Design und Evaluation von Verfahren zur Ereigniserkennung in Sozialen Datenströmen[1]

Andreas Weiler[2]

Abstract: Die Dissertation beschreibt Forschungsergebnisse aus der Entwicklung und Evaluation von Verfahren zur Ereigniserkennung in Sozialen Datenströmen. Im ersten Teil präsentieren wir eine auf die besonderen Merkmale von Datenströmen fokussierte Technik zur Erkennung von Ereignissen in Echtzeit. Zusätzlich wird eine Technik zur Erkennung von Ereignissen in vordefinierten geographischen Bereichen vorgestellt. Im zweiten Teil analysieren wir die aufgabenbezogene sowie die Laufzeitperformance von mehreren modernen Techniken zur Ereigniserkennung unter Verwendung von realen Twitter Datenströmen. Um die Laufzeitperformance reproduzierbar zu vergleichen, basiert unsere Methode auf einem Datenstrommanagementsystem. Die aufgabenbezogene Performance wird basierend auf einer Reihe von neuartigen Maßen evaluiert. Diese Maße wurden speziell zur Gegenüberstellung der quantitativen und der qualitativen Performance entworfen. Der letzte Teil beschreibt das Design zweier Visualisierungen zur Unterstützung der visuellen Erkennung von Ereignissen. Mit Stor-e-Motion präsentieren wir eine Visualisierung zur Überwachung der fortlaufenden Entwicklung von Wichtigkeit, Stimmung und Kontext in benutzerdefinierten Themen. Mit SiCi Explorer präsentieren wir eine Visualisierung zur Überwachung von Ereignissen, Themen und Stimmungen über die Zeit und Raum für benutzerdefinierte geographische Bereiche. Für diese Visualisierung wird abschließend eine Benutzerstudie vorgestellt.

1 Einführung und Motivation

Der beispiellose Erfolg und die aktive Verwendung von Social Media Diensten führt zu einer gewaltigen Menge an nutzergenerierten Daten. Ein führender Akteur bei der Erzeugung von großen Datenvolumen als kontinuierlichen Datenstrom bestehend aus Kurznachrichten, den sogenannten Tweets, ist das soziale Netzwerk Twitter. Die Kürze der Tweets machen sie zu einem idealen Medium für die mobile Kommunikation. Deshalb steigt die Beliebtheit von Twitter als Quelle für aktuelle Nachrichten und Informationen über aktuelle Ereignisse stetig. Als Reaktion auf diesen Trend wurden zahlreiche Forschungsarbeiten über Techniken zur Ereigniserkennung, welche auf den Datenstrom von Twitter angewendet werden, vorgestellt. Jedoch weisen die meisten Techniken zwei Hauptmängel auf.
Erstens, tendieren sie dazu, sich exklusiv auf den Aspekt der Informationsgewinnung zu konzentrieren und ignorieren des öfteren die besonderen Eigenschaften von Datenströmen. Obwohl alle vorgestellten Arbeiten Nachweise über die Qualität der erkannten Ereignisse erbringen, stellt zweitens keine Arbeit einen Bezug zwischen dieser aufgabenbezogenen Performance und der Laufzeitperformance hinsichtlich Verarbeitungsgeschwindigkeit oder Datendurchsatz her. Insbesondere wurde bis heute keine quantitative oder vergleichende Evaluation dieser Aspekte durchgeführt.

[1] Englischer Titel der Dissertation: Design and Evaluation of Event Detection Techniques for Social Media Data Streams

[2] Universtität Konstanz, andreas.weiler@uni-konstanz.de

Die wissenschaftlichen Beiträge der Dissertation lassen sich in folgende Punkte aufgliedern.

- Wir haben zwei Verfahren zur Ereigniserkennung in Twitter Datenströmen entwickelt und in unsere Evaluationen eingebunden. Ein Verfahren ist spezialisiert für das Auffinden von nicht spezifizierten Ereignissen in Echtzeit, das andere Verfahren für Orts spezifische Ereignisse.
- Wir haben einen umfassenden Überblick über verwandte Arbeiten im Kontext von Verfahren zur Ereigniserkennung und existierende Evaluationsmethoden der Verfahren erstellt.
- Wir haben existierende Ereigniserkennungsverfahren in einem Datenstrommanagement System implementiert, um eine konsistente Evaluation der Verfahren zu ermöglichen.
- Wir haben zwei detaillierte Studien und Evaluationen der Verfahren in Bezug auf Qualität und Verarbeitungsgeschwindigkeit durchgeführt.
- Wir haben unsere Evaluationen auf einer allgemein verfügbaren Plattform implementiert, um weitere Evaluationen von neuartigen Erkennungsverfahren zu ermöglichen.
- Wir haben zwei Visualisierungsverfahren entworfen, entwickelt und evaluiert, welche die Erkennung von Ereignissen im Twitter Datenstrom unterstützen. Die erste Visualisierung erkennt Ereignisse in definierten Themen in Echtzeit und die zweite spezifische Ereignisse in definierten geographischen Bereichen.

2 Stand der Technik

In den letzten Jahren wurde das Thema Twitter und die aus dem Dienst entstehenden Datenmengen ein fester Bestandteil der wissenschaftlichen Community. Die Anzahl an Publikationen zum Thema Twitter wie aber auch zum Thema Ereigniserkennung ist in den letzten Jahren ständig gestiegen. Die Dissertation stellt einen umfassenden Überblick über die bestehenden wissenschaftlichen Arbeiten im Bereich Ereigniserkennung auf Twitter wie aber auch Arbeiten im Bereich Evaluationen der Verfahren zu Ereigniserkennung bereit. Hierbei haben wir wissenschaftliche Arbeiten untersucht, welche in den bereits existierenden Übersichten [FK15, Gu13, BR14] nicht vorhanden waren.
Ab dem Jahr 2013 kamen 14 neue Publikationen zum Thema Ereigniserkennung hinzu. Insgesamt kann man erkennen, dass der Hauptanteil an Verfahren die open-domain Erkennung von Ereignissen anstrebt (30 von 42). Zusätzlich gibt es Verfahren, welche sich auf die Erkennung von Ereignissen bei Epidemien oder Katastrophen spezialisiert haben. Die verwendeten Techniken der Verfahren umfassen statistische Modelle, Clustering und Wavelet Analysen. Hierbei werden stets verschiedene Dimensionen für die Analysen verwendet, wobei am häufigsten die Zeit- und Geographische Dimension oder eine Kombination der beiden verwendet wird. Aus dieser Kollektion an Verfahren an haben wir uns zwei bekannte Verfahren [WL11, Co12] für unsere Evaluationen ausgewählt, welche wir in einem Data Stream Management System implementiert haben.
Weiterhin haben wir alle Arbeiten auf ihre Evaluationsmethoden untersucht und diese in

einer umfangreichen Übersicht dargestellt. Als Ergebnis aus den Untersuchungen hat sich ergeben, dass trotz der großen Anzahl (42) an wissenschaftlichen Arbeiten keine umfangreichen Evaluationen hinsichtlich der Qualität der Ergebnisse oder der Verarbeitungsgeschwindigkeit der Verfahren durchgeführt wurden. Zum Beispiel haben nur zwei der untersuchten Arbeiten ihre Verfahren hinsichtlich Verarbeitungsgeschwindigkeit evaluiert. Es wurde zusätzlich ersichtlich, dass viele Arbeiten unrealistische Annahmen treffen. Diese Annahmen sind unter anderem dass die Verfahren auf Sammlungen von Daten mehrere Monate ausgeführt werden oder sämtliche Parameter der Verfahren statisch gesetzt werden ohne jegliche Evaluation ob der Datenstrom hiermit verarbeitet werden kann.
Weiterhin gibt es sehr große Unterschiede in den Evalutionskonfigurationen und Evaluationsmethoden. Die Anzahl an Tweets, welche als Eingabedaten für die Verfahren verwendet wird, liegen im Bereich von 0,6 Mio. bis zu über 1,2 Mrd. Eine Schwierigkeit hierbei ist zusätzlich, dass diese Sammlungen von Tweets nicht öffentlich zugänglich gemacht werden dürfen. Deshalb argumentieren wir, dass es nötig ist, die Verfahren in einem Data Stream Management System zu implementieren und auf denselben Eingangsdaten auszuführen, um vergleichbare Ergebnisse zu erhalten.

3 Wissenschaftliche Beiträge

3.1 Verarbeitung der Twitter Datenströme

Für die Verarbeitung der Twitter Datenströme wurde eine Plattform (siehe Abb.1) entwickelt, welche die Tweets von Twitter per Streaming API empfängt, diese auf der einen Seite persistiert und auf der anderen Seite live durch das Data Stream Management System leiten kann. Durch die Verbindungen mit der Streaming API erhalten wir stündlich bis zu 3 Mio. Tweets aus den öffentlichen Kanälen von Twitter. Der Stream Manager konsolidiert die Tweets aus den verschiedenen Streams und leitet diese in den Data Store. Der Data Store speichert die Tweets und streamt diese gleichzeitig weiter in das Data Stream Management System. Weiterhin kann der Data Store auch historische Daten als Stream in das Data Stream Management System leiten.

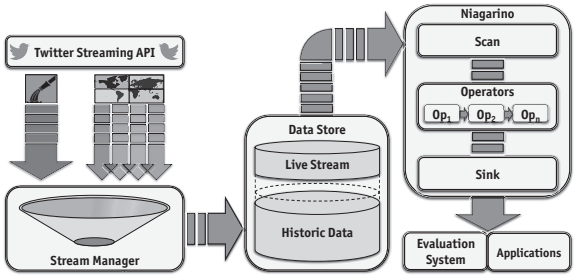

Abb. 1: Plattform und Data Stream Management System für die Verarbeitung der Tweets.

Im Data Stream Management System wird eine Anfrage als gerichteter azyklischer Graph $Q = (O,S)$ repräsentiert, wobei O die Menge der Operatoren der Anfrage und S die Menge

der Streams, welche verwendet werden um die Operatoren zu verknüpfen. Das Datenmodell basiert auf relationalen Tupeln, welche der ersten Normalform folgen und somit keine Verschachtlungen erlauben. Basierend auf dem relationalen Datenmodell umfasst die Menge der Operatoren unter anderem die Selektion (σ) und die Projektion (π), welche genau wie ihre Gegenstücke in relationalen Datenbanksystemen funktionieren. Weitere tuple-basierte Operatoren sind der Derive (f) und der Unnest (μ) Operator. Der Derive Operator führt eine Funktion auf einem Tupel aus und fügt den Ergebniswert an das Tupel an. Der Unnest Operator teilt ein Attribut und sendet ein neues Tupel pro neuem Wert in den Stream. Ein typischer Anwendungsfall für den Unnest Operator ist es eine Zeichenkette zu zerteilen und für jede Teilzeichenkette ein extra Tupel zu senden. Zusätzlich zu diesen allgemeinen Operatoren bietet das Daten Stream Management System eine Vielzahl von stream-spezifischen Operatoren, welche dazu verwendet werden können den Stream in verschiedenste Bereiche zu unterteilen. Zusätzlich zu den bekannten Zeit- und Tupelbasierten Window Operatoren (ω), welche tumbling oder sliding sein können, gibt es auch datengetriebene Datenfenster, welche als Frames [Ma12] bekannt sind. Segmente können mit Join ($\bowtie$) und Aggregationsoperatoren (Σ) verarbeitet werden. Die Implementierung der Anfragepläne der von uns entwickelten und der evaluierten Verfahren ist in der Abbildung 4 dargestellt.

3.2 Verfahren zu Ereigniserkennung

Die Dissertation stellt den Entwurf und die Entwicklung von zwei Verfahren zu Ereigniserkennung vor. Das erste Verfahren (*Shifty* [WGS14a]) beschäftigt sich mit der Erkennung von Ereignissen ohne Einschränkung der Anwendungsdomäne und in Echtzeit. Das Verfahren beruht auf der Analyse der Frequenz von Termen über die Zeit in verschiedenen Intervallen. Dabei wird die Veränderung der Frequenz überwacht und beim überschreiten von gewissen Schwellwerten ein Ereignis ausgelöst. Da dieses Ereignis auf einem einzelnen Term basiert werden dem Ereignis weitere zusammen auftretende Terme hinzugefügt. Diese Zusammenfassung beschreibt das resultierende Event.

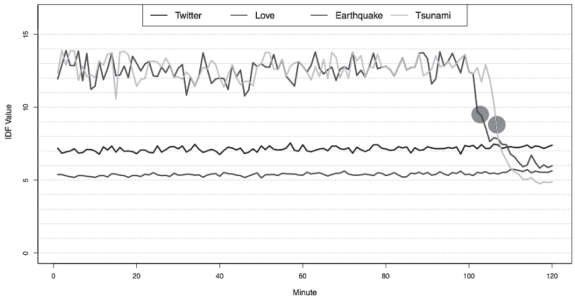

Abb. 2: Erkennung der Ereignisse *Earthquake* und *Tsunami* mit dem Shifty Verfahren.

In Abbildung 2 kann man erkennen, wie sich die Frequenz der Terme *Twitter*, *Love*, *Earthquake* und *Tsunami* über eine Zeit von zwei Stunden verändert. Die beiden Terme *Twitter* und *Love* zeigen fast keinerlei Veränderung ihrer Frequenz. Im Gegensatz hierzu erhöht

sich die Frequenz (invers dargestellt) der Terme *Earthquake* und *Tsunami* plötzlich zwischen den Minuten 100 und 110. An den roten Markierungen greifen die statistischen Regeln des Verfahrens und melden die jeweiligen Terme als Ereignisse. Weiterhin werden den Termen ihre zusammen auftretenden Terme hinzugefügt und somit steht für den Term *earthquake* das Ereignis *earthquake, epicenter, aceh, banda, tsunami*. Hierbei kann man zusätzlich die Verwandtschaft der beiden Ereignisse erkennen.
Das zweite Verfahren (*LLH* [We13]) beschäftigt sich mit der Erkennung von Ereignissen in definierten geographischen Bereichen. Das Verfahren beruht auf der Kombination der Berechnung eines kombinierten des log-likelihood Verhältnis. Die Kombination umfasst das Verhältnis zwischen der Frequenz von Termen innerhalb der definierten Region und außerhalb. Zusätzlich wird das log-likelihood Verhältnis zwischen der Frequenz von Termen innerhalb der definierten Region des aktuellen Zeitfensters gegenüber der historischen Frequenz der Terme in der definierten Region analysiert. Wenn das kombinierte log-likelihood Verhältnis eines Termes einen gewissen Schwellenwert überschreitet wird dieser Term als Ereignis für den spezifizierten geographischen Bereich und das aktuelle Zeitfenster gemeldet. Auch bei diesem Verfahren wird dieser Term mit weiteren gemeinsam auftretenden Termen angereichert, um ein aussagekräftiges Ereignis zu bilden. Ein Beispiel hierfür ist in Abbildung 3 für das geographische Gebiet um Boston dargestellt.

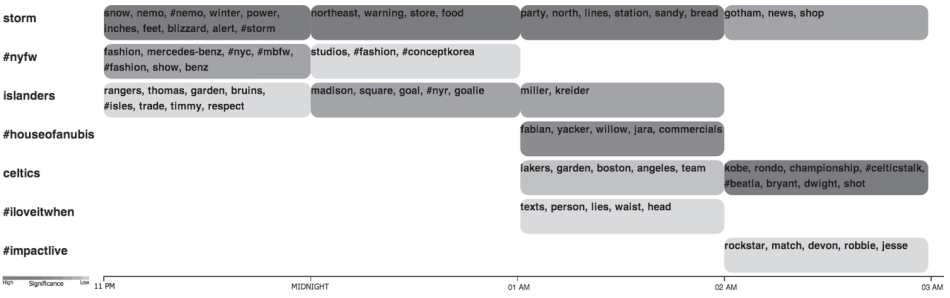

Abb. 3: Erkennung von Ereignissen für das geographische Gebiet um Boston.

3.3 Evaluation zu Verfahren zur Ereigniserkennung

Die Dissertation präsentiert Methoden, um die Qualität und die Verarbeitungsgeschwindigkeit von aktuellen wie aber auch zukünftigen Verfahren zur Ereigniserkennung zu evaluieren. Zusätzlich zu den speziellen Erkennungsverfahren wurden einige Techniken implementiert, welche als Basis verwendet werden, um bessere Vergleiche zu ermöglichen. Um eine möglichst gute Vergleichbarkeit zu gewährleisten wurden alle Verfahren in einem bestehenden Data Stream Management System implementiert (siehe Abb. 4). Zusätzlich wurden einige skalierbare Kennzahlen definiert, um die Qualität der Ergebnisse der Verfahren automatisch zu evaluieren ohne manuell eingreifen zu müssen.

In den Evaluationen [WGS15b, WGS15a] werden die Ergebnisse von vier speziellen Erkennungsverfahren und fünf Basisverfahren untersucht. Aus der großen Sammlung wur-

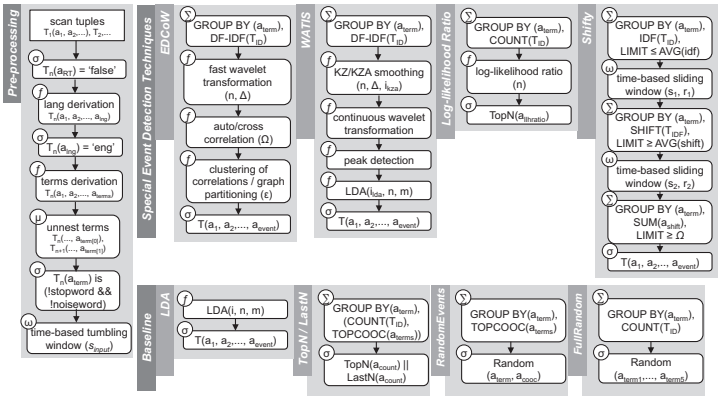

Abb. 4: Anfragepläne des Data Stream Managementsystems für die Verfahren zu Ereigniserkennung.

den hierzu die beiden meist zitierten und damit erfolgversprechendsten Kandidaten *EDCoW* [WL11] und *WATIS* [Co12] ausgewählt. Diese beiden Verfahren verwenden komplexe Wavelet Analysen, um Ereignisse in Datenströmen zu erkennen. Im Gegensatz zu diesen beiden Verfahren wurden zusätzlich unsere selbst entwickelten, weniger komplexere Verfahren Log-likelihood Ratio (LLH) und Shifty evaluiert. Als Basisverfahren wurden *LDA* [BNJ03], ein bekanntes Topic Detection Verfahren, die Top bzw. Letzten N Treffer aus der Menge an Termen, und zwei Verfahren welche zufällige Terme als Ereignisse melden. Um eine faire Evaluation zu gewährleisten, wurden alle Verfahren auf denselben Eingangsdaten mit denselben Vorverabeitungsschritten ausgeführt (siehe Pre-Processing in Abb. 4). Zusätzlich zu den genannten Verfahren wurden ebenfalls die von Twitter selbst generierten Trending Topics (TT) in die Qualitätsevaluation aufgenommen.

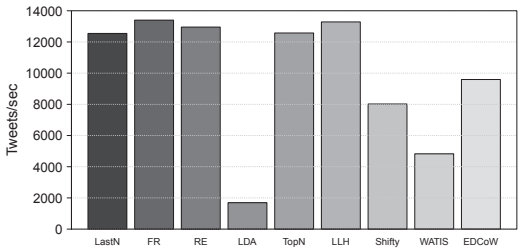

Abb. 5: Vergleich der Laufzeitperformance der Verfahren.

Beim Vergleich der Laufzeitperformance (siehe Abb. 5) kann man erkennen, dass LLH vor EDCoW, Shifty und WATIS in der Rangliste liegt. Am wenigstens Tweets pro Sekunde kann das Topic Detection Verfahren LDA verarbeiten. Dies liegt daran, dass LDA mehrere Iterationen über die Sammlung der Tweets ausführt und somit einen höheren Verabeitungsaufwand hat. Zusätzlich muss man erwähnen, dass Shifty iterativ auf kleineren Zeitfenstern als die anderen Verfahren arbeitet und somit hier nachteilig behandelt wird.

Zusammenfassend kann man aussagen, dass alle Verfahren bis auf LDA und WATIS in der Lage sind, den Datenstrom von Twitter in Echtzeit abzuarbeiten.

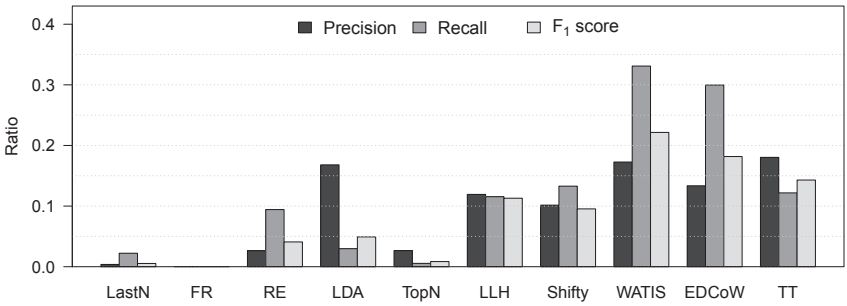

Abb. 6: Qualitätsvergleich der Ergebnisse der Verfahren.

Für den Qualitätsvergleich (siehe Abb. 6) wurden einige neue Maße entwickelt, welche in Kombination zur Anwendung kommen. Die Präzision wird z.B. anhand eines Abgleichs mit Google und dem Newsarchiv von NYTimes berechnet. Hierbei wird das Verhältnis der gemeldeten Ereignisse zu den gefundenen Ereignissen in Google und dem NYTimes Archiv analysiert. Der Recall wird z.B. anhand eines Archivs der Nachrichtenseite Reuters berechnet und analysiert ebenfalls die Gesamtmenge der gemeldeten Ereignisse zu den im Newsarchiv vorhandenen. Weiterhin wird bei der Evaluation die Menge an doppelt gemeldeten Ereignissen berücksichtigt. Für die Ergebnisse der Qualitätsevaluation kann folgendes festgehalten werden. Die berechneten Werte sind für alle der getesteten Verfahren eher niedrig. Da wir jedoch nur an relativen Ergebnissen interessiert sind stellt dies kein Problem der Evaluation dar. Im Bezug auf Präzision erreichen LDA, WATIS und TT die besten Werte. Beim Recall überzeugen ebenfalls WATIS und das weitere komplexe Erkennungsverfahren EDCoW. Beide Werte in Kombination stellen das F_1 Score dar. Hierbei sehen wir, dass die komplexen Erkennungsverfahren und die von Twitter selbst bereitgestellten Trending Topics die besten Werte erzeugen. Die von uns entwickelten weniger komplexen Verfahren erreichen zwar viel höhere Werte als die Basis Verfahren, können jedoch nicht mit den komplexen Verfahren mithalten.

3.4 Visuelle Ereigniserkennung

Die Dissertation stellt zwei Visualisierungsverfahren zur Erkennung von Ereignissen im Twitter-Datenstrom vor. Das erste Verfahren (*Stor-e-Motion*[We14, WGS15c]) stellt den Verlauf von vordefinierten Themen dar und lässt den Benutzer erkennen, falls plötzliche Veränderungen in der Frequenz oder den Emotionen über das Thema stattfinden. Wie in Abbildung 7 zu erkennen ist, kann man hierbei den Tagesverlauf des Themas live verfolgen und wird über unübliche Veränderungen visuell auf dem Laufenden gehalten. Zusätzlich wird stündlich eine Zusammenfassung der Terme der letzten Stunde festgehalten und in der Visualiserung dargestellt. Hierdurch kann verfolgt werden, welche Terme in den Themen wichtig sind und zu den Veränderungen in der Frequenz und den Emotionen geführt haben.

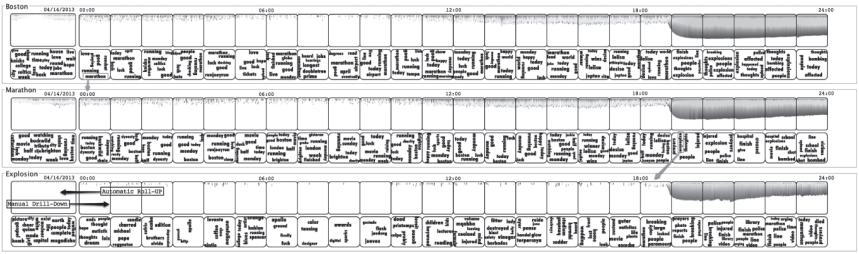

Abb. 7: Monitoring der Topics Boston, Marathon und Explosion während des Boston Marathons im Jahre 2013 mit der Stor-e-Motion Visualisierung.

Das zweite Verfahren (*SiCi Explorer* [WGS16, WGS14b]) stellt den Verlauf von Emotionen und Frequenz von Tweets in vordefinierten geographischen Bereichen dar. Diese Visualisierung lässt den Benutzer erkennen, falls plötzliche Veränderungen in den Emotionen der Tweets gibt, die von Personen innerhalb des geographischen Bereichs oder von Personen, die über den geographischen Bereich (z.B. Boston) schreiben, stammen. Das Beispiel in Abbildung 8 stellt den Verlauf der Tweets aus Boston und über Boston während des Marathons im Jahre 2013 dar. Man erkennt, dass in den letzten 10 Minuten der zweiten Stunde (erste Zeile in der Mitte) die Emotionen von positiv (grün) auf negativ (rot) umschwenken. Dies ist der Zeitpunkt, an dem der Anschlag auf den Marathon in Boston ausgeübt wurde. Weiterhin werden in den einzelnen Segmenten der Visualisierung die beliebtesten Terme eingeblendet und die dazugehörigen Tweets können angezeigt werden. Eine Benutzerstudie zu dieser Visualisierung hat gezeigt, dass die Visualisierung intutiv verstanden wird und die Benutzer in der Wahrnehmung von Ereignissen unterstützt.

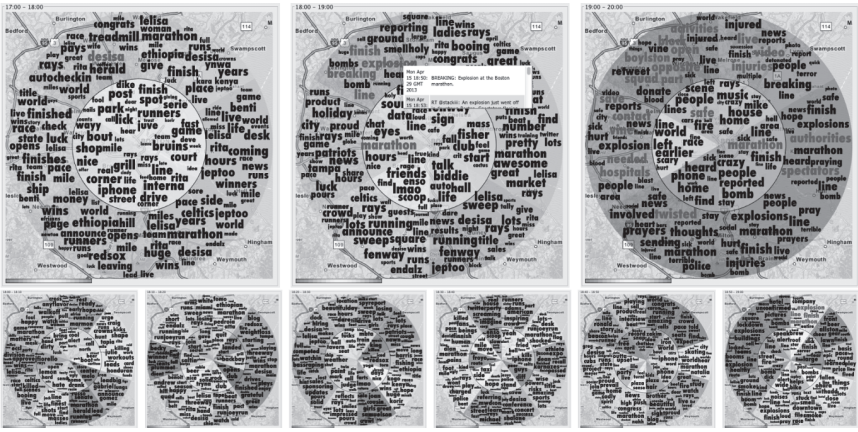

Abb. 8: Monitoring der Situation in Boston während des Boston Marathons im Jahre 2013 mit dem SiCi Explorer.

3.5 Fazit

Die Dissertation [We16] stellt die Ergebnisse von drei Hauptforschungsfragen dar. Das erste Ziel war es, Verfahren zu Erkennung von Ereignissen in den Datenströmen zu entwerfen und entwickeln. Das zweite Ziel leitet sich aus dem Ersten ab und entwickelte sich daraus, dass keine ausreichenden Evaluationen und Evaluationskriterien für Erkennungsverfahren von Ereignissen in Twitter existierten. Das dritte Ziel war es Visualisierungen zu entwerfen und zu entwickeln, welche die Erkennung von Ereignissen in den Twitter Datenströmen aktiv unterstützen.
Zusammenfassend diskutiert und präsentiert die Dissertation neue, noch offene Fragen für zukünftige Forschung. Hierbei wird z.B. aufgezeigt, wie die Evalutionen auf weitere Verfahren ausgeweitet werden können und diese Verfahren auf einfach Art und Weise in das Evaluationsframework eingebunden werden können. Weiterhin wird eine völlig neue Forschungsfrage aufgeworfen. Da die meisten Verfahren zu Ereigniserkennung auf vielen Parametern, wie z.B. die Größe der Zeitfenster, verschiedenste Schwellwerte oder aber die Anzahl an Verarbeitungszyklen, muss dies in weiteren Evaluationen untersucht werden. Dabei ist es spannend zu untersuchen ob diese Parameter fortlaufend während der Ausführung der Verfahren verändert und angepasst werden können. Dieses Untersuchungen könnten dann zu dynamischen, adaptiven und skalierbaren Verfahren zur Ereigniserkennung in Twitter Datenströmen führen.

Literaturverzeichnis

[BNJ03] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I.: Latent Dirichlet Allocation. J. Mach. Learn. Res., 3:993–1022, 2003.

[BR14] Bontcheva, Kalina; Rout, Dominic: Making Sense of Social Media Streams through Semantics: a Survey. Semantic Web, 5(5):373–403, 2014.

[Co12] Cordeiro, Mário: Twitter Event Detection: Combining Wavelet Analysis and Topic Inference Summarization. In: Proc. Doctoral Symposium on Informatics Engineering (DSIE). 2012.

[FK15] Farzindar, Atefeh; Khreich, Wael: A Survey of Techniques for Event Detection in Twitter. Computational Intelligence, 31(1):132–164, 2015.

[Gu13] Guille, Adrien; Favre, Cécile; Hacid, Hakim; Zighed, Djamel A.: Information Diffusion in Online Social Networks: A Survey. SIGMOD Rec., 42(2):17–28, 2013.

[Ma12] Maier, David; Grossniklaus, Michael; Moorthy, Sharmadha; Tufte, Kristin: Capturing Episodes: May the Frame Be with You (Invited Paper). In: Proc. Intl. Conf. on Distributed Event-Based Systems (DEBS). S. 1–11, 2012.

[We13] Weiler, Andreas; Scholl, Marc H.; Wanner, Franz; Rohrdantz, Christian: Event Identification for Local Areas Using Social Media Streaming Data. In: Proc. Workshop on Databases and Social Networks (DBSocial) in conjunction with Intl. Conf. on Management of Data (SIGMOD). S. 1–6, 2013.

[We14] Weiler, Andreas; Grossniklaus, Michael; Wanner, Franz; Scholl, Marc H.: The Stor-e-Motion Visualization for Topic Evolution Tracking in Social Media Streams. In: Proc. Eurographics Conference on Visualization (EuroVis): Posters. 2014.

[We16] Weiler, Andreas: Design and Evaluation of Event Detection Techniques for Social Media Data Streams. Dissertation, University of Konstanz, Konstanz, 2016.

[WGS14a] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: Event Identification and Tracking in Social Media Streaming Data. In: Proc. Workshop on Multimodal Social Data Management (MSDM) in conjunction with Intl. Conf. on Extending Database Technology (EDBT). S. 282–287, 2014.

[WGS14b] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: SiCi Explorer: Situation Monitoring of Cities in Social Media Streaming Data. In: Proc. Workshop on Mining Urban Data (MUD) in conjunction with Intl. Conf. on Extending Database Technology (EDBT). S. 369–370, 2014.

[WGS15a] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: Evaluation Measures for Event Detection Techniques on Twitter Data Streams. In: Proc. British Intl. Conf. on Databases (BICOD). S. 108–119, 2015.

[WGS15b] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: Run-time and Task-based Performance of Event Detection Techniques for Twitter. In: Proc. Intl. Conf. on Advanced Information Systems Engineering (CAiSE). S. 35–49, 2015.

[WGS15c] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: The Stor-e-Motion Visualization for Topic Evolution Tracking in Text Data Streams. In: Proc. Intl. Conf. on Information Visualization Theory and Applications (IVAPP). S. 29–40, 2015.

[WGS16] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: Situation Monitoring of Urban Areas Using Social Media Data Streams. Information Systems, 57:129–141, 2016.

[WL11] Weng, Jianshu; Lee, Bu-Sung: Event Detection in Twitter. In: Proc. Intl. Conf on Weblogs and Social Media (ICWSM). S. 401–408, 2011.

Andreas Weiler wurde am 20. März 1983 in Stuttgart Bad-Cannstatt geboren. Er studierte im Studiengang Information Engineering an der Universität Konstanz. Im Jahr 2006 erhielt er den Abschluss des Bachelor of Science mit einer Spezialisierung in Datenbanken und Informationssysteme. Seine Bachelorarbeit widmete sich dem Thema „Design und Entwicklung eines visuellen Frontends zur Exploration von multi-dimensionalen OLAP Daten“. Hierbei wurde der Fokus speziell auf die interaktive Ausführung der Anfragen und eine innovative Ergebnisdarstellung gelegt. Im Jahr 2010 erhielt er den Abschluss des Masters of Science in Information Engineering mit dem Fokus Computer Science. In seiner Masterarbeit arbeitete er am Design und der Entwicklung einer Client/Server Architektur für die Open-Source XML Datenbank BaseX. Im Jahr 2011 wurde er als Doktorand Teil der Forschungsgruppe für Datenbank und Informationssysteme von Prof. Dr. Marc H. Scholl. Seine Forschungsinteressen liegen in den Gebieten der Verarbeitung und Analyse von Informationen aus sozialen Datenströmen. Speziell evaluierte er Ereigniserkennungsverfahren in Bezug auf Performanz in der Laufzeit und Qualität. Andreas Weiler erhielt seinen Doktortitel für die Dissertation mit dem Titel „Design and Evaluation of Event Detection Techniques for Social Media Data Streams“ im Jahre 2016. Die Betreuer und Gutachter der Dissertation waren Prof. Dr. Marc H. Scholl und Prof. Dr. Michael Grossniklaus. Aktuell arbeitet er in Konstanz als Senior Data Engineer und Data Scientist bei coliquio.de dem größten Experten-Netzwerk für Ärzte.