Storyteller: Visual Analytics of Perspectives on Rich Text Interpretations

Maarten van Meersbergen
Janneke van der Zwaan
Willem van Hage
Netherlands eScience Center
Science Park 140
Amsterdam, The Netherlands
Email: m.vanmeersbergen@esciencecenter.nl
j.vanderzwaan@esciencecenter.nl
w.vanhage@esciencecenter.nl

Piek Vossen
Antske Fokkens
Inger Leemans
Isa Maks
VU University Amsterdam
De Boelelaan 1081
Amsterdam, The Netherlands
Email: piek.vossen@vu.nl
antske.fokkens@vu.nl
inger.leemans@vu.nl
isa.maks@vu.nl

Abstract

Complexity of event data in texts makes it difficult to assess its content, especially when considering larger collections in which different sources report on the same or similar situations. We present a system that makes it possible to visually analyze complex event and emotion data extracted from texts. We show that we can abstract from different data models for events and emotions to a single data model that can show the complex relations in four dimensions. The visualization has been applied to analyze 1) dynamic developments in how people both conceive and express emotions in theater plays and 2) how stories are told from the perspective of their sources based on rich event data extracted from news or biographies.

1 Introduction

People frequently write about the changes in the world and about the emotions that these events arouse. Events and emotions typically have complex structures and are difficult to detect and represent. According to our estimation, standard news articles contain about 200 event mentions on average (Vossen et al., 2016). These events stand in complex temporal and causal relations to each other, while the text can also present different perspectives on their impact. Especially when considering event data from many different sources that may report on the same events, the extracted data quickly becomes very complex.

Most systems that handle large collections of text use some kind of topic clustering. Documents are grouped together on the basis of a similarity measure and clustering technique. In the case of streaming text, such as news and tweets, topic modeling and clustering is also done dynamically over time, indicating when a cluster appears and dies out. Dynamic clustering can be seen as a rough approximation of the temporal bounding of a real world event. Although topic modeling works well as an indicator of trending real world events, it does not tell the story in detail as a sequence of events with participants and causal/temporal relations across these events.

In this paper we present Storyteller, a visual analytics tool for complex multidimensional textual data. Storyteller groups events with multiple participants into storylines. As such, it can give insight into complex relations by providing multiple perspectives on time-based textual data. We explain the capabilities of Storyteller by applying it to semantically linked news data from the NewsReader project,[1] using different perspectives on the data to visualize complex relations between the events found by its Natural Language Processing (NLP) pipeline. These visualizations give more insight into the performance of the system and how well complex event relations approximate the storylines people construct when reading news. We further show the usability of the Storyteller visualization for general purpose event-based textual data by applying it to other use cases, namely the *Embodied Emotions* and *BiographyNet* projects provided by other humanities experts.

The paper is structured as follows. In Section 3, we present the semantic model for events used in NewsReader. Section 4 explains the Storyteller visualization tool that loads the NewsReader data and provides different views and interactions. We show the capacity of data generalization by the tool by applying it to other projects with biograph-

[1] http://www.newsreader-project.eu/

Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism, pages 37–45
Copenhagen, Denmark, September 7, 2017. ©2017 Association for Computational Linguistics

ical data and emotions in Dutch 17th century theater plays in Section 5. Section 2 explains how our work differs from others. Section 6 concludes with future plans.

2 Related work

Interactive graphics have been used for before to analyze high dimensional data (Buja et al., 1996; Martin and Ward, 1995; Buja et al., 1991), but fast, web-based and highly interactive visualizations with filtering are a fairly new development. With the advent of the open source libraries *D3.js* (Bostock et al., 2011), *Crossfilter.js* (Square, 2012) and *DC.js* (dcj, 2016), we now have the tools to rapidly develop custom visual applications for the web.

The *egoSlider* (Wu et al., 2016) uses a similar visualization to our co-participation graph for Egocentric networks. Our visualization can however display co-participation of all participants rather than just for one. The interactive filtering, our other views on the multidimensional data, and the immediate link to the original data are also not present in egoSlider.

The TIARA system (Liu et al., 2012) visualizes news stories as theme rivers. It also has a network visualization of actor-actor relations. This can be used when the corpus consists of e-mails, to show who writes about what to whom. In StoryTeller, the relations are not based on metadata but are relations in the domain of discourse extracted from text.

3 Multi-dimensional event data from text

The NewsReader system automatically processes news and extracts *what* happened, *who* is involved, and *when* and *where* it happened. It uses a cascade of NLP modules including named entity recognition and linking, semantic role labeling, time expression detection and normalization and nominal and event coreference. Processing a single news article results in the semantic interpretation of mentions of events, participants and their time anchoring in a sequence of text.

In a second step, the mention interpretations are converted to an instance representation according to the Simple Event Model (Van Hage et al., 2011) or SEM. SEM is an RDF representation that abstracts from the specific mentions within a single or across multiple news articles. It defines individual components of events such as the action, the participants with their roles and the time-anchoring. A central notion in NewsReader is the event-centric representation, where events are represented as instances and all information on these events is aggregated from all the mentions in different sources. For this purpose, NewsReader introduced the Grounded Annotation Framework (GAF, (Fokkens et al., 2013)) as an extension to SEM through *gaf:denotedBy* relations between instance representations in SEM and their mentions represented as offsets to different places in the texts. Likewise, information that is the same across mentions in different news articles gets deduplicated and information that is different gets aggregated. For each piece of information, the system stores the source and the perspective of the source on the information. The result is a complex multidimensional data set in which events and their relations are defined according to multiple properties (Vossen et al., 2015). The Storyteller application exploits these dimensions to present events within a context that explains them, approximating a story.

The following dimensions from the NewsReader data are used for the visualization. **Event** refers to the SEM-RDF ID: the instance identifier. The **actors** in the news article, which are described using role labels that come from different event ontologies, such as PropBank (Kingsbury and Palmer, 2002), FrameNet (Baker et al., 1998), and ESO (Segers et al., 2015). A **climax** score indicating the relevance of the event (normalized between 0 and 100) for a story. The climax score is a normalized score based on the number of mentions of an event and the prominence of each mention, where early mentions count as more prominent. A **group** label that uniquely identifies the event-group to which the event belongs. In NewsReader, groups are formed by connecting events by topic overlap of the articles in which they are mentioned and by sharing the same actors. Each group also has a **groupScore** which indicates the relevance of the group or storyline for the whole collection. For NewsReader, this is the highest climax score within the group of events normalized across all the different groups extracted from a data set. The group's **groupName** consists of the most dominant topic within the group in comparison with all other groups based on IDF*TF. Event groups are the basis for event-centric story visualizations.

The **labels** represent all the different wordings used to mention the event. The **prefLabel** is the most-frequent label for the event. **Time** refers to the date to which the event is anchored. **Mentions** is a list of mentions of the event in source texts. A mention consists of a **snippet**, the offsets of the label in that snippet (**snippet_char**), the **URI** of the source text, and **char**, the character offsets for the raw text inside the source. Next we show an abbreviated example of a NewsReader event in the JSON format used in Storyteller:

```
{ "timeline":
  "events": [{
    "actors": {
    "actor:": [
      "dbp:European_Union",
      "dbp:United_Kingdom",
      "dbp:The_Times",
      "dbp:Government_of_the_United_Kingdom"
    ]
  },
  "prefLabel": ["stop"],
  "time": "20140622",
  "climax": 89,
  "event": "ev194",
  "groupName": "[Community sanction]",
  "groupScore": "099",
  "labels": [
    "stop",
    "difficulty",
    "pressure"
  ],
  "mentions": [{
    "char": [ "5665", "5673" ],
    "perspective": [{
      "source": "author:FT_reporters"
    }],
    "snippet": [" Sunday Times, said
        they were extremely concerned
        about the UK's difficulties in
        stopping the EU from introducing
        measures that continue to
        erode Britain's competitiveness"],
    "snippet_char": [ 81, 89 ],
    "uri": ["http://www.ft.com/thing/
    f2bc1380-fa32-11e3-a328-00144feab7de"]
  }, {
(...)
```

4 Storyteller

The Storyteller application consists of 3 main *co-ordinated views* (Wang Baldonado et al., 2000): a participant-centric view, an event-centric view and a data-centric view.

User interaction is a key component in the design. The rich text data are too complex to visualize without it, as they contain numerous interconnected events, participants and other properties. Given that the estimated processing capacity for accurate data of human sight is limited to around 500 kbit per second (Gregory and Zangwill, 1987), a large number of connections can make the visual exploration very difficult without filters to slice the data into humanly manageable portions. This section presents the 3 views and describes how filtering can be applied.

The interactivity focuses mostly on analysis through filtering. The user can apply filters by clicking through various components of the charts. These filters are then dynamically applied to the other parts of the application, reducing the amount of data on the screen. This allows a user to *drill-down* into the data, gaining knowledge of its composition in the process.

4.1 Participation-centric view

The participation-centric view (Figure 1), which is our own graph design that we have dubbed a *Co-Participation graph*, is an interactively filterable version of the *XKCD Movie Narrative Charts* (Munroe, 2009). It is placed at the top of the Storyteller page and visualizes the participants of all the events. The major participants are placed on the Y-axis and the events they participate in are placed (as ovals) on a timeline. Each participant's timeline has a different color. If different participants take part in the same event, the lines are bent towards and joint in this event, showing the co-participation through this intersection. Events receive descriptive labels. Hovering the mouse cursor over an event will show further details, such as the mentions of that specific event.

4.1.1 Axis ordering

The X-axis is a timeline stretching between the first of the events shown and the last. The Y-axis is slightly more complicated. The ordering of the participants is of great importance to the legibility of the resulting chart. If this is done improperly, the resulting chart will be cluttered because of many curved lines crossing each other unnecessarily. We solved this legibility problem by ordering the elements in such a way that the participant lines travel straight for as long as possible. We do this by re-ordering the elements on the Y-axis in order of co-appearance on the timeline, from bottom to top. We start by determining the first and bottom-most line. For this, we select the first event on the timeline and determine the participants of this event. We then loop over all events that share these participants in order of appearance on the timeline. Every time a new co-participant is found

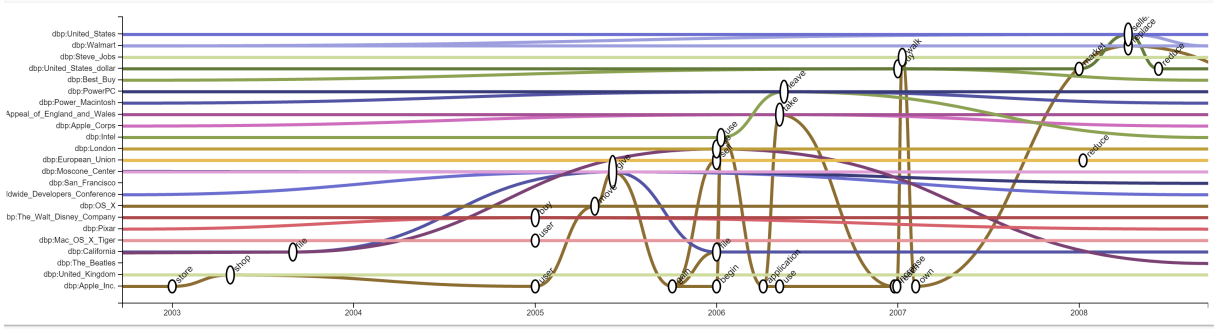


Figure 1: The Co-Participation graph in its unfiltered state, with a NewsReader dataset loaded

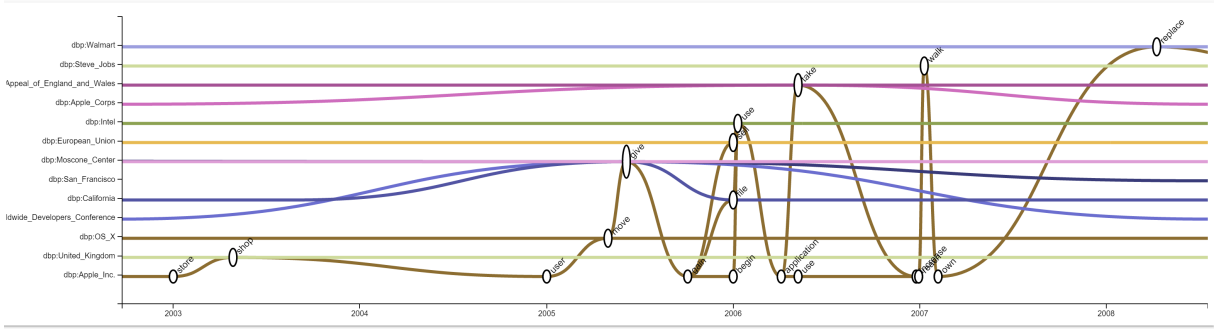


Figure 2: The Co-Participation graph in a filtered state, showing only those events where Apple_Inc. co-participates

in one of these events, it is added to the list. Once all events have been processed in this manner, the algorithm has clustered events that share participants making the resulting graph much easier to read.

4.1.2 Filtering

An alphabetic list of all of the participants present in the dataset is displayed left of the Co-participants graph. The length of the colored bar indicates the number of events in which the participant occurs. This number is shown when hovering the mouse over the bar.

Clicking on one of the items in the participant list applies a filter to the dataset. The Co-participation graph only shows the lines with events that involve the selected participant. This means that the line of the selected participant and those of participants that co-participate with the selected participant are displayed. Selecting more than one participant reduces the graph to those events in which all selected participants co-participate.

4.2 Event-centric view

Figure 3 shows the event-centric view. This second view in the Storyteller demo shows time ordered sequences of events grouped in different rows. The grouping can be determined in different ways and according to different needs.

In NewsReader, each row in the graph approximates the structure of a story, as defined in (Vossen et al., 2015): consisting of one or more climax events, preceded by events that build up towards the climax and following events that are the consequence of the climax defined by prominence (see section 3). Preceding and following events are selected on the basis of co-participation and topic-overlap: so-called bridging relations. The size (and color) of the event bubbles represents the climax score of the event.

A climax event together with all bridged events approximate a story, where we expect events to increase in bubble size when getting closer to the climax event (the biggest bubble in a row) and then gradually decrease after the climax event. The size of the events thus mimics the topical development of for example streaming news, while we still show details on the events within such a topic. The first row presents the story derived from the

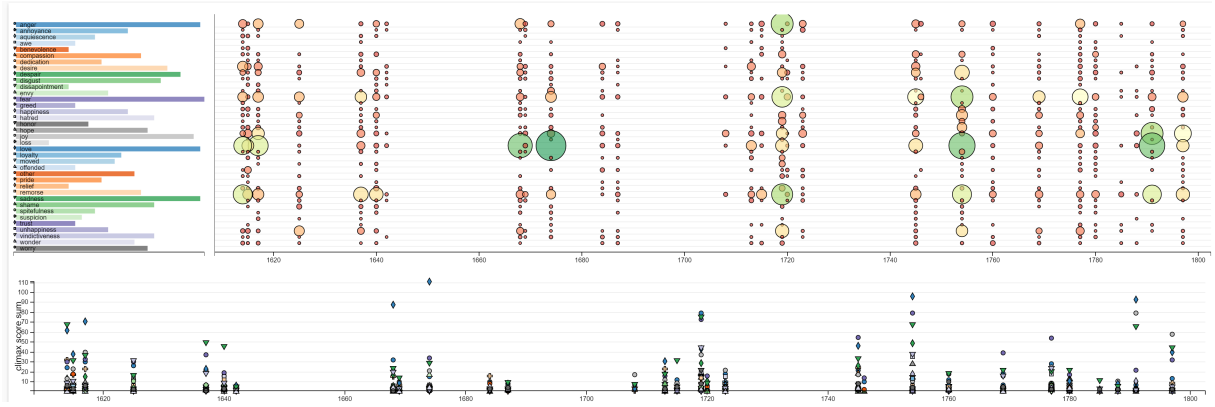


Figure 3: The event-centric view, with data of the Embodied Emotions project (See: Use Cases).

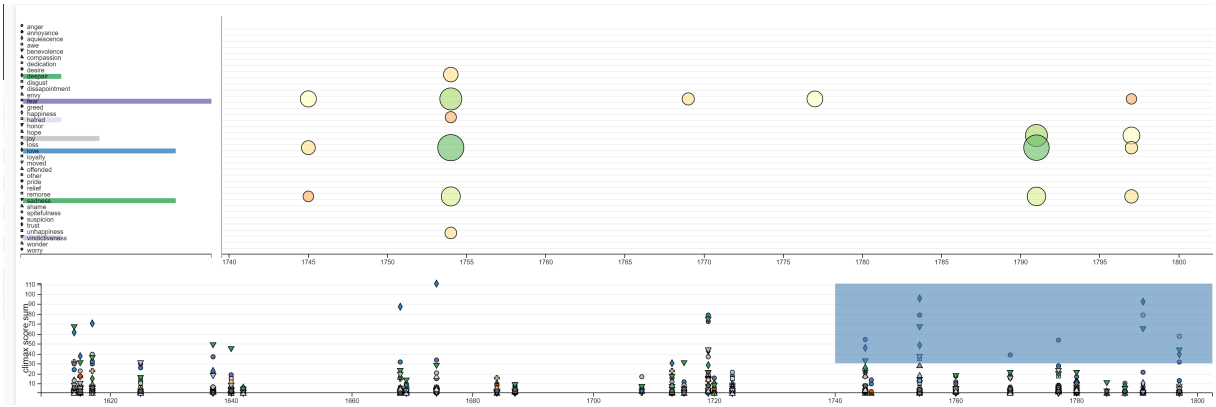


Figure 4: The same event-centric view, with a filter applied to only show high-climax events from the year 1740 and onwards.

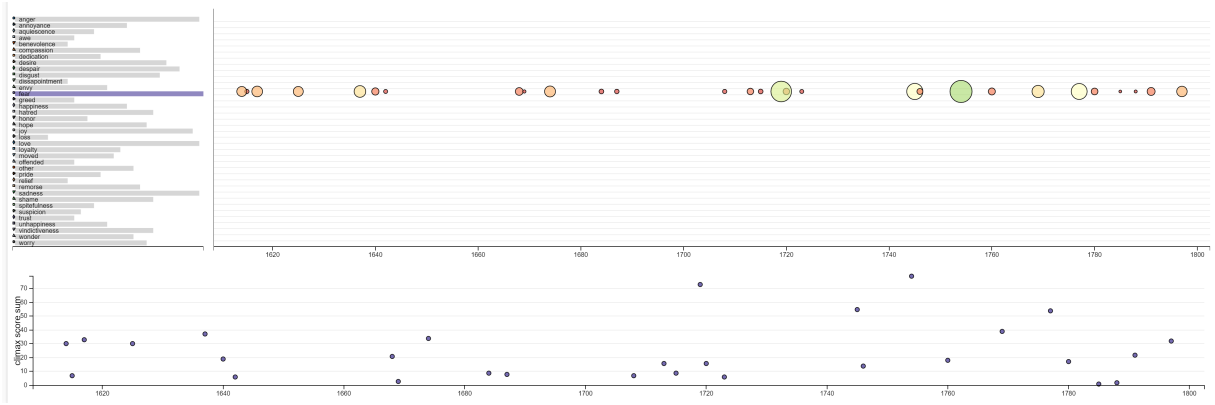


Figure 5: The same event-centric view, with a filter applied to only show events with the fear emotion.

climax event with the highest score (normalized to 100). The next rows show stories based on lower scoring climax events that are not connected to the main story. We thus apply a greedy algorithm in which the most prominent events has the first chance to absorb other events.

Stories are labeled by the climax event's groupName (consisting of the topic that is dominant for the set of events making up the story and the highest climax score within the event). In addition to the label, each row has a colored bar that indicates the cumulative value of the climax scores of all the events within a story. Note that the story with the highest climax event does not necessarily have the largest cumulative score. If an event is mentioned a lot but poorly connected to other events, the cu-

mulative score may still be lower than that of other stories.

The bottom graph of the second view plots individual events for climax score on the Y-axis so that events from the same story may end up in different rows. Group membership is indicated by symbols representing the events. This view shows how events from a story are spread over climax scores and time.

4.2.1 Filtering

The user can select a story by clicking on the index of stories to the left. Unlike the Participation-centric view, selecting more than one group or row adds data to the representation. This is intentionally different from the co-participation graph, where multiple selections intersect (groups can, in this context, by definition not intersect), because stories are separated on the basis of lack of overlap.

In the bottom graph of the event-centric view, the user can make selections by dragging a region in the Y/X space with the mouse. A region thus represents a segment in time and a segment of climax scores. This enables both selecting time intervals (by dragging a full height box between two particular time frames) and selection of the most (or least) influential events, which can be used to exclude outliers in the data but also to select and inspect them.

All filters are applied globally: this means that participant selection in the top view influences the event-centric visualization and vice-versa.

4.3 Data-centric view

At the bottom of the Storyteller page, we see text snippets from the original texts that were used to derive the event data. It lists all events visualized for a given set of filters. Events are presented with the text fragments they are mentioned in and the event word is highlighted. The event labels are given separately as well, where synonyms are grouped together. Furthermore, the table shows the scores, the date and the group name or story label. No selections can be made through this view.

4.4 The full system

The seven tasks of the Visual Information Seeking Mantra

Overview Gain an overview of the entire collection.

Zoom Zoom in on items of interest.

Filter Filter out uninteresting items.

Details-on-demand Select an item or group and get details when needed.

Relate View relationships among items.

History Keep a history of actions to support undo, replay, and progressive refinement.

Extract Allow extraction of sub-collections and of the query parameters.

(Shneiderman, 1996)

We designed Storyteller following the (authoritative) Taxonomy for data visualizations (Shneiderman, 1996), phrased as the 7 tasks of the Visual Information Seeking Mantra:

We initiate the visualization with the **Overview** task presenting the initial (unfiltered) view. The **Filter** and **Zoom** tasks allow the selection of subsets of items in multiple ways through the different views. The **Details-on-demand** task provides detailed mouse-over information boxes as well as a view into the raw data that is maintained while filtering. The co-participation graph supports the main **Relate** task, the filter state displays the **History**, and finally, the data-view allows users to **Extract**.

Storyteller is built to be as generic, reusable and maintainable as possible. We have used only Free and Open Source (FOSS) software web visualization tools and libraries and made Storyteller itself fully FOSS as well. The code is available on github[2] and a demo is also available.[3]

5 Expert Panel Use Cases

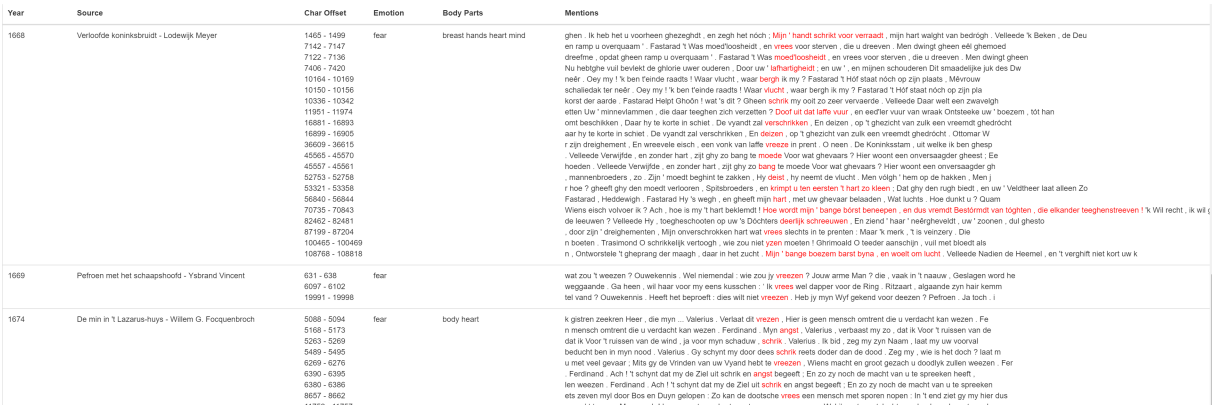
The Storyteller demo has been used in two use cases other than NewsReader. We briefly describe both use cases in this section.

5.1 Embodied Emotions

Embodied Emotions (Zwaan et al., 2015) is a digital history project that investigates the relation

[2] https://github.com/NLeSC/UncertaintyVisualization/

[3] http://nlesc.github.io/UncertaintyVisualization/



Year	Source	Char Offset	Emotion	Body Parts	Mentions
1668	Verloofde koninksbruidt - Lodewijk Meyer	1465 - 1499 7142 - 7147 7122 - 7136 7406 - 7420 10164 - 10169 10150 - 10156 10336 - 10342 11951 - 11974 16881 - 16893 16899 - 16905 36609 - 36615 45565 - 45570 45557 - 45561 52753 - 52758 53321 - 53358 56840 - 56844 70735 - 70843 82462 - 82481 87199 - 87204 100465 - 100469 108768 - 108818	fear	breast hands heart mind	ghen . Ik heb het u voorheen ghezeghdt , en zegh het nóch ; Mijn ' handt schrikt voor verraadt , mijn hart waight van bedrógh . Velleede 'k Beken , de Deu en ramp u overquaam ' . Fastarad 't Was moed'loosheidt , en vrees voor sterven , die u dreeven . Men dwingt gheen eêt ghemoed dreefme , opdat gheen ramp u overquaam ' . Fastarad 't Was moed'loosheidt , en vrees voor sterven , die u dreeven . Men dwingt gheen Nu hebtghe vuil bevlekt de ghlorie uwer ouderen , Door uw ' lafhartigheidt ; en uw ' , en mijnen schouderen Dit smaadelijke juk des Dw neêr . Oey my ! 'k ben t'einde raadts ! Waar vlucht , waar bergh ik my ? Fastarad 't Hóf staat nóch op zijn plaats . Mévrouw schalledak ter neêr . Oey my ! 'k ben t'einde raadts ! Waar vlucht , waar bergh ik my ? Fastarad 't Hóf staat nóch op zijn pla korst der aarde . Fastarad Helpt Ghoôn ! wat 's dit ? Ghneen schrik my ooit zo zeer vervaerde . Velleede Daar welt een zwavelgh etten Uw ' minnevlammen , die daar teeghen zich verzetten ? Doof uit dat laffe vuur , en eedler vuur van wraak Ontsteeke uw ' boezem , tót han omt beschikken , Daar hy te korte in schiet . De vyandt zal verschrikken , En deizen , op 't ghezicht van zulk een vreemdt ghedrócht aar hy te korte in schiet . De vyandt zal verschrikken , En deizen , op 't ghezicht van zulk een vreemdt ghedrócht . Ottomar W r zijn dreighement , En wreevele eisch , een vonk van laffe vreeze in prent . O neen . De Koninkstam , uit welke ik ben ghesp . Velleede Verwijfde . en zonder hart , zijt ghy zo bang te moede Voor wat ghevaars ? Hier woont een onversaagder gheest ; Ee hoeden . Velleede Verwijfde , en zonder hart , zijt ghy zo bang te moede Voor wat ghevaars ? Hier woont een onversaagder gh , mannenbroeders , zo . Zijn ' moedt beghint te zakken , Hy deist , hy neemt de vlucht . Men vólgh ' hem op de hakken , Men j r hoe ? gheeft ghy den moedt verlooren , Spitsbroeders , en krimpt u ten eersten 't hart zo kleen ; Dat ghy den rugh biedt , en uw ' Veldtheer laat alleen Zo Fastarad , Heldewigh . Fastarad Hy 's wegh , en gheeft mijn hart , met uw ghevaar beisaden , Wat luchts . Hoe dunkt u ? Quam Wiens eisch volvoer ik ? Ach , hoe is my 't hart beklemdt ! Hoe wordt mijn ' bange bórst beneepen , en dus vremdt Bestórmdt van tóghten , die elkander teeghenstreeven ! 'k Wil recht , ik wil de leeuwen ? Velleede Hy , toeghescholten op uw 's Dóchters deerlijk schreeuwen , En ziend ' haar ' neêrgheveldt , uw ' zoonen , dul ghesto , door zijn ' dreighementen , Mijn onverschrokken hart wat vrees slechts in te prenten ; Maar 'k merk , 't is veinzery . Die n boeten . Trasimond O schrikkelijk vertoogh , wie zou niet yzen moeten ! Ghrimoald O teeder aanschijn , vuil met bloedt als n , Ontworstele 't gheprang der maagh , daar in het zucht . Mijn ' bange boezem barst byna , en woelt om lucht . Velleede Nadien de Heemel , en 't verghift niet kort uw k
1669	Pefroen met het schaapshoofd - Ysbrand Vincent	631 - 638 6097 - 6102 19991 - 19998	fear		wat zou 't weezen ? Ouwekennis . Wel niemendal : wie zou jy vreezen ? Jouw arme Man ? die , vaak in 't naauw , Geslagen word he weggaande . Ga heen , wil haar voor my eens kusschen : ' Ik vrees wel dapper voor de Ring . Ritzaart , algaande zyn hair kemm tel vand ? Ouwekennis . Heeft het beproeft : dies wilt niet vreezen . Heb jy myn Wyf gekend voor deezen ? Pefroen , Ja toch . i
1674	De min in 't Lazarus-huys - Willem G. Focquenbroch	5088 - 5094 5168 - 5173 5263 - 5269 5489 - 5495 6269 - 6276 6390 - 6395 6380 - 6386 8657 - 8662	fear	body heart	k gistren zeekren Heer , die myn ... Valerius . Verlaat dit vrezen , Hier is geen mensch omtrent die u verdacht kan wezen . Fe n mensch omtrent die u verdacht kan wezen . Ferdinand . Myn angst , Valerius , verbaast my zo , dat ik Voor 't ruissen van de dat ik Voor 't ruissen van de wind , ja voor myn schaduw , schrik , Valerius . Ik bid , zeg my zyn Naam , laat my uw voorval beducht ben in myn nood . Valerius . Gy schynt my door dees schrik reels dooder dan de dood . Zeg my , wie is het doch ? laat m u met veel gevaar ; Mits gy de Vrinden van uw Vyand hebt te vreezen , Wiens macht en groot gezach u doodlyk zullen weezen . Fer . Ferdinand . Ach ! 't schynt dat my de Ziel uit schrik en angst begeeft ; En zo zy noch de macht van u te spreeken heeft , len weezen . Ferdinand . Ach ! 't schynt dat my de Ziel uit schrik en angst begeeft ; En zo zy noch de macht van u te spreeken ets zeven myl door Bos en Duyn gelopen : Zo kan de dootsche vrees een mensch met sporen nopen : In 't end ziet gy my hier dus

Figure 6: The data-centric view, showing the raw data and offsets for the text snippets, as well as the (colored) labels as found in the text. The data displayed here is from 17th century dutch theatre plays

between body parts and emotions in Dutch stage plays over time (i.e. in what body parts do people experience emotions, and through which bodily acts do they express emotions). In our Storyteller visualization, emotions are events and body parts are participants. The visualization can be applied at two levels. First, the X-axis of the Co-participation graph can be used to represent the acts of a single play. Second, the X-axis can be used as a timeline where we plot which emotions and body parts are linked to each other in plays published in a specific year.

Intersections indicate when different body parts are associated with the same emotion. The current online demo shows the second type of visualization: emotion/body-part pairs occurrence per year.[4] The climax score in this visualization reflects the prominence of the emotion (the occurrence of the emotion related to the number of sentences).

5.2 BiographyNet

The BiographyNet (Ockeloen et al., 2013) use case involves the visual exploration of biographies of prominent people in Dutch history. The biographies have been processed by a pipeline similar to the one used for NewsReader. However, the data is radically different: biographies mainly contain events related to the subject of the biography and we do not find many instances of cross-document event co-reference.

The BiographyNet demo[5] therefore does not focus on event instances but on event types: we extract all events related to a person that can be linked to a specific year from the biographical text and meta data. The biography subjects are the participants in our visualization, the event types are the events. If two people are born in the same year, study during the same year or die in the same year, etc., their lines intersect. The BiographyNet visualization thus allows historians to perform generation analyses: what other events have people born in a specific year have in common? The climax score reflects the number of participants involved in an event. The event-centric visualization thus shows which events are prominent in a given year.

6 Conclusion

The fast and efficient web-based visualizations of the output of NLP pipelines allows us to visually analyze complex storylines spanning years. This means that we can now analyze news articles with much more context, be it by looking at co-participants, co-occurrence or other such criteria. This changes literature study from first having to read all the text to build an overview to first having a rough overview that allows you to decide which text to read later.

Storyteller's event-based input format makes it easy to port to new data sources. We described two other projects with different graph-like data (biographies and emotions) that can be served with the same application thanks to this generic input format.

There are many directions in which this research could be advanced in the future. First, the interaction model should be completed. Although Storyteller allows the user to select and filter the data in almost all graphs, other interactivity could

[4] http://nlesc.github.io/EmbodiedEmotions/

[5] http://nlesc.github.io/BiographyNet/

be added. For example, currently, the data table (view 3) only shows the results of filters and selections applied in other views. This table could be made interactive by adding search functionality. This means that the data can be filtered based on user queries. Another possibility for improved interaction is to allow the user to re-order events, participants and groups according to different criteria (e.g., based on frequency, or alphabetically).

While we now have a visualization capable of displaying a decently sized data, it cannot handle the sheer volume of *all* available news data. We are currently implementing an interface that generates manageable data structures on the basis of user queries from a triple store that contains massive event data from the processed news in RDF (i.e. possibly millions of news articles and billions of triples). The user queries are translated into SPARQL requests and the resulting RDF is converted to the JSON input format. This solution requires that some structures, e.g. the climax score and the storylines, need to be computed beforehand. The user should make a visually supported selection *overview, zoom and filter* before querying the database to obtain all required data and displaying the current views.

The NewsReader data exhibits various degrees of uncertainty and source perspectives on the event data (e.g. whether the source believes the event has actually happened, or whether it is a positive or negative speculation or expectation of the source). These are modeled through the RDF representation as well but have not yet been considered in the tool. In the next version of the Storyteller application, we aim to visualize this data layer as well.

Feedback from domain experts who explored data from the three different use cases indicates that a proper understanding of the tool and the data is required in order to get meaningful results. In addition to developing tutorials that help researchers to get the most out of Storyteller, we propose two types of user studies. First, we need to evaluate Storyteller's usability. Second, we need to evaluate to what extent Storyteller generates results (data and views) that are useful for the different domains.

Acknowledgment

The authors would like to thank the Netherlands Organisation for Scientific Research (NWO) and the Netherlands eScience Center for sponsoring this research.

References

2016. DC.js dimensional charting javascript library.

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet project. In *Proc. of ACL*, pages 86–90.

M. Bostock, V. Ogievetsky, and J. Heer. 2011. D^3 data-driven documents. *Visualization and Computer Graphics*, 17(12):2301–2309.

A. Buja, D. Cook, and D. F. Swayne. 1996. Interactive high-dimensional data visualization. *Journal of computational and graphical statistics*, 5(1):78–99.

A. Buja, J. A. McDonald, J. Michalak, and W. Stuetzle. 1991. Interactive data visualization using focusing and linking. In *Proc. of Visualization'91*, pages 156–163. IEEE.

Antske Fokkens, Marieke van Erp, Piek Vossen, Sara Tonelli, Willem Robert van Hage, Luciano Serafini, Rachele Sprugnoli, and Jesper Hoeksema. 2013. GAF: A grounded annotation framework for events. In *The 1st Workshop on Events*, Atlanta, USA.

R. L. Gregory and O. L. Zangwill. 1987. *The Oxford companion to the mind.* Oxford University Press.

P. Kingsbury and M. Palmer. 2002. From treebank to propbank. In *LREC*. Citeseer.

S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian. 2012. Tiara: Interactive, topic-based visual text summarization and analysis. *Transactions on Intelligent Systems and Technology*, 3(2):25.

A. R. Martin and Matthew O. Ward. 1995. High dimensional brushing for interactive exploration of multivariate data. In *Proc. of Visualization'95*, page 271.

R. Munroe. 2009. Xkcd# 657: Movie narrative charts.

N. Ockeloen, A. Fokkens, S. ter Braake, P. Vossen, V. De Boer, G. Schreiber, and S. Legêne. 2013. Biographynet: Managing provenance at multiple levels and from different perspectives. In *Proc. of LISC2013*, pages 59–71. CEUR-WS. org.

R. Segers, P. Vossen, M. Rospocher, L. Serafini, E. Laparra, and G. Rigau. 2015. ESO: A frame based ontology for events and implied situations. In *Proceedings of MAPLEX 2015*, Yamagata, Japan.

Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Symposium on Visual Languages*, pages 336–343.

Square. 2012. Crossfilter.js fast multidimensional filtering for coordinated views.

W. R. Van Hage, V. Malaisé, R. Segers, L. Hollink, and G. Schreiber. 2011. Design and use of the simple event model (sem). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136.

P. Vossen, T. Caselli, and Y. Kontzopoulou. 2015. Storylines for structuring massive streams of news. In *Proc. of CNewS 2015*, Bejing, China.

Piek Vossen, Rodrigo Agerri, Itziar Aldabe, Agata Cybulska, Marieke van Erp, Antske Fokkens, Egoitz Laparra, Anne-Lyse Minard, Alessio Palmero Aprosio, German Rigau, Marco Rospocher, and Roxane Segers. 2016. Newsreader: Using knowledge resources in a cross-lingual reading machine to generate more knowledge from massive streams of news. *Knowledge-Based Systems*, 110:60 – 85.

M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. 2000. Guidelines for using multiple views in information visualization. In *Proc. of AVI*, pages 110–119. ACM.

Y. Wu, N. Pitipornvivat, J. Zhao, S. Yang, G. Huang, and H. Qu. 2016. egoslider: Visual analysis of egocentric network evolution. *Visualization and Computer Graphics*, 22(1):260–269.

J. M. van der Zwaan, I. Leemans, E. Kuijpers, and I. Maks. 2015. Heem, a complex model for mining emotions in historical text. In *Proc. of IEEE eScience*, pages 22–30.