software-Entwurfsmuster wurde zum Konzept automatisch konfigurierbarer Workflows weiterentwickelt. Die Software-Module für die Workflows sind in der separaten Bibliothek `CommandManager` ausgelagert, die Open Source (<https://github.com/hinneburg/CommandManager>) verfügbar ist.

Als Ergebnis der Nutzung von automatisch konfigurierbaren Workflows sind alle Funktionalitäten des `TopicExplorer`-Systems in kleinen, übersichtlichen Modulen, sogenannten Kommandos, implementiert. Die Kommandos wiederum werden zu bereichsüberspannenden Plugins gebündelt. Somit können ganze inhaltlich zusammenhängende Analysebereiche an- und abgeschaltet werden. So können aufwendige Workflows entstehen, die sich inkrementell aus den Abhängigkeiten der einzelnen Kommandos ergeben. Abbildung 1 zeigt als Beispiel den automatisch konfigurierten Workflow für den Bereich Vorverarbeitung.

Die meisten Funktionalitäten sind mithilfe eines Datenbanksystems in der Programmiersprache SQL deklarativ programmiert. Ein Vorteil des Datenbanksystems ist, dass die Datenstrukturen des probabilistischen Themenmodells als relationales Modell durch Tabellen explizit ausgedrückt werden. In zukünftigen Versionen sollen die rechenintensiven Abschnitte der Vorbereitung in `SparkSQL` (Armbrust et al. 2015) umgesetzt werden.

Der Client ist als Web-Applikation in JavaScript implementiert. Das Benutzer-Interface des `TopicExplorers` ist mit einem modularisierbaren Web-Client-Programmiersatz auf der Basis von `KnockOutJs` umgesetzt. Die entwickelte Lösung ist flexibel genug, um Javascript-Bibliotheken z. B. für Informationsvisualisierungen mit `D3js`, auf einfache Weise einzubinden.

## 2.2 Visuelle, interaktive Benutzerschnittstelle

Das `TopicExplorer`-System umfasst in der derzeitigen Ausbaustufe folgende Funktionen: (1) Themen-Rankings von Dokumenten, (2) Dokumentansichten mit Themenzuordnungen der einzelnen Wörter, (3) durchgängige Themenzuordnungen in allen Ansichten durch Farbe und symbolische Nummerierung,

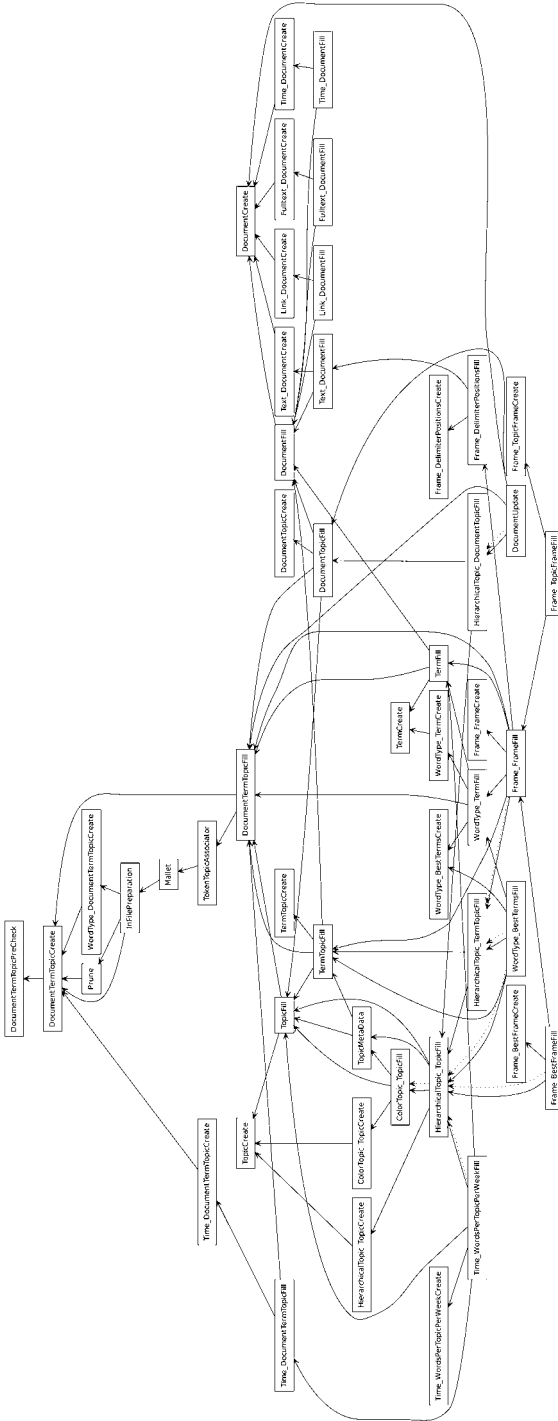


Abbildung 1: Automatisch konfigurierter Workflow-Graph für den Bereich Vorverarbeitung.

(4) automatische Vervollständigung von Suchwörtern mit thematischer Zuordnung, (5) ähnlichkeitsorientierte lineare Anordnung der Themen, (6) thematische Wort-Rankings, (7) zeitliche Analyse und Entwicklung von Themen, (8) Themendarstellung durch bestimmte Wortarten und TopicFrames, (9) themenbasierte Stichwortsuche und (10) interaktives Zusammenfassen und Aufspalten von verwandten Themen.

Basisfunktionen des TopicExplorer-Systems sind Stichwortsuche, Anzeige von Dokument-Rankings sowie die Themendarstellung und -navigation. Abbildung 2 zeigt die Benutzeroberfläche des TopicExplorers am Beispiel einer Dokumentsammlung der 10.000 längsten englischsprachigen Wikipedia-Artikel. Die Themen werden als Wortlisten dargestellt. Ähnliche Themen werden benachbart angeordnet – so gehören beispielsweise in Abbildung 2 die vier gelben nebeneinander liegenden Themen von links alle in den Bereich Unterhaltung. Mittels des horizontalen Sliders kann der Anwender das Spektrum der Themen durchsehen, ohne dass inhaltliche Brüche auftreten. Die Farbe kann aus diesem Grund als Farbverlauf von links nach rechts den Themen zugeordnet werden. Der Farbcode dient dem schnellen approximativen visuellen Referenzieren der Themen in den weiteren Ansichten des TopicExplorers. Das Klicken auf den fett gedruckten Titel eines Themas öffnet einen Browser-Tab mit einem Dokument-Ranking, das wichtige Dokumente des Themas zeigt, z. B. Thema 48 zu Musik im mittleren Teil von Abbildung 2. Ein Dokument wird im Browser-Tab als Kasten dargestellt, der den Titel, den Textbeginn und mittels farbiger Kreise die vier dominanten Themen des Dokuments zeigt. Die farbige Referenz eines solchen Kreises kann genauer untersucht werden, wenn man mit der Maus darüberfährt. Durch Klicken auf einen Kreis kann das entsprechende Thema nachgeschlagen werden; es wird dann in der unteren Themenansicht mittig fokussiert.

Durch Klicken auf den Titel eines Dokumentes, z. B. auf „Queen (Band)“ in Abbildung 2, wird ein neuer Tab mit einer Dokument-Ansicht geöffnet (Abbildung 3), der den Inhalt des Dokuments zeigt, wobei die Wörter durch farbige Unterstreichungen die Themenzuordnungen anzeigen, die bei der automatischen Themenanalyse ermittelt wurden. Somit kann das Ergebnis der Themenanalyse direkt an den Dokumentinhalten verifiziert werden. Ebenso wie in der Dokument-Browser-Ansicht können die farbigen Themenzuordnungen durch Darüberfahren mit der Maus konkretisiert werden. Ein Klick auf ein farbiges Wort fokussiert das Thema in der Themenansicht.

Mit Klicken auf das Zeitreihen-Icon eines Themas (in der farbigen Box des Themas rechts oben) wird die Zeitansicht des TopicExplorers für dieses Thema in einem neuen Tab geöffnet. Ein Beispiel wird in Abbildung 4 für die Cäsium-Rindfleisch-Daten dargestellt, die Teil der japanischen Blog-Sammlung zum Stichwort „Kernkraft“ (*genpatsu*) sind. Zeitreihen zu weiteren Themen sowie die Durchschnittszeitreihe aller Themen können als Vergleich mittels Checkboxes

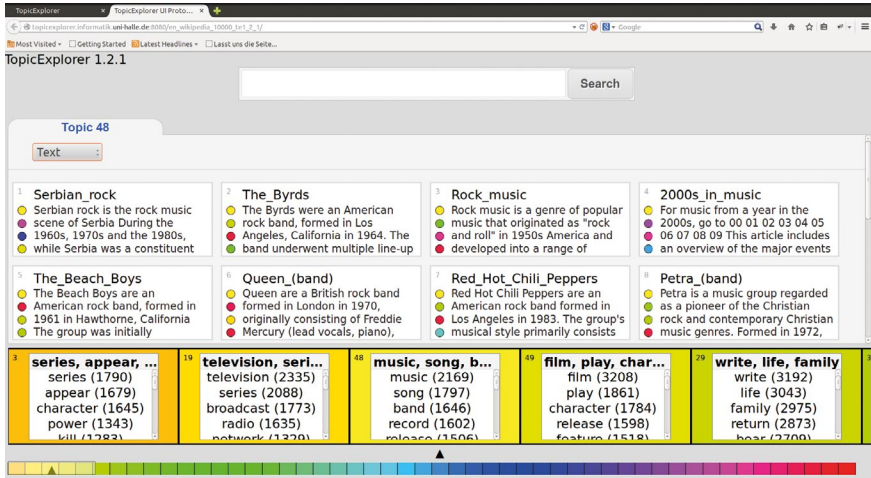


Abbildung 2: Überblick TopicExplorer mit Bedienelementen zur Suche, zum Dokument-Ranking und zur Themendarstellung und -navigation (von oben nach unten).

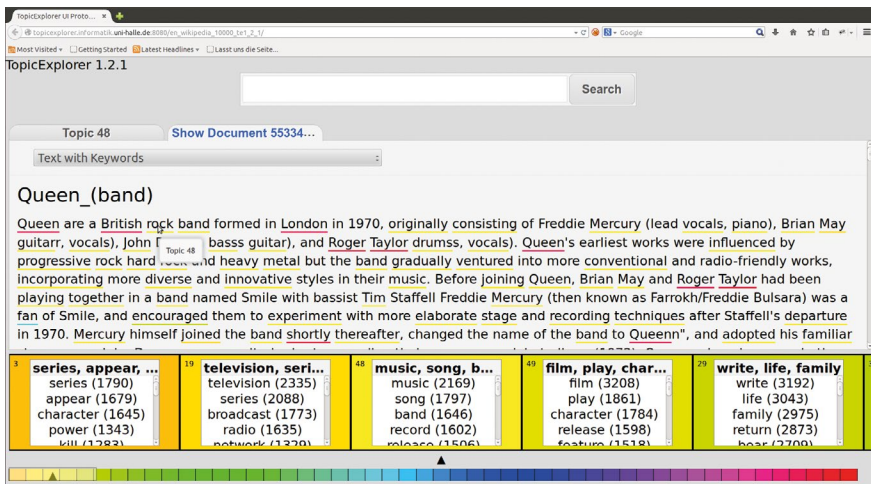


Abbildung 3: Dokumentansicht des TopicExplorer: Die farbigen Unterstreichungen der Wörter zeigen die Themenzuordnungen. Die nicht unterstrichenen Wörter wurden als Stopp-Wörter (sehr häufig) oder als zu seltene Wörter aus der Themenanalyse ausgeschlossen.

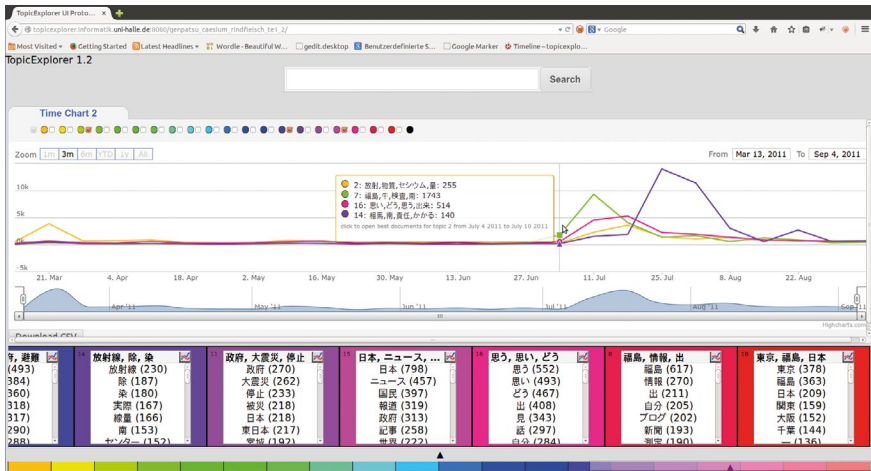


Abbildung 4: Zeitanzeige des TopicExplorer: Die Entwicklung mehrerer Themen über die Zeit wird anhand der Anzahl der ihnen zugeordneten Wörter durch die farbigen Zeitreihen dargestellt.

interaktiv eingblendet werden. Die Legende zeigt die wichtigsten Wörter der gewählten Themen für die Woche, über der sich die Maus befindet. Somit kann die Entwicklung der Themen interaktiv verfolgt werden. Ein Klick auf eine Woche in der Zeitreihe öffnet ein Dokument-Ranking in einem neuen Tab, das die wichtigsten Dokumente zu dem zuerst gewählten Thema in der mit der Maus gewählten Woche zeigt. Die thematischen Zeitreihen geben z. B. Auskunft, ob, wann und in welcher Abfolge Themen einen Höhepunkt erreichen.

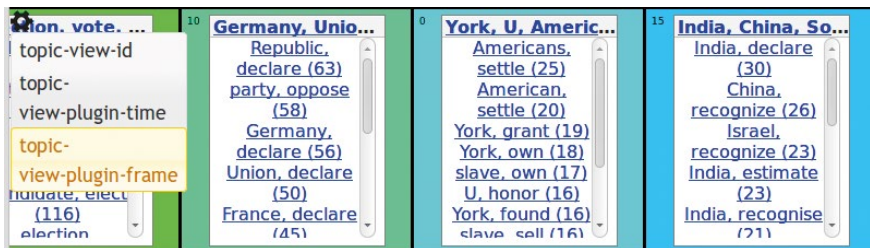
Eine zentrale Herausforderung bei der Gestaltung des TopicExplorers ist die Darstellung der Themen, die auf eine Weise erfolgen muss, dass Anwender die Themen interpretieren und von irrelevanten statistischen Artefakten unterscheiden können. Dafür schlagen wir vor, Themen durch Substantiv-Verb-Kombinationen (TopicFrames) darzustellen (Hinneburg et al. 2014). Die Repräsentation eines Themas durch häufige, themenspezifische Substantiv-Verb-Kombinationen hilft dem Benutzer, den Kontext der wahrscheinlichsten Wörter des Themas zu erfassen bzw. einzugrenzen und somit das Thema zutreffend zu interpretieren (Abbildung 5a). Besonders hilfreich sind TopicFrames, wenn die wahrscheinlichsten Wörter eines Themas ausschließlich aus Substantiven bestehen und das Thema aufgrund dieser Wortlisten mehrere Interpretationen haben könnte. Die Kombinationen mit Verben helfen, den Kontext genauer zu bestimmen. Ein TopicFrame besteht aus einem Substantiv und einem Verb, die im gleichen Satz vorkommen und demselben Thema zugeordnet sind. Abbildung 5b zeigt die häufigsten TopicFrames der oben eingeführten Themen. Das Asienthema (rechts)



Abbildung 5: Wechsel der Themenrepräsentation.



(5a) Wahrscheinlichste Wörter.

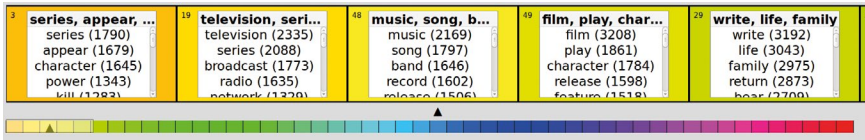


(5b) Häufigste TopicFrames.

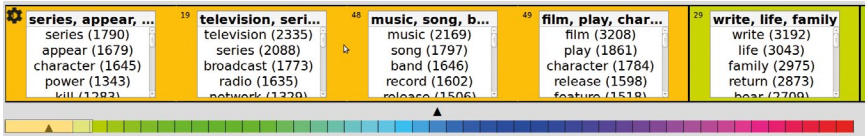
kann durch die Kombination von *declare* bzw. *recognize* mit Ländernamen als asiatische Außenpolitik erkannt werden.

Für das Berechnen der TopicFrame-Repräsentationen führt der TopicExplorer eine Wortartbestimmung („Part-Of-Speech-Tagging“) in der Vorverarbeitung durch. In weiteren Schritten wurde für den TopicExplorer das TopicFrame-Konzept dahingehend erweitert, dass neben Substantiv-Verb-Kombinationen weitere frei wählbare Wortarten für die Repräsentation der Themen kombiniert werden können. Diese können bis zu den Dokumenten zurückverfolgt werden, sodass Anwender leicht auf Belegstellen für TopicFrames zugreifen können. Die verschiedenen Repräsentationen der Themen lassen sich interaktiv umschalten, sodass Anwender verschiedene Sichten auf ein Thema erhalten. Mit der Bestimmung der Wortarten für jedes einzelne Wort in den Dokumenten ergibt sich die Möglichkeit, die Themen durch die wahrscheinlichsten Wörter einer bestimmten Wortart zu repräsentieren. Gerade weil die Wahrscheinlichkeiten für die unterschiedlichen Wortarten sehr variieren, ist es bei der Interpretation eines Themas oft hilfreich, eine weniger häufige Wortart in den Fokus zu rücken. So sind Adjektive und Adverbien im Vergleich zu Substantiven und Verben seltener. Durch die Repräsentation eines Themas durch Adjektive

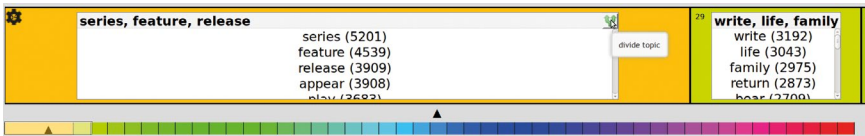
Abbildung 6: Interaktive, hierarchische Themennavigation.



(6a) Durch Themenanalyse berechnete Themen.



(6b) Vorschau des Zusammenfassens von Themen aufgrund der Hierarchie.



(6c) Zusammengefasstes Thema, interaktives Aufteilen ist möglich.

lässt sich außerdem erkennen, ob und in welcher Weise ein Thema emotional gefärbt ist.

Hierarchische Themen ermöglichen Anwenderinnen und Anwendern, die Themengrundlage an jeweils eigene Fragestellungen anzupassen. Dieses interaktive Definieren der Themen durch die Anwender wird durch Vorschauansichten, Ranking- und Dokumentansichten sowie Zeitdiagramme zur Themenentwicklung unterstützt, die sich alle den aktuell gewählten Themen anpassen. Somit sind die Anwender nicht nur passive Konsumenten der Themen, sondern können diese während des Explorationsprozesses selbst definieren.

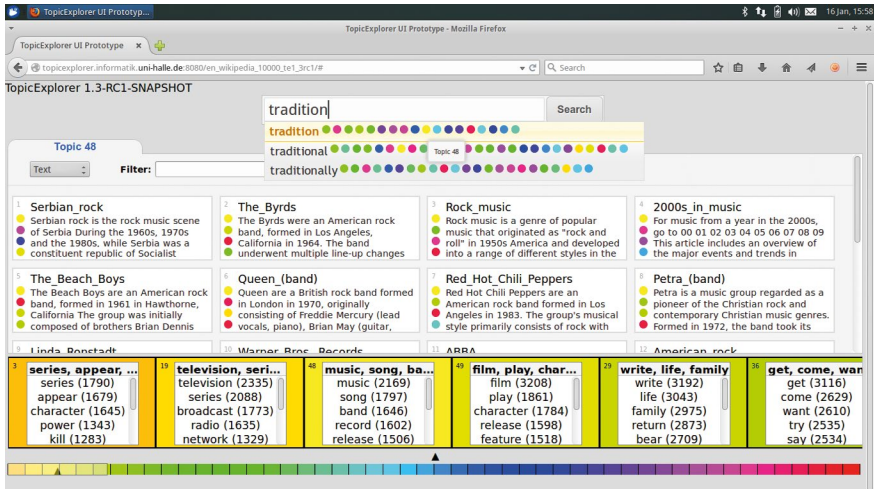
Um dies interaktiv zu ermöglichen, werden die bei der Themenanalyse berechneten Themen während der Vorverarbeitung mittels eines hierarchischen Clustering-Verfahrens aufgrund ihrer Ähnlichkeit in den Wortverteilungen zusammengefasst. Als Ergebnis des Clusterings entsteht ein binärer Baum, dessen Blätter die durch die Themenanalyse berechneten Themen sind. Für jedes Thema, das in der Clusteranalyse durch Zusammenfassen entsteht, werden in der Vorverarbeitung ebenfalls alle relevanten Informationen und TopicFrames berechnet. Die Anwender/innen können im TopicExplorer interaktiv in diesem binären Baum der Themen navigieren, indem sie die Maus auf die Trennlinie zwischen den farbigen Themenspalten positionieren. Dabei ändert sich als Vorschau

die Hintergrundfarbe der Themen, die durch einen Klick auf die Trennlinie zu einem einzigen Thema zusammengefasst würden. Es können zwei oder mehr Themen durch eine solche Interaktion zusammengefasst werden in Abhängigkeit davon, wie weit die beiden Themen rechts und links der Trennlinie im binären Baum auseinanderliegen. Abbildungen 6a und 6b zeigen durch den Wechsel der Themenfarbe an, dass die vier Themen von links über Serien, Fernsehen und Radio, Musik und Filme zu einem Thema über Unterhaltung zusammengefasst würden, wenn zwischen die beiden mittleren Themen geklickt würde. Abbildung 6c zeigt das zusammengefasste Thema. Das Zusammenfassen kann durch Klicken auf den Button rückgängig gemacht werden, der rechts oben bei dem zusammengefassten Thema erscheint. Die Information des Zusammenfassens wird sofort an alle anderen Ansichten des TopicExplorers (Dokument-Browser, Dokumentansicht, Zeitanzeige) weitergeleitet und dort dargestellt. Gerade in der Zeitanzeige können durch das Zusammenfassen von Themen zeitliche Häufungen sichtbar werden, die sonst in mehrere kleine Themen zerlegt sind. Somit wird die Auswirkung des Zusammenfassens interaktiv erfahrbar.

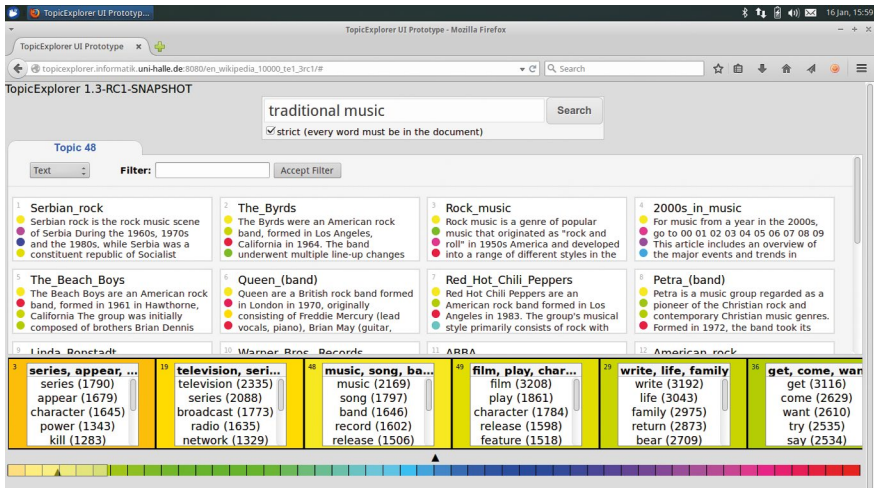
Das TopicExplorer-System bietet weiterhin Volltextsuche sowie Filterfunktionen, um Anwendern den Einstieg in ein Korpus anhand von konkreten Suchwörtern zu erleichtern. Es können einzelne oder mehrere Wörter im Verbund als Suchbedingung angegeben werden. Solange ein einzelnes Wort im Suchfeld steht, werden durch die Auto-Complete-Funktion die Themen mit absteigender Wahrscheinlichkeit angezeigt, die zu diesem Wort passen (siehe Abbildung 7a). So können interessante Themen ausgehend von diesem einen Wort rasch gefunden werden. Wenn mehrere Wörter eingegeben werden, kann zwischen normaler und strikter Suche gewählt werden (siehe Abbildung 7b). Bei strikter Suche müssen alle Wörter in einem Trefferdokument vorkommen, während die normale Suche nur sicherstellt, dass die Trefferdokumente möglichst viele, aber nicht unbedingt alle Wörter der Anfrage enthalten. In der Dokument-Browser-Ansicht wird für jedes Treffer-Dokument angezeigt, welche Wörter der Anfrage im Dokument wie oft vorkommen.

Die thematische Stichwortsuche findet Dokumente, die das gewünschte Stichwort enthalten, jedoch wird zusätzlich geprüft, ob das Wort im Trefferdokument auch dem gewünschten Thema zugeordnet ist. Mit dieser Suchmethode können auch mit allgemeinen Suchwörtern relevante Trefferdokumente gefunden werden, wenn das Thema entsprechend eingeschränkt wird. Beispielsweise findet die allgemeine Stichwortsuche nach *traditional* Dokumente aus sehr verschiedenen Bereichen. Mit einer thematischen Stichwortsuche nach *traditional* für das Thema Musik werden jedoch nur Dokumente zu Volks- und Regionalmusik gefunden, die *traditional* enthalten und dem Thema Musik zuzuordnen sind.

Abbildung 7: Stichwortsuche.



(7a) Autocomplete mit Themendarstellung.



(7b) Mehrere Stichwörter.

Ausgestattet mit den interaktiven Möglichkeiten des TopicExplorers können nicht technisch ausgebildete Anwender ohne Programmierkenntnisse Topic Models nutzen, um Inhalte in großen Dokumentensammlungen mit wenig Aufwand vorzustrukturieren. Weiterhin können die Effekte von Parametereinstellungen wie „Anzahl der Themen“, „Auswahl der analysierten Wortarten“ und „Weglassen von häufigen und seltenen Wörtern“ auf die Themen studiert werden, um deren semantische Qualität zu prüfen und gegebenenfalls zu verbessern.

### 2.3 Stand der Technik im Vergleich

In den letzten Jahren wuchs das Interesse an Werkzeugen, die automatische Text-Analyse und -Visualisierung miteinander verbinden. Zu diesem Thema wurden mehrere Workshops (z. B. Chuang et al. 2014) veranstaltet sowie einzelne Konferenzbeiträge veröffentlicht. Parallel zum TopicExplorer wurden daher international auch eine Reihe anderer Systeme entwickelt und vorgestellt. Um den TopicExplorer relativ zum Stand der Technik auf diesem Gebiet zu verorten, vergleichen wir ihn – trotz erheblicher Unterschiede in den anwendungsspezifischen Zielsetzungen – mit den anderen Systemen (Tabelle 1). Die Tabelle fasst die wichtigsten methodischen Ansätze zusammen, die sich auf Topic Models stützen, und stellt deren Eigenschaften dar.

Zu den international bekannten Systemen gehört beispielsweise das Serendip-System (Alexander et al. 2014), das Informatiker und Anglisten an der Universität Wisconsin-Madison, USA zur Exploration englischsprachiger Korpora entwickelt haben. Ähnliche weniger ausgebaute Ansätze sind LDAvis (Sievert & Shirley, 2014) und topic-explorer (Murdock & Allen, 2015), ein einfacher Themen-Browser mit gleichem Namen. Der Hallenser TopicExplorer bietet zu fast allen Möglichkeiten hinsichtlich Ranking und Themenrepräsentation ähnliche Funktionen wie Serendip an, doch basieren sie auf anderen Visualisierungs- und Interaktionstechniken. Darüber hinaus hält der TopicExplorer aber auch wichtige zusätzliche Funktionen vor, wie z. B. Themen interaktiv zusammenzufassen und ihren zeitlichen Verlauf darzustellen. Diese zusätzlichen Möglichkeiten, die von den Anwendern hoch geschätzt werden, sind auf dem internationalen Stand der Technik ausgeführt, dessen Aspekte im Folgenden kurz umrissen und mit den Systemen in Tabelle 1 verknüpft werden. Wesentlich sind die Aspekte: (1) maschinelles Lernen der zeitlichen Themenentwicklung, (2) deren Visualisierung, (3) die hierarchische Themengliederung und (4) die Software-Organisation.

Die zeitliche Themenentwicklung wird am einfachsten modelliert, indem die Themen zuerst ohne Berücksichtigung der Zeit gelernt werden und danach deren Entwicklung durch die gemeinsame Analyse der Zeitstempel der Dokumente und der Zuordnungen der Wörter zu den Themen charakterisiert wird. Dieser

Tabelle 1: Überblick der Visualisierungswerkzeuge für Topic Models mit ihren Eigenschaften.

System	Features																				
	Ranking				Representation						Hierarchy		Time		Search		Software				
	Document given Topic	term given topic	topic given document	topic given term	word list	POS	frames	token topic assignment	document groups	topic concentration	document authors	overview	interactive merge/split	overview	evolving relations	interactive zoom	keyword	topic	config. preprocessing	config. user-interface	
TopicExplorer (Hinneburg et al. 2012, Hinneburg et al. 2014)	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓			✓	✓		✓	✓	
Seredip (Alexander et al. 2014)	✓	✓			✓			✓	✓	✓											
LDavis (Sievvert & Shirley, 2014)	✓	✓			✓								✓								
topic-explorer (Murdock & Allen, 2015)	✓	✓			✓																
TopicPanorama (Liu et al. 2014)	✓	✓			✓								✓								
TIARA, (Liu et al. 2009, Pan et al. 2013)	✓	✓			✓							✓				✓					
EvoRiver (Sun et al. 2014)	✓	✓			✓							✓				✓					
RoseRiver (Cui et al. 2014)	✓	✓			✓							✓				✓	✓				
HierarchicalTopics (Dou et al. 2013)	✓	✓			✓							✓				✓	✓				
Dvita (Derntl et al. 2014)	✓	✓			✓							✓				✓		✓	✓		
Footprint (Isaacs et al. 2014)	✓	✓			✓							✓				✓		✓	✓		
HierarchicalTopicGrid (Jojić et al. 2016)	✓	✓			✓							✓				✓		✓	✓		

Ansatz wird von fast allen Visualisierungssystemen TopicExplorer, TIARA (Liu et al. 2009, Pan et al. 2013), EvoRiver (Sun et al. 2014), RoseRiver (Cui et al. 2014) in diesem Bereich verfolgt. Eine Ausnahme ist Dvita (Derntl et al. 2014), das ein deutlich rechenintensiveres Themenmodell nutzt und die zeitlichen Änderungen von Themenpräferenzen durch Wahrscheinlichkeitsverteilungen modelliert. Bisher zeigte jedoch keiner der beiden Ansätze (Lernen der Themen mit oder ohne Zeitinformation) besondere Vorteile für die Inhaltsanalyse. Neben einer Übersichtsdarstellung der Themenentwicklung und verschiedener Interaktionsmöglichkeiten, die alle genannten Systeme bieten, wurden in TIARA, EvoRiver und RoseRiver prototypisch die Entwicklungen von Beziehungen zwischen Themen visualisiert. Diese Möglichkeit soll ebenfalls in zukünftigen Versionen des TopicExplorers eingebaut werden, um neuartige Verbindungen zwischen Themen zu erkennen.

Die Visualisierungsansätze zur hierarchischen Themengliederung wurden in LDAvis, TopicPanorama (Liu et al. 2014) und HierarchicalTopicGrid (Jojic et al. 2016) prototypisch umgesetzt, um speziell entwickelte Visualisierungs- und Interaktionstechniken zu untersuchen. Die Kombination von Hierarchie und zeitlicher Entwicklung wurde in TopicExplorer, HierarchicalTopics (Dou et al. 2013) und RoseRiver umgesetzt, wobei HierarchicalTopics eine nur prototypische Studie ist. RoseRiver ist dagegen eine spezielle Visualisierung für wenige vorausgewählte Themen, die beide Aspekte verbindet und nach einer Evaluation ebenfalls im TopicExplorer umgesetzt werden kann.

Software-Design, Wartung, Erweiterbarkeit und Konfigurierbarkeit als letzter Aspekt wurden nur im TopicExplorer und Dvita explizit adressiert. Für die anderen Systeme wurde dieser Aspekt in den Veröffentlichungen nicht angesprochen.

Footprint (Isaacs et al. 2014), das interessante Funktionen im Bereich Suche bietet, fällt ansonsten aus dem Vergleichsrahmen heraus, weil es keine Topic Models benutzt. Stattdessen werden hier Dokumente durch einen externen Informationsservice mit Themen annotiert, der nur für Englisch verfügbar ist. Deshalb können dort keine Wortzuordnungen zu Themen gemacht werden und daher enthält das System keine weiteren darauf aufbauenden Funktionen.

Zusammenfassend kann festgestellt werden, dass der Hallenser TopicExplorer im Vergleich mit allen anderen Systemen einen großen Funktionsumfang bietet und alle in der Forschung auftretenden Aspekte der Visualisierung und Interaktion mit Topic Models abdeckt. Einzelne innovative Ideen für weitere Funktionalitäten können durch die erweiterbare Software-Architektur ergänzt werden.

### 3. Sozialwissenschaftliche Anwendung am Beispiel Fukushima

#### 3.1 Fukushima und neue Medien als Artikulationsraum

Als ein Beispiel für die Anwendung des TopicExplorers werden in diesem Aufsatz exemplarisch japanische Blogs über den Skandal des mit radioaktivem Cäsium verseuchten Rindfleisches, das einige Monate nach dem Atomunfall von Fukushima zunächst in Supermarktregalen in Tokio, später dann auch in anderen Teilen Japans auftauchte, untersucht. Im Falle dieses Skandals spielte das Internet eine zentrale Rolle. Denn die Gefahren, die von Radioaktivität ausgehen, sind für den Einzelnen in der realen Umwelt weder direkt erfahrbar noch sichtbar. Folglich ist das Bewusstsein für die Gefährlichkeit der Lage in der Gesellschaft zunächst nur latent vorhanden. Das Ausmaß der Bedrohung wird den Menschen erst durch gesellschaftliche Instanzen, insbesondere durch die Medien, vermittelt und somit sichtbar und „begreifbar“ gemacht. Die Menschen nutzen bei diesem, nach Latour (2004) auch als „Artikulation“ bezeichneten Vorgang in erheblichem Maße das Internet als „Artikulationsraum“ (Kuchinskaya 2011: 406). Denn durch die neuen Medien kann der Einzelne seine Ansichten, Einschätzungen und Informationen zu gesellschaftlich relevanten Themen im Internet formulieren und rasch verbreiten. So verschafft das Internet dem Einzelnen die Möglichkeit, im Austausch mit anderen ein Verständnis des Unglücksgeschehens zu entwickeln, das bis zu einem gewissen Grade unabhängig von den Interpretationen ist, die ihm die konventionellen Medien anbieten. Um die Rolle des Internets als Artikulationsraum einschätzen zu können, ist u. a. die Frage bedeutungsvoll, in welchem Umfang die Berichterstattung der konventionellen Medien die Inhalte z. B. von Blogs prägt und bis zu welchem Grad Blogs davon unabhängige Themen aufgreifen. Im folgenden Abschnitt werden der Skandal und seine Darstellung in japanischen Blogs anhand eines Blog-Korpus und seiner automatisierten Inhaltsanalyse mit dem TopicExplorer beschrieben. Einem Vergleich der so erzielten Ergebnisse mit den Inhalten der Printmedien folgt eine kurze Diskussion und Zusammenfassung.

#### 3.2 Der Cäsium-Skandal nach Fukushima in der TopicExplorer-Analyse von japanischen Blogs

Nach dem Großen Ostjapan-Erdbeben vom 11. März 2011 wurden aus dem Kernkraftwerk Fukushima-1 große Mengen radioaktiver Substanzen in die Umwelt freigesetzt, die nicht nur die nähere Umgebung, sondern stellenweise auch entfernte Landesteile Japans verseuchten. Die japanische Regierung tat sich schwer damit, die radioaktive Kontamination einzudämmen und die Sicherheit von



Lebensmitteln, besonders aus der direkt betroffenen Region Tohoku, sicherzustellen. Politiker und Beamte ergriffen nur zögerlich Maßnahmen, um die Verseuchung von Japans Nahrungsmittelversorgung zu verhindern. Unter anderem versäumten sie es, den Vertrieb von Reisstroh aus den kontaminierten Regionen wirkungsvoll zu verbieten. Dieses Reisstroh wurde folglich an Züchter im ganzen Land als Rinderfutter verkauft, sodass auch Rinderbestände außerhalb des ursprünglich betroffenen Gebiets mit radioaktivem Cäsium verseucht wurden. Außerdem verließ sich die japanische Regierung zunächst auf freiwillige Vertriebssperren für Rindfleisch aus radioaktiv gefährdeten Gebieten, nur um dann im Juli 2011 aus Berichten von Kommunen und Medien zu erfahren, dass Rindfleisch mit einem erhöhten Gehalt von radioaktivem Cäsium in den Regalen von Tokioter Supermärkten auslag. Erst bei Bekanntwerden dieses Skandals reagierte die Regierung mit einem Vertriebsverbot für Rindfleisch aus der betroffenen Region. Diese Nachlässigkeiten schädeten der Gesundheit der Konsumenten und störten das Vertrauen, das über Jahre in hochpreisiges japanisches Rindfleisch aufgebaut worden war (Kingston 2012).

Für die TopicExplorer-Studie wurden Blog-Einträge in japanischer Sprache aus dem Internet maschinell heruntergeladen, die drei Stichworte, nämlich „Rind“, „Cäsium“ und „Atomkraft“ (*genpatsu*), enthalten. Letzteres Stichwort wurde hinzugefügt, um den thematischen Fokus der Texte sicherzustellen. Für den Untersuchungszeitraum von März bis September 2011 konnten so 1931 Texte mit einer durchschnittlichen Länge von 2402 Worten gesammelt und anonymisiert werden. Das Korpus von Blog-Texten wurde nach linguistischer Vorverarbeitung dann einer automatischen Inhaltsanalyse durch den TopicExplorer unterzogen. Die so ermittelten Themen sind durch eine Liste ihnen zugeordneter Worte charakterisiert, die vom TopicExplorer noch durch eine Aufstellung der häufigsten im jeweiligen Thema vorkommenden Kollokationen ergänzt wird (Tabelle 2). Die anhand von Stichworten und Kollokationen interpretierten Themen, denen je ein Thementitel zugeordnet wird, können dann anhand der Textpassagen, in denen die Themen behandelt werden, überprüft werden. Dadurch lässt sich die Plausibilität der Themenanalyse direkt an den Blog-Inhalten verifizieren, und die wichtigsten Textpassagen zu einem bestimmten Thema können schnell aufgefunden werden. In der separaten Zeitanzeige (Abbildung 8) kann die Entwicklung eines Themas im Untersuchungszeitraum oder einem Ausschnitt davon interaktiv verfolgt werden, z. B. auch im Vergleich zu Zeitreihen weiterer Themen oder zur Durchschnittszeitreihe aller Themen.

Die zwanzig Themen, die der Algorithmus im Blog-Korpus zum Cäsium-Skandal gebildet hat, sind in Tabelle 2 in der Reihenfolge aufgeführt, in der sie vom TopicExplorer arrangiert werden. In den Spalten sind nach der Nummer des jeweiligen Themas die Spitzeneinträge aus den Listen der häufigsten Stichworte

Tabelle 2: Zwanzig Themen im Blog-Korpus

Thema	2	5	9	7	12
Charakt. Worte	<p>2</p> <p>放射, 物質, セシ... 放射 (800) 物質 (665) セシウム (456) 素 (333) ヨウ (329) 放射線 (327) 影響 (277) 半減 (208) 体内 (201)</p>	<p>5</p> <p>放射, セシウム, ... 放射 (676) セシウム (615) ベクレル (588) 検出する (578) 基準 (503) 断定 (418) 物質 (363) 超える (325) 当たり (248)</p>	<p>9</p> <p>稲, わら, セシウム 稲 (714) わら (682) セシウム (586) 農家 (585) 牛 (542) 与え (411) 肉牛 (365) 汚染 (327) 産卵 (309)</p>	<p>7</p> <p>福島, 牛, 出荷 福島 (702) 牛 (682) 出荷 (502) 検査 (488) 検出する (457) 出荷する (453) 肉 (434) 農家 (381) セシウム (367)</p>	<p>12</p> <p>牛, セシウム, 食べ 牛 (981) セシウム (979) 食べ (794) 牛肉 (674) 被害 (485) 汚染する (412) 肉 (373) 風評 (316) 汚染 (308)</p>
Charakt. Kollokationen	<p>2</p> <p>放射, 物質, セシ... 体内, 取り込ま (30) 大気, 放出する (27) 物質, 拡散する (21) 物質, 飛散する (21) 物質, 放出する (21) 物質, 吸収する (20) 物質, 付着する (19) 骨, 蓄積する (19) 物質, 含ま (18)</p>	<p>5</p> <p>放射, セシウム, ... セシウム, 検出する (133) ベクレル, 検出する (128) 放射, 検出する (119) 基準, 超える (97) ベクレル, 超える (81) 基準, 超え (47) 物質, 検出する (44)</p>	<p>9</p> <p>稲, わら, セシウム わら, 与え (162) 肉牛, 与え (58) 農家, 出荷する (45) セシウム, 汚染する (38) 飼料, 与え (35) 肉牛, 出荷する (32) 牛, 与え (32)</p>	<p>7</p> <p>福島, 牛, 出荷 セシウム, 検出する (86) 農家, 出荷する (80) 相馬, 出荷する (71) 放射, 検出する (70) 規制, 超える (60) 牛, 出荷する (53) 断定, 超える (41) 区域, 出荷する (36)</p>	<p>12</p> <p>牛, セシウム, 食べ 牛肉, 食べ (103) 牛, 食べ (82) セシウム, 汚染する (67) セシウム, 食べ (59) 肉, 食べ (55) ワラ, 食べ (32) 牛肉, 流通する (30) 悪, 食べ (26)</p>
Thementitel: Charakt. Worte Charakt. Kollokationen (u. a.)	<p>Ausbreitung u. Aufnahme radioaktiver Substanzen: Strahlung, Substanz, Cäsium (456), Element, Jod; in den Körper aufgenommen werden, [in die] Atmosphäre freisetzen, Substanz verstreuen</p>	<p>Nachweis von Cäsium: Strahlung, Cäsium (615), Becquerel, feststellen, Grundlage; Cäsium feststellen, Becquerel feststellen, Strahlung feststellen</p>	<p>Ursache: Reis, Stroh, Cäsium (586), Landwirt, Rind (542); Stroh geben, Schlachtvieh [Futter] geben, Futter geben, Landwirte verschicken [Stroh]</p>	<p>Entdeckung/ Kontrolle von Cäsium-verseuchtem Rindfleisch: Fukushima, Rind (682), Lieferung, Untersuchung, Cäsium (367); Landwirte liefern [Fleisch], Strahlung feststellen</p>	<p>Schäden durch Cäsium-verseuchtes Rindfleisch: Rind (981), Cäsium (979), essen, Rindfleisch, Schaden, verseuchen; Rindfleisch essen, Rind essen, [mit] Cäsium verseuchen</p>

Thema	17	0	10	13	4
Charakt. Worte	<p>17 牛, 農家, 畜産</p> <p>牛 (939) 農家 (237) 畜産 (233) 口 (222) 生産 (188) 事故 (184) 家畜 (184) 消費 (182) 豚 (180)</p>	<p>0 原発, 菅, 東電</p> <p>原発 (502) 菅 (344) 東電 (342) 責任 (313) 政治 (287) 首相 (275) 国民 (272) 政府 (240) 民主党 (197)</p>	<p>10 原発, 事故, 汚染</p> <p>原発 (1305) 事故 (570) 汚染 (403) でき (399) 考え (237) 自然 (189) 地域 (175) 環境 (171) 経路 (170)</p>	<p>13 原発, 福島, 事故</p> <p>原発 (1033) 福島 (670) 事故 (607) 原子力 (316) 安全 (304) 発電 (289) 東電 (265) 核 (207) 原子 (186)</p>	<p>4 国, 検査, 調査</p> <p>国 (419) 検査 (410) 調査 (330) 対応 (279) 対策 (279) 受け (263) 不安 (248) 安全 (228) 対象 (226)</p>
Charakt. Kollokationen	<p>17 牛, 農家, 畜産</p> <p>牛, 殺 (17) 家畜, 殺 (12) 豚, 殺 (10) 豚, 殺 (9) 牛, 処分する (8) 牛, 飼育する (8) 事故, 伴う (7) 疫, 発生する (7) 豚, 飼育する (7)</p>	<p>0 原発, 菅, 東電</p> <p>責任, 問わ (13) 総理, 辞め (10) 首相, 辞め (8) 原発, 推進する (7) 責任, とる (6) 東電, 認め (5) 首相, 示し (5) 社会, 目指す (4) 国, 示さ (4)</p>	<p>10 原発, 事故, 汚染</p> <p>原発, 離れ (46) 原発, 考え (11) 原発, でき (10) 想像, でき (9) 原発, 推進する (7) 事故, でき (7) 原発, 得 (6) 原発, 流れ (6) 事故, 起こし (6)</p>	<p>13 原発, 福島, 事故</p> <p>運転, 停止する (11) 事故, 受け (8) 燃料, 使う (7) 原発, 推進する (7) 事故, 起こし (7) 事故, 起き (6) 原発, 決定する (6) 原発, 停止する (6) 再臨, 許さ (5)</p>	<p>4 国, 検査, 調査</p> <p>検査, 実施する (27) 調査, 実施する (17) 方針, 示し (17) 検査, 行う (14) 対応, 求める (10) 検査, 求める (9) 検査, 決め (8) 検査, 受け (8) 車豆, 電力, 受け (8)</p>
Thementitel:	<b>Probleme der Rinderzüchter:</b>	<b>Verantwortung für Unfälle folgen:</b>	<b>Zukunft der Atomkraft in Japan:</b>	<b>Atomkraft-wiederbetrieb und -aufnahme:</b>	<b>Forderungen:</b>
Charakt. Worte (u. a.):	Rind (939), Landwirt,	Atomkraftwerk, [Premiermin.] Kan, TEPCO,	Atomkraftwerk, Unfall, Verseuchung;	Atomkraftwerk, Fukushima, Unfall;	Staat, Messung/
Charakt. Kollokationen (u. a.)	Viehzucht; Maul, Cäsium (82); Rinder/Haustiere/ Schweine töten, Seuche tötet [Tiere], Rinder züchten, [mit dem Atom] Unfall einhergehen, Seuche auslösen	Verantwortung prüfen, [als] Premierminister zurücktreten, [als] Regierungschef zurücktreten	[von] Atomkraftwerk entfernen, [über] Atomkraftwerk nachdenken, Atomkraftwerk [nicht kontrollieren] können	Atomkraftwerk, Fukushima, Unfall; [Kraftwerks]Betrieb anhalten, Unfall erleiden, Brennstoff einsetzen	Untersuchung, Maßnahme, Rind (23) Untersuchungen durchführen, Plan vorlegen, Maßnahmen fordern

Thema	1	19	3	6	14
Charakt. Worte	<p>1 汚染, 放射能, 食品 汚染 (961) 放射能 (781) 食品 (461) 基準 (440) 汚染する (431) 安全 (431) 野菜 (320) とうもろこし (248) 米 (211)</p>	<p>19 放射能, 影響, 事故 放射能 (514) 影響 (399) 事故 (356) チェルノブイリ (324) 健康 (253) レベル (185) データ (183) 科学 (177)</p>	<p>3 子供, 福島, 被曝 子供 (383) 福島 (292) 被曝 (280) 学校 (211) 子ども (209) 先生 (188) 内部 (186) 放射線 (186) 言う (178)</p>	<p>6 放射線, 政府, 避難 放射線 (493) 政府 (384) 避難 (360) 福島 (318) 内部 (317) 被曝 (290) 原発 (288) 総量 (248) 住民 (187)</p>	<p>14 放射線, 除, 染 放射線 (230) 除 (187) 染 (180) 実際 (167) 総量 (166) 南 (153) センター (152) 内部 (144) 子ども (139)</p>
Charakt. Kollokationen	<p>2 汚染, 放射能, 食品 放射能汚染する (83) 汚染汚染する (26) 基準決める (24) 汚染発覚する (16) 基準超える (16) 食品流通する (12) 汚染流通する (11) 魚汚染する (9) 食守る (8)</p>	<p>19 放射能, 影響, 事故 放射能 及び (11) 研究発表する (11) 影響受け (11) 放射能汚染する (11) 事故起き (9) チェルノブイリ起き (9) データ公開する (9) 影響評価する (8)</p>	<p>3 子供, 福島, 被曝 子供守る (32) 子供被曝する (25) 給食使われ (23) 子供連れ (22) 先生おら (22) 子供心配する (20) 子供汚染する (20) 子供でき (19) 小学校言わ (19)</p>	<p>6 放射線, 政府, 避難 区域指定する (15) ミリンヘルト超え (14) 放射線浴び (13) 住民避難する (11) 原発超え (10) 時点超え (9) がん認め (9) 放射線測定する (8)</p>	<p>14 放射線, 除, 染 原爆相当する (48) 総量降下する (43) 原爆低下する (42) 状態認識する (41) 尿糞まじり (40) 粒子放出する (40) 日本用い (40) 総量かかる (39) 東京陸上 (39)</p>
Thementitel: Charakt. Worte (u. a.); Charakt. Kollokationen (u. a.)	<p><b>Verseuchung von Lebensmitteln:</b> Verseuchung, Radioaktivität, Lebensmittel, Cäsium (90); [mit] Radioaktivität verstrahlen, verseuchen, [Mess]Grundlage festlegen</p>	<p><b>Erfahrung aus Tschernobyl:</b> Radioaktivität, Unfall, Einfluss, Tschernobyl, Gesundheit; Radioaktivität ausgesetzt sein, Forschung veröffentlichen, Einfluss erleiden</p>	<p><b>Schäden durch radioaktives Cäsium bei Kindern:</b> Kinder, Fukushima, Strahlenschäden, Schule, Cäsium (129); Kinder schützen, Kinder erleiden Strahlenschäden, bei der Schulspeisung [verstrahlte Lebensmittel] einsetzen,</p>	<p><b>Messwerte radioaktiver Strahlung und Gefahr eines Aufenthalts:</b> Radioaktive Strahlung, Regierung, Flucht; [Sperr]Gebiet festlegen, [Werte von X] Milli-Siebert übersteigen, radioaktiver Strahlung ausgesetzt sein, Einwohner flüchten</p>	<p><b>Radioaktivität und Dekontamination:</b> Radioaktive Strahlung, Dekontamination, Strahlenmenge, Cäsium (63); [so und so vielen] Atombomben entsprechen, Strahlungs-menge niedergehen, Atombomben [Strahlung] reduziert [sich]</p>

Thema	11	15	16	8	18
Charakt. Worte	<p>11 政府, 大震災, 停止 政府 (270) 大震災 (262) 停止 (233) 日本 (218) 被災 (218) 東日本 (217) 宮城 (192) 福島 (186) 出発 (182)</p>	<p>15 日本, ニュース, ... 日本 (798) ニュース (457) 国民 (397) 報道 (319) 政府 (313) 記事 (258) 世界 (222) セブン (215) テレビ (205)</p>	<p>16 思う, 思い, どう 思う (552) 思い (493) どう (467) 出 (408) 見 (343) 話 (297) 自分 (284) 感じ (246) 本当 (1739)</p>	<p>8 福島, 情報, 出 福島 (617) 情報 (270) 出 (211) 自分 (205) ブログ (202) 新聞 (193) 測定 (190) でき (178) 記事 (165)</p>	<p>18 東京, 福島, 日本 東京 (378) 福島 (363) 日本 (209) 関東 (159) 大阪 (152) 千葉 (144) 一 (136) 原発 (131) 国 (130)</p>
Charakt. Kollokationen	<p>11 政府, 大震災, 停止 停止, 解除する (13) 災害, 基づき (11) 制限, 指示する (9) 停止, 指示する (9) 大震災, 伴う (8) 出発, 指示する (8) 対策, 基づき (8) 制限, 解除する (6) 復原, 向け (6)</p>	<p>15 日本, ニュース, ... 編集, 放送する (19) ニュース, 見 (14) マスコミ, 報道する (13) 日本, 言う (12) テレビ, 放送する (11) 日本, 見 (10) ニュース, 報道する (10)</p>	<p>16 思う, 思い, どう 話, 聞く (16) 自分, 思う (13) 話, 出 (13) 外, 出 (11) 話, 聞く (10) 話, 思う (8) 人間, 思い (7) 言葉, 出 (7) 皆さん, 思い (7)</p>	<p>8 福島, 情報, 出 測定, 測定する (11) 福島, 住む (11) 市内, 住む (10) 自分, 測定する (8) 理由, 挙げ (7) 新聞, 強調する (7) 記事, 紹介する (7) 高校, 続け (7) 測定, 出 (7)</p>	<p>18 東京, 福島, 日本 日本, 来 (3) 大阪, 住ん (3) 福島, 出 (3) 日本, 来る (3) 京都, 使う (3) 大阪, 来 (3) 福島, 買う (2) 関東, 被爆する (2) 日本人, 死ぬ (2)</p>
Themen- titel:	<p>Maßnahmen nach Atomunfall, u. a. Rind- fleischlieferstopp:</p>	<p>Kommentare/Zsfg. zur Medienberichter- stattung:</p>	<p>Persönliche Wahrneh- mung:</p>	<p>Selbstmessungen von Radioaktivität:</p>	<p>Atomkraft- werkunfall u. seine Folgen:</p>
Charakt. Worte (u. a.): Charakt. Kollokationen (u. a.)	<p>Regierung, Großes Erd- beben, Stopp, Erleiden einer Katastrophe; Stopp aufheben, Katas- trophe [...] basieren auf, Begrenzung anordnen, Stopp anordnen</p>	<p>Japan, Nachrichten, japanisches Volk, Berichterstattung; Cäsium (215); Edition senden, Nach- richten sehen, Massen- medien berichten</p>	<p>Denken, sehen, sagen, selbst, fühlen; Erzählung hören, selbst denken, Erzählung entstehen</p>	<p>Fukushima, Information, selbst, Blog, Zeitung, Messung; Messung, durchführen, [in] Fuku- shima wohnen, [in der] Stadt wohnen, selbst messen</p>	<p>Tokio, Fukushima, Japan, Kantō, Osaka; [nach] Japan kommen, [in] Osaka wohnen, Fukushima verlassen</p>

und Kollokationen für das jeweilige Thema in japanischer Sprache entsprechend der jeweiligen TopicExplorer-Ansicht wiedergegeben. Diese Listen haben oft einen Gesamtumfang von einigen Hundert Einträgen und werden von der Analysesoftware für jedes Thema erstellt. In der folgenden Zeile ist eine deutsche Benennung des Themas (=Thementitel) eingetragen, die durch Interpretation der Stichwortliste, der Kollokationen und der repräsentativen Texte des Themas gewonnen wurde. Es folgen die deutschen Übersetzungen ausgewählter Stichworte und Kollokationen. Bei den Themen, die u. a. auch durch die Stichworte „Cäsium“ oder „Rind“ charakterisiert sind, ist zusätzlich die Häufigkeit dieser Stichworte vermerkt. Beispielsweise sind dem Thema 9 „Ursache“, „Cäsium“ und „Rind“ 586-mal bzw. 542-mal zugeordnet.

Die Themen 7, 9, 12 und 17 sind am engsten mit dem eigentlichen Skandalgeschehen verbunden. Alle vier Themen zeichnen sich durch häufiges Auftreten der Worte „Cäsium“ und insbesondere „Rind“ aus. Bei Thema 7 weisen die Stichworte und besonders die typischen Kollokationen darauf hin, dass es die Entdeckung und nachfolgende Überwachung von Cäsium-verseuchtem Rindfleisch zum Gegenstand hat. Die genauere Durchsicht der repräsentativen Textstellen bestätigt, dass hier die Aufdeckung des Skandals, dass von radioaktivem Cäsium verseuchtes Rindfleisch aus Fukushima zu Lebensmitteln verarbeitet worden war, thematisiert wird.

Unter den häufigsten Stichworten zu Thema 9 findet sich „Reis-Stroh“ an oberster Stelle. Die für dieses Thema repräsentativsten Texte beschreiben die bereits früh erkannte Ursache des Skandals, nämlich dass Reis-Stroh im Freien gelagert worden und dort radioaktivem Fallout ausgesetzt war, bevor es dann an Rinder verfüttert wurde. In Thema 12 sind Textpassagen zusammengefasst, die über die Folgen bzw. Schäden, die der Konsum von radioaktiv verseuchtem Rindfleisch mit sich bringen kann, berichten. Ebenfalls auf den Cäsium-Skandal im engeren Sinne nehmen auch die typischen Textstellen des Themas 17 Bezug, die die Probleme und das Vorgehen der Rinderzüchter behandeln, die teilweise schon vor der Nuklearkatastrophe mit der Maul-und-Klauenseuche zu kämpfen hatten und nun erneut erhebliche Verluste ihrer Bestände hinnehmen müssen.

Unter den Themen 4 und 11 sind Textstellen gebündelt, die die Maßnahmen der Regierung und die Reaktionen darauf wiedergeben. Die typischen Kollokationen zu Thema 4, dem auch das Wort „Rind“ an 23 Stellen zugeordnet ist, lauten „Untersuchungen/Nachforschungen durchführen“, „Plan vorlegen“ und „Maßnahmen fordern“. Hier wird u. a. die Lebensmittelkontrolle der Rindfleischbestände und die Forderung nach ihrer Verschärfung, um die Konsumenten zu schützen, Vertrauen wiederherzustellen und dadurch die Auslieferung von Fleisch wieder zu ermöglichen, angesprochen. Unter Thema 11 taucht das Stichwort „Rind“ zwar nicht auf, aber wie die Stichworte „Regierung“, „anhalten“ und „Katastrophe erleiden“ sowie die Kollokationen „Stopp aufheben“, „Begrenzung

anordnen“ und „Stopp anordnen“ zeigen, thematisieren die hier gesammelten Blog-Auszüge die Maßnahmen im Gefolge der Reaktorkatastrophe, zu denen u. a. auch der Auslieferstopp für bestimmte Lebensmittel gehörte.

Die sechs Themen 1, 2, 3, 5, 14 und 15 haben das Auftreten von „Cäsium“ unter den charakteristischen Stichworten gemeinsam, ohne dass das Wort „Rind“ präsent wäre. Thema 1 zu den Auswirkungen auf „Lebensmittel“ wird u. a. charakterisiert durch die Stichworte „Verseuchung“, „Radioaktivität“, „Lebensmittel“, „Grenzwert“, „verseuchen“, „Sicherheit“, „Gemüse“ und, mit etwas geringerer Häufigkeit, „Cäsium“. Die häufigste Kollokation für dieses Thema lautet „[mit] Radioaktivität verstrahlen“. Die charakteristischen Einträge für Thema 1 illustrieren die Sorgen der Blogger, die in Folge des Rindfleischskandals die radioaktive Verseuchung der Lebensmittelversorgung als persönliche Bedrohung erkennen. Beispielsweise kommentiert ein Blogger am 20. Juli 2011 ausdrücklich: „[Man muss] den eigenen Körper selbst schützen. So ein Zeitalter ist angebrochen“ (*jibun no mi wo jibun de mamoru. so iu jidai ni natta*).

Die Detailanalyse repräsentativer Textpassagen ergibt für Thema 1 außerdem, dass hier vielfach Medienauftritte von Koide Hiroaki (\*1949) wiedergegeben werden. Insgesamt wird Koide in 108 Textstellen mit Bezug zu Thema 1 erwähnt. Koide, ein auf Nukleartechnik spezialisierter Ingenieur, war Assistenzprofessor am Kyoto University Research Reactor Institute (KURRI). Er ist einer der wenigen japanischen Nuklearspezialisten, die schon frühzeitig Japans Ausstieg aus der Kernenergie gefordert haben – eine Unabhängigkeit, für die er mit der Stagnation seiner Universitätskarriere auf der Stufe des „ewigen Assistenzprofessors“ (*mannen no jokyō*) bis zu seiner Pensionierung 2015 bezahlte. Am 23. Mai 2011 sagte Koide als Experte vor einem Ausschuss des japanischen Parlaments aus. Obwohl an diesem Tag insgesamt vier prominente Kritiker der Atomenergie angehört wurden, übertrug das japanische Fernsehen, das sonst regelmäßig *live* aus dem Parlament sendete, diese Sitzung nicht, und auch die Zeitungen berichteten darüber nur mit kurzen Zusammenfassungen. Im Blog-Korpus dagegen waren erste Berichte schon am Folgetag der Ausschusssitzung zu finden, die auch danach noch oft zitiert wurden. Besonders häufig wurde Koides Aussage wiedergegeben, dass es auch unter seinen Kollegen viele gebe, denen gesagt worden sei, dass sie alarmierende Messergebnisse nicht veröffentlichen sollten (*watashi no dōryō demo, kenshutsu shita dēta wo kōhyō shinai you iwareta hitotachi ga nannin mo imasu*), um „Panik“ zu vermeiden.

Thema 15 umfasst Passagen aus 215 Texten, die das Wort „Cäsium“ enthalten, und bündelt Kommentare und Zusammenfassungen zur japanischen Medienberichterstattung über Fukushima und seine Folgen, wie die charakteristischen Stichworte „Japan“, „Nachrichten“, „japanisches Volk“ und „Berichterstattung“ sowie die Kollokationen „Freiheit [von] Berichterstattung“, „Edition senden“, „Nachrichten sehen“ und „Massenmedien berichten“ erkennen lassen. Thema 14

schließlich nimmt nur an 63 Stellen direkten Bezug auf Cäsium. Wie die Stichworte „radioaktive Strahlung“, „Dekontamination“ und „Strahlenmenge“ sowie die Kollokationen „[so und so vielen] Atombomben entsprechen“, „Strahlungsmenge niedergehen“ und „Atombomben[strahlung] reduziert [sich]“ vermuten lassen, behandeln diese Textstellen die dramatische Menge freigesetzter Radioaktivität und die dadurch notwendig gewordenen Dekontaminationen.

Die übrigen Einzelthemen wurden ohne das Stichwort „Cäsium“ gebildet. Sie behandeln zentrale Aspekte des Kernkraftwerkunfalls, die teilweise aber auch mit dem Cäsium-Rindfleisch-Skandal in Verbindung gebracht werden. Beispielsweise sind unter Thema 0, das u. a. durch die Wörter „Atomkraft“, „[Premierminister] Kan“, „TEPCO“, „Verantwortung“ und „Politik“ charakterisiert wird, Fragen nach den Konsequenzen und nach der Verantwortung u. a. für den konkreten Umgang mit dem Kraftwerksunfall erfasst. Darauf weisen auch die am häufigsten verwendeten Kollokationen hin wie z. B. „Verantwortung prüfen“, „Premierminister zurücktreten“ und „Premierminister [Amt] aufgeben“. Dabei werden auch in Verbindung mit dem Cäsium-Skandal Anschuldigungen gegen die damals amtierende Regierung erhoben. So behauptete ein Blogger am 21. Juli 2011, „die Ausbreitung Cäsium-verseuchter Rinder ist ein Verbrechen der Kan-Regierung“ (*seshiumu osen gyu no kakudai ha Kan seiken no hanzai da*). Auf die übrigen Themen wird aus Platzgründen hier nicht näher eingegangen.

### 3.3 Die zeitliche Entwicklung der Themen und der Einfluss einzelner Akteure

Der zeitliche Verlauf der Einzelthemen unterstützt ihre aus Stichworten, Kollokationen und Beispieltexten gewonnene Interpretation. Abbildung 8 zeigt den zeitlichen Verlauf aller Themen, die in den Blogs zum Skandal um das mit radioaktivem Cäsium verseuchte Rindfleisch identifiziert wurden. Insgesamt markieren die Themen 7 „Entdeckung“ und 9 „Ursache“ deutlich das Bekanntwerden des Cäsium-Skandals in der zweiten Juliwoche. Die früheste Spitze erreicht naturgemäß Thema 7, da es die Entdeckung des Skandals zum Gegenstand hat. Neun der zehn repräsentativsten Textstellen dieses Themas datieren auf den 9. Juli bzw. die Tage unmittelbar danach, als die Medien erstmals über den Skandal berichteten. Auch das Thema 5 „Nachweis...“, das die Bestimmung der Cäsiumwerte beschreibt und insofern mit Thema 7 eng verwandt ist, erreicht in dieser Woche seinen Spitzenwert. Die für Thema 9 „Ursache“ repräsentativsten Texte datieren hingegen überwiegend auf die zweite Juli-Hälfte, als man Reis-Stroh als „Überträger“ von radioaktivem Cäsium auf Rinder erkannte. Es überrascht nicht, dass die Zahl der Textstellen von Thema 9 daraufhin besonders stark abnimmt: Nachdem die Ursache des Skandals erst einmal gefunden worden war, stieß sie bald weniger auf Interesse als ihre Folgen. Folglich treten mit Thema 12



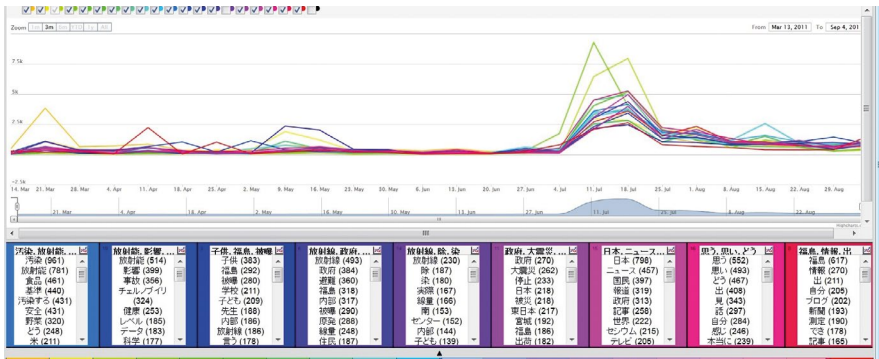


Abbildung 8: 19 Themen im Zeitverlauf (ohne Thema 14).

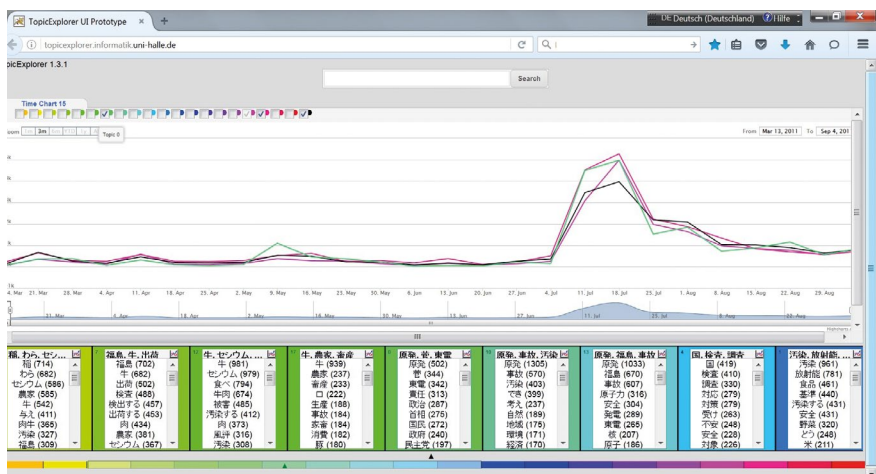


Abbildung 9: Themen 0, 15 und 16 im Vergleich zum Durchschnitt Zeitverlauf.

„Schäden...“ schon in der zweiten Juli-Hälfte Textpassagen, die von den Folgen bzw. Schäden, die der Konsum von radioaktiv verseuchtem Rindfleisch mit sich bringen kann, berichten, besonders häufig auf.

Mit dem Bekanntwerden des Skandals weisen alle übrigen Themen der Blogs einen Anstieg auf, der aber weniger prononciert ausfällt als bei den o.g. vier Themen und in seiner Größenordnung weitgehend dem Durchschnitt der Themen im Gesamtkorpus entspricht. Die Themen 0, 15 und 16 treten dabei noch am stärksten, nämlich überdurchschnittlich, hervor (Abbildung 9). Die Zunahme von Thema 16 „persönliche Wahrnehmung“, das persönliche Eindrücke zusammenfasst, lässt die große emotionale Anteilnahme am Geschehen vermuten. In der darauffolgenden Woche bleibt nicht nur dieses Thema prominent, sondern

auch Thema 15 „Berichterstattung und Kommentare“ schließt dicht auf, das auf eine intensive Auseinandersetzung mit der medialen Berichterstattung schließen lässt. Schließlich steigt Thema 0 „Verantwortung...“ im Zuge des Skandals ähnlich steil an wie Thema 15, sodass zu vermuten ist, dass der Skandal die Bewertung der Rolle des Kraftwerksbetreibers von Fukushima und der japanischen Regierung beeinflusst.

### 3.4 Schlussfolgerungen: Neue Medien als Artikulationsraum

Yanagisawa (2012) hat die Berichterstattung über den Cäsium-Skandal anhand von ca. 1000 japanischen Zeitungsartikeln für den gleichen Zeitraum detailliert untersucht. Ein Vergleich seiner Ergebnisse mit den Themen, die mit Hilfe des TopicExplorers im Blog-Korpus identifiziert wurden, zeigt, dass sowohl inhaltlich als auch in ihrer spezifischen Abfolge die Themen weitgehend korrespondieren (Tabelle 3). Beispielsweise hat Thema 9 im Blog-Korpus, das die Ursache des Skandals beschreibt, einen ganz ähnlichen Gegenstand wie Cluster 8 bei Yanagisawa. In den Blog-Texten werden sämtliche Themen, die in den Printmedien behandelt werden, ebenfalls aufgegriffen. Die bei der Beschreibung der einzelnen Themen behandelten Beispiele illustrieren, dass die konventionellen Medien auf die Inhalte der Blogs einen großen Einfluss ausüben – schon alleine deshalb, weil mediale Inhalte vielfach zitiert und die darin enthaltenen Informationen in den Blogs oft kommentiert werden.

Im Blog-Korpus finden sich darüber hinaus jedoch auch zahlreiche Gegenstände, die die Printmedien nicht thematisieren. Sie sind in den unteren Zeilen von Tabelle 3 aufgeführt. So findet im Internet eine kritische Auseinandersetzung mit der Berichterstattung der konventionellen Medien statt (Thema 15), es wird über die Notwendigkeit und die Ergebnisse unabhängiger Radioaktivitätsmessungen berichtet (Thema 8), und Erfahrungen aus Tschernobyl (Thema 19) werden ebenso weitergegeben wie Appelle für eine rasche und flächendeckende Dekontamination (Thema 14). An diesen Punkten zeigen sich die Interessen und Sorgen der Bevölkerung, auf die die konventionellen Medien nicht eingehen und die daher im Sinne einer unabhängigen „Artikulation“ im Internet besprochen werden.

Tabelle 3: Entsprechungen von Clustern nach Yanagisawa (2012: 71) mit Topic-Explorer-Themen (Quelle: Eigene Zusammenstellung).

Cluster	Cluster-Nr.	Themen-Nr.
Schaden	1	12
Atomkraftwerksunfall und seine Folgen	2	6, 10, 13, 18
Forderungen	3	4
Lebensmittel	4	1
Auswirkungen auf den menschlichen Körper	5	2, 3
Gegenwärtige Lage	6	5, 7
Gegenmaßnahmen	7	11
Ursachen	8	9
Verantwortung für Unfallfolgen		0
(Selbst)Messungen von Radioaktivität		8
Dekontamination		14
Medienkritik		15
Persönliche Wahrnehmung		16
Probleme der Rinderzüchter		17
Erfahrungen von Tschernobyl		19

Aber auch innerhalb der Themen, die Blogs und Printmedien gemeinsam haben, zeigen sich unterschiedliche Schwerpunktsetzungen. Dies ist oft auf den besonderen Einfluss zurückzuführen, den bestimmte Persönlichkeiten auf das Internet ausüben. So ist davon auszugehen, dass die Kritik, die z. B. fachliche Autoritäten wie Koide Hiroaki am Handeln der Regierung äußerten und die im Internet weite Verbreitung fand, das Vertrauen der Internetnutzer in die Lebensmittelsicherheit beeinflusst hat. Es ist daher anzunehmen, dass trotz des Einflusses, den die traditionellen Medien besitzen, Personen mit ausgewiesener Expertise ihren Ansichten durch die Verbreitung von Blogbeiträgen zu einer großen gesellschaftlichen Präsenz verhelfen können.

#### 4. Fazit und Perspektiven

Mit dem TopicExplorer wurde ein Werkzeug entwickelt, das Personen ohne Programmierkenntnisse die explorierende Analyse großer Textsammlungen ermöglicht und dadurch einer breiten Gruppe von Anwendern Zugang zu neuen Forschungsmöglichkeiten mit Big Data eröffnet. In zukünftigen Arbeitsschritten soll insbesondere die Benutzerfreundlichkeit weiter erhöht und die Steuerung des Systems vereinfacht werden. Außerdem sollen in den TopicExplorer Qualitäts- und Kohärenzmaße für jene automatisch berechneten Themen eingebaut

werden, die mit menschlichen Bewertungen von Themen hinsichtlich ihrer Interpretierbarkeit eng korrelieren (Röder et al. 2015). Somit können auch Nutzer ohne vertiefte Kenntnisse statistischer Verfahren Topic Models intuitiv bewerten und aussagekräftige Themen identifizieren.

Die Beispielanalyse japanischer Blogs mithilfe des TopicExplorers hat gezeigt, dass viele Japanerinnen und Japaner den Lebensmittelskandal um Cäsium-verseuchtes Rindfleisch in den alten und neuen Medien kritisch verfolgen. Außerdem lässt die Analyse Aspekte in den Blog-Einträgen erkennen, die in den gedruckten Medien nicht in vergleichbarer Weise zum Ausdruck kommen, wie etwa die große Sorge der Bevölkerung vor gesundheitlichen Schäden und das Interesse an Informationen, die als regierungsunabhängig eingeschätzt werden. Der Einfluss der sozialen Medien geht dabei bereits so weit, dass sich das japanische Industrie- und Wirtschaftsministerium METI veranlasst sah, ein Unternehmen zu beauftragen, die Diskussion über Kernenergie auf dem Forum „2channel“ und auf Twitter zu überwachen – angeblich, um wirtschaftlichen Schäden durch Verbreitung falscher Gerüchte vorzubeugen.

## 5. Bibliografie

- Alexander, Eric, Joe Kohlmann, Robin Valenza, Michael Witmore und Michael Gleicher. 2014. “Serendip: Topic Model-Driven Visual Exploration of Text Corpora.” In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 173–182, <https://doi.org/10.1109/VAST.2014.7042493>.
- Anthes, Gary. 2010. Topic models vs. unstructured data, *Commun. ACM* 53, no. 12 (December): 16–18. <https://doi.org/10.1145/1859204.1859210>.
- Armbrust, Michael, Reynold S. Xin, Cheng Lian, Yin Huai, Davies Liu, Joseph K. Bradley, Xiangrui Meng, Tomer Kaftan, Michael J. Franklin, Ali Ghodsi und Matei Zaharia. 2015. “Spark SQL: Relational data processing in spark.” In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (SIGMOD '15)*, New York, NY, USA, 2015, 1383–1394. <https://doi.org/10.1145/2723372.2742797>.
- Blei, David M. 2012. “Probabilistic Topic Models.” *Commun. ACM* 55, no. 4 (April): 77–84.
- Chang, Jonathan, Jordan L. Boyd-Graber, Sean Gerrish, Chong Wang und David M. Blei, 2009. “Reading Tea Leaves: How Humans Interpret Topic Models.” In *Proceedings of the 22nd International Conference on Neural Information Processing Systems (NIPS 2009)*, Vancouver, British Columbia, Canada, December 07–10, 2009, 288–296.
- Chuang, Jason, Spence Green, Marti Hearst, Jeffrey Heer und Philipp Koehn, Herausgeber. 2014. *Proceedings of the Workshop on Interactive Language*

- Learning, Visualization, and Interfaces*. Association for Computational Linguistics, Baltimore, Maryland, USA, June 2014.
- Cui, Weiwei, Shixia Liu, Zhuofeng Wu und Hao Wei. 2014. “How Hierarchical Topics Evolve in Large Text Corpora.” *IEEE Transactions on Visualization and Computer Graphics*, 20, no. 12 (December): 2281–2290. <https://doi.org/10.1109/TVCG.2014.2346433>.
- Derntl, Michael, Nikou Günnemann, Alexander Tillmann, Ralf Klamma und Matthias Jarke. 2014. “Building and Exploring Dynamic Topic Models on the Web.” In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM '14)*, New York, NY: ACM, 2012–2014. <http://dx.doi.org/10.1145/2661829.2661833>.
- Dou, Wenwen, Li Yu, Xiaoyu Wang, Zhiqiang Ma und W. Ribarsky. 2013. “Hierarchical Topics: Visually Exploring Large Text Collections Using Topic Hierarchies.” *IEEE Transactions on Visualization and Computer Graphics*, 19, no. 12 (December): 2002–2011. <https://doi.org/10.1109/TVCG.2013.162>.
- Evangelopoulos, Nicholas und Lucian Visinescu. 2012. “Text-Mining the Voice of the People.” *Commun. ACM* 55, no. 2 (February): 62–69.
- Gohr, André, Myra Spiliopoulou und Alexander Hinneburg. 2010. “Visually Summarizing the Evolution of Documents under a Social Tag.” In *Proceedings of LWA2010 – Workshop-Woche: Lernen, Wissen & Adaptivität*, Kassel, Germany 2010, 85–94, [http://users.informatik.uni-halle.de/~hinnebur/PS\\_Files/kdir2010\\_TT.pdf](http://users.informatik.uni-halle.de/~hinnebur/PS_Files/kdir2010_TT.pdf).
- Gohr, André, Myra Spiliopoulou und Alexander Hinneburg. 2013. “Visually Summarizing Semantic Evolution in Document Streams with Topic Table.” In *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, herausgegeben von A. Fred, J. L. G. Dietz, K. Liu und J. Filipe. Berlin: Springer, 136–150 (Communications in Computer and Information Science).
- Hinneburg, Alexander, Rico Preiss und René Schröder. 2012. “TopicExplorer: Exploring Document Collections with Topic Models.” In *Machine Learning and Knowledge Discovery in Databases*, herausgegeben von Peter A. Flach, Tijn De Bie, Nello Cristianini. Berlin: Springer, 838–841 (Lecture Notes in Computer Science 7524).
- Hinneburg, Alexander, Frank Rosner, Stefan Pessler und Christian Oberländer. 2014. “Exploring Document Collections with Topic Frames.” In *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM '14)*, New York, NY: ACM, 2084–2086.
- Isaacs, E., K. Damico, S. Ahern, E. Bart und M. Singhal. 2014. “Footprints: A Visual Search Tool that Supports Discovery and Coverage Tracking.” *IEEE Transactions on Visualization and Computer Graphics* 20, no. 12 (December): 1793–1802.

- Jojic, Nebojsa, Alessandro Perina und Dongwoo Kim. 2016. "Hierarchical learning of grids of microtopics." In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence (UAI'16)*, New Jersey, USA, June 25–29, Arlington, Virginia: AUAI Press, 299–308.
- Kingston, Jeff. 2012. "The Politics of Disaster, Nuclear Crisis and Recovery." In *Natural Disaster and Nuclear Crisis in Japan. Response and Recovery after Japan's 3/11*, herausgegeben von Jeff Kingston. London: Routledge, 188–206.
- Kuchinskaya, Olga. 2011. "Articulating the signs of danger: Lay experiences of post-Chernobyl radiation risks and effects." *Public Understanding of Science* 20, no. 3: 405–21.
- Liu, Shixia, Michelle X. Zhou, Shimei Pan, Weihong Qian, Weijia Cai und Xiaoxiao Lian. 2009. "Interactive, Topic-Based Visual Text Summarization and Analysis." In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)*, New York, NY: ACM, 543–552, <http://dx.doi.org/10.1145/1645953.1646023>.
- Liu, Shixia, Xiting Wang, Jianfei Chen, Jun Zhu und Baining Guo. 2014. "Topic-panorama: A Full Picture of Relevant Topics." In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, October 2014, 183–192.
- Murdock, Jaimie und Colin Allen. 2015. "Visualization Techniques for Topic Model Checking." In *29th AAAI Conference on Artificial Intelligence (AAAI-15)*, 4284–4285.
- Pan, Shimei, Michelle X. Zhou, Yangqiu Song, Weihong Qian, Fei Wang und Shixia Liu. 2013. "Optimizing Temporal Topic Segmentation for Intelligent Text Visualization." In *Proceedings of the 2013 International Conference on Intelligent User Interfaces (IUI'13)*, 339–350, New York, NY: ACM, 339–350, <https://doi.org/10.1145/2449396.2449441>.
- Ramage, Daniel, Evan Rosen, Jason Chuang, Christopher D. Manning und Daniel A. McFarland. 2009. "Topic Modeling for the Social Sciences." In *Workshop on Applications for Topic Models (NIPS 2009): Text and Beyond*, Whistler, Canada, December 2009. <http://vis.stanford.edu/files/2009-TopicModels-NIPS-Workshop.pdf>.
- Röder, Michael, Andreas Both und Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*, New York, NY: ACM, 399–408.
- Sievert, Carson und Kenneth Shirley. 2014. "LDAvis: A Method for Visualizing and Interpreting Topics." *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, Maryland, USA, June 2014. Association for Computational Linguistics, 63–70. <https://nlp.stanford.edu/events/illvizo2014/papers/sievert-illvizo2014.pdf>.

- Sun, Guodao, Yingcai Wu, Shixia Liu, Tai-Quan Peng, J.J.H. Zhu und Ronghua Liang. 2014. “Evoriver: Visual Analysis of Topic Coopetition on Social Media.” *IEEE Transactions on Visualization and Computer Graphics* 20, no. 12 (December): 1753–1762.
- Yanagisawa, Toru. 2012. “Keiryō tekisuto bunseki ni yoru media-furēmu no tan-sakuteki kentō – ‘hōshasei seshiumu osen-gyū mondai’ no shinbun kiji wo tōshite.” (An Exploratory Study of Media Frames in Quantitative Text Analysis. Through the Newspaper Report on the Problems with Radioactive Cesium-contaminated Beef). *Shakai jōhō-gaku* 1, no. 2: 61–76.

