

# Event Detection and Semantic Storytelling: Generating a Travelogue from a large Collection of Personal Letters

Georg Rehm\*, Julian Moreno Schneider\*, Peter Bourgonje\*, Ankit Srivastava\*,  
Jan Nehring\*, Armin Berger<sup>+</sup>, Luca König<sup>+</sup>, Sören Räuchle<sup>+</sup>, Jens Gerth<sup>+</sup>

DFKI GmbH, Language Technology Lab\*  
Alt-Moabit 91c, 10559 Berlin, Germany

3pc GmbH Neue Kommunikation<sup>+</sup>  
Prinzessinnenstraße 1, 10969 Berlin, Germany

Corresponding author: georg.rehm@dfki.de

## Abstract

We present an approach at identifying a specific class of events, movement action events (MAEs), in ca. 2,800 personal letters exchanged by the German architect Erich Mendelsohn and his wife, Luise. A backend system uses these and other semantic analysis results as input for an authoring environment that curators can use to produce new pieces of content. The human expert will receive recommendations from the system with the goal of putting together a travelogue, i. e., a description of the trips and journeys undertaken by the couple. We describe the components and also apply the system to news data.

## 1 Introduction

Robust event detection coupled with text analytics can lead to a multitude of innovative solutions to contribute to the decades-old “information overflow” challenge, but also to address more specialised, sector-specific needs. While many researchers concentrate on identifying meaningful stories, story paths or storylines in collections of news documents we propose an approach that bundles a flexible set of semantic services for the *production of digital content*, especially to recommend interesting storylines to human experts who process large collections of documents. We call this approach *Semantic Storytelling*.

The activities reported in this paper are carried out in the context of the research and technology transfer project Digital Curation Technologies, in which a research centre collaborates with four SME companies that operate in four sectors. We develop and deploy, in prototypically implemented use cases, a flexible platform that provides generic curation services such as, e. g.,

summarisation, named entity recognition, entity linking and machine translation (Bourgonje et al., 2016a,b). These are integrated into the in-house systems of the partner companies and customised to their domains so that the knowledge workers, journalists, experts, museum planners and digital curators who use these systems can do their jobs more efficiently, more easily and with higher quality. Their tasks involve the processing, analysis, skimming, sorting, summarising, evaluating and making sense of large amounts of digital content, out of which a new piece of digital content is created, e. g., an exhibition catalogue, a news article or an investigative report. The curation technology platform is meant to simplify the content curation task significantly.

This paper is structured as follows: Section 2 describes the Semantic Storytelling use case in more detail, i. e., the authoring environment and the data set. Section 3 focuses upon the approach, defines Movement Action Events (MAEs), and describes the curation services, e. g., temporal analysis, entity recognition, and event detection. Section 4 sketches the results of initial experiments on news data, while Section 5 summarises related work. Section 6 concludes the paper.

## 2 Use Case: Semantic Storytelling

The generic Semantic Storytelling use case involves processing a coherent and self-contained collection of documents in order to identify and to suggest, to the human expert, one or more potential story paths that can then be used to structure an actual story around them or, generally, a new piece of content (Schneider et al., 2016). One example are millions of leaked documents, in which an investigative journalist wants to find the interesting nuggets of information, i. e., surprising relations between different entities, say, politicians and off-

shore banks. The semantic technologies involved do not necessarily have to exhibit perfect performance because, in our use cases, humans are always in the loop. We want to provide, ideally, robust and generic technologies with broad coverage. For some services this goal can be fulfilled while for others, it must be considered ambitious.

## 2.1 Smart Authoring Environment

One of the partner companies is currently designing and developing an authoring environment, enabled by the curation technology platform and its semantic services.<sup>1</sup> Many of its projects involve a client, e. g., a company, a museum or a political party, that approaches the company with a set of digital content and a rough conception how to structure and visualise these assets in the form of a website or app. An authoring environment that can semantically process such a collection to enable the efficient authoring of flexible, professional, convincing, visually appealing content products that provide engaging stories would significantly reduce the effort on the side of the agency and, at the same time, improve their flexibility. From the same set of semantically enhanced content different output formats could be generated (e. g., web app, iOS or Android app, ebook etc.). Example screens of the authoring environment’s user interface (“Redaktionstool” in German) are shown in Figure 3. With regard to the look and feel, it was a conscious design decision to move beyond the typical notion of a “web page” that is broken up into different “modules” using templates. The clear focus are engaging stories told through the content.

With this tool the curator can interactively put together a story based on the content that has previously been enriched through the curation services and that act as building blocks. Figure 3 shows examples from the set of ca. 2,800 letters exchanged between the German architect Erich Mendelsohn (1887-1953) and his wife Luise, both of whom travelled frequently. We decided to focus upon the use case of identifying all *movement action events*, i. e., all trips undertaken by the author of the respective letter from location A to location B using a specific mode of transport. We want to construct, ideally automatically, a *travelogue* from this analysis layer, that provides an engaging story to the reader and that also enables ad-

<sup>1</sup>This company, 3pc GmbH, is a digital agency, founded in 1995, that has completed more than 2,000 projects.

ditional modes of access, e. g., through map-based or timeline-based visualisations. The goal is to process multiple interconnected instances of the text type *letter* in order to generate one instance of the text type *travelogue*.

## 2.2 Data Set: The Mendelsohn Letters

The collection contains 2,796 letters, written between 1910 and 1953, with a total of 1,002,742 words (avg. number of words per letter: 358.6, incl. addresses) on more than 11,000 sheets of paper; 1,410 of the letters were written by Erich and 1,328 by Luise Mendelsohn.<sup>2</sup> Most are in German (2,481), the rest is written in English (312) and French (3). The letters were scanned, transcribed and critically edited; photos and metadata are available. This research was carried out in a project that the authors of the present paper are not affiliated with (Bienert and de Wit, 2014). In the letters the Mendelsohns discuss their private and professional lives, their relationship, meetings with friends and business partners, and also their travels. One result of (Bienert and de Wit, 2014) is an online version of the Mendelsohn collection. In the present project we explore to what extent it is possible to automate the production of an online version of an arbitrary document collection.

## 3 Approach

We attempt to detect movement events to generate the backbone of a travelogue. Typically, in linguistics, the definition of “event” (vs. “state”) is so broad and implicit that it is, for the time being, not feasible to implement a corresponding general-purpose event detection system. In NLP, on the other hand, events are usually defined as words or phrases (typically verbs, sometimes nouns) that clearly signal, on the linguistic surface, the existence of a specific action, activity, or change of state. Event detection is related to information and relation extraction (IE, RE). While IE and RE are focused on specific relations or template-like IE, event detection is more general. As open domain event detection is not feasible yet, we focus on Movement Action Events (MAEs). With regard to the text type “letter”, an MAE mention relates to a currently happening or upcoming trip or journey announced or mentioned in a letter. A few examples, taken from two letters from Erich to Luise,

<sup>2</sup>There are also several duplicates and letters without any textual content in the collection.

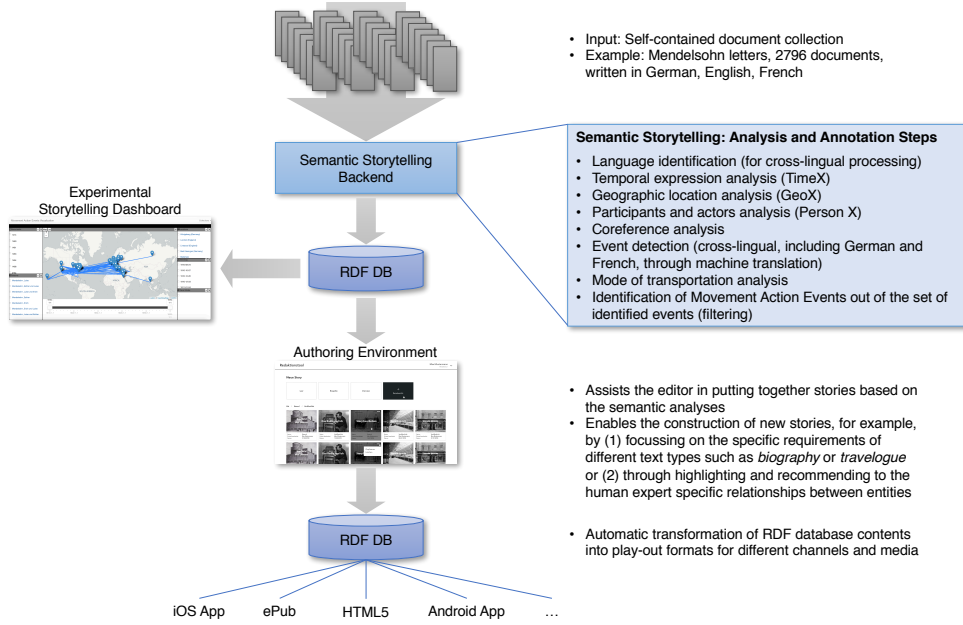


Figure 1: The Semantic Storytelling architecture and workflow

written on March 14, 1944, and March 10, 1949, respectively:

“The hectic days of St. Louis, my beloved, are drawing to their close. *I am leaving tonight for Davenport.*”

“Temple Washington affair promising. Have been there on Tuesday night from 9.30 to 1, returned to Baltimore at 2 A.M. [...] *Due in St. Louis around midnight.*”

MAEs imply physical motion events that occur when a person is travelling from one location (e. g., town, city) to another using a medium or long distance mode of transport. An MAE consists of the six-tuple  $MAE = \langle P, L_O, L_D, t_d, t_a, m \rangle$  with  $P$  a reference to the participant (E. or L. Mendelsohn),  $L_O$  and  $L_D$  references to the origin and destination locations (named locations, GPS coordinates),  $t_d$  and  $t_a$  the time of departure and arrival and  $m$  the mode of transport. Each component is optional as long as the MAE contains at least one participant and a destination. If multiple people travel together  $P$  can refer to a set of persons. For consecutive MAEs, we assume that  $L_D$  is  $L_O$  of the next trip:

$$MAE_1 = \langle P, L_a, L_b, t_i, t_j, m_x \rangle$$

$$MAE_2 = \langle P, L_b, L_c, t_k, t_l, m_y \rangle$$

$$MAE_3 = \langle P, L_c, L_d, t_m, t_n, m_z \rangle$$

We detect MAEs through triggers, locations, temporal expressions, participants and the mode of transport. Out of the instantiated sets of six-

tuples we attempt to construct a travelogue as a list of six-tuples (see Figure 1).

Many researchers working on, among others, text linguistics have emphasised the relationship between generalised text structure patterns and their respective text types or genres. Recently, (Caselli and Vossen, 2016) proposed the Storyline Annotation and Representation Scheme, which is primarily aimed at news articles to “identify salient events (climax events) as the central elements around which a specific topic develops”. With regard to the travelogue example, the notion of one “climax event”, “rising actions” and “falling actions” is not applicable, also see (Pang et al., 2011; Ye et al., 2011). Storyline applications tailored to specific text types (*news articles* vs. *letters* and *travelogue*) have different requirements regarding their storyline abstraction models. Accordingly, we focus on the identification of consecutive instantiations of MAE six-tuples.

### 3.1 Temporal Expressions

We use two tools for extracting temporal expressions: TimeX and HeidelTime. TimeX is our own implementation for recognising and normalising temporal expressions. It is based on a regular expression grammar and available for English and German. TimeX covers concrete (“11th of March, 2014”) and relative mentions (“last week”). All expressions are normalised into a machine-readable format.

	TimeX			HeidelTime		
WikiwarsDE	P: 0.72	R: <b>0.90</b>	F: 0.80	P: <b>0.98</b>	R: 0.85	F: <b>0.91</b>
Mendelsohn letters	P: <b>0.91</b>	R: <b>0.60</b>	F: <b>0.72</b>	P: 0.71	R: 0.44	F: 0.54

Table 1: Comparison of the performance of TimeX and HeidelTime

A typical date notation used in the letters is “12.IV.26” (“12 April 1926”), with roman-style numerals for the number of the month; the extraction grammar can be adapted to cover alternative notations. A brief comparison of the performance of TimeX and HeidelTime on two data sets is shown in Table 1. TimeX achieves reasonable results but is outperformed by HeidelTime (Strötgen and Gertz, 2010) on the general domain corpus. TimeX scores better on the Mendelsohn collection. The WikiwarsDE corpus (Strötgen and Gertz, 2011) consists of German documents describing military conflicts. Customising HeidelTime’s grammar requires significant modifications on different levels; having direct control over our own system worked better for us regarding the Mendelsohn collection and other data sets. After recognising and normalising temporal expressions, we also want to position documents on a timeline. This requires calculation of the average time stamp of a document including the spread over the timeline. The time stamp is computed on the basis of average milliseconds before or after java epoch (1st January 1970); standard deviation is also calculated. For the Mendelsohn experiments we use TimeX. For processing general domain texts and languages not covered by TimeX, we integrated HeidelTime into our platform.

### 3.2 Geolocations

Our geolocation extraction tool, GeoX, is based upon the OpenNLP NameFinder (Apache Software Foundation, 2016) trained on Wikipedia locations (Nothman et al., 2012). After the identification of locations we use DBpedia Spotlight or a domain-specific ontology (GeoNames for the Mendelsohn experiments) to retrieve a URI for every location entity. Once a URI is available, latitude and longitude can be obtained. Similar to TimeX, the average latitude and longitude value is calculated for every document, so that documents (rather than locations mentioned in them) can be pinpointed on a map. Adaptability to new domains is an important requirement. In addition to a general model, we allow uploading key-value-based

dictionaries for pattern-based entity spotting. The key is the pattern to look for, the value a URI in an ontology. If it allows SPARQL queries, we can include ontology-specific queries to retrieve related information (e. g., latitude, longitude, country etc.). For the Mendelsohn experiments we had access to a database that includes a list of location names and their GeoNames URIs. Table 2 shows GeoX’s performance using the Wikipedia model, based on 10-fold cross-validation on part of the data from (Nothman et al., 2012) using 120,000 sentences with 101,540 locations.

	GeoX	PersonX
Precision	93.68	96.89
Recall	69.50	74.00
F-score	79.80	83.91

Table 2: Performance of GeoX and PersonX

### 3.3 Participants and Actors

Similar to GeoX, we implemented a tool (PersonX) for extracting persons by training a corresponding model. For the general model, the same data is used as for the location model as it was also annotated for person-type entities. We also perform entity linking to retrieve an ontology URI (DBpedia by default, unless a domain-specific ontology is plugged in). For the Mendelsohn experiments we had access to a list of persons linked to a URI at Deutsche Nationalbibliothek. Table 2 shows evaluation results using the same procedure as for the location model, using 120,000 sentences containing 56,086 persons.

### 3.4 Crosslingual Event Detection

The Mendelsohn data set is multilingual with the majority of the letters written in German. Most of our processing tools are language dependent, several are available for English only. Therefore, we implemented a crosslingual event detection system, i. e., translating German and French documents into English through Moses Statistical Machine Translation (Koehn et al., 2007) and detect-

ing events in the translated documents. We implemented a dedicated pre-processing module for cleaning the German letters before we were able to send them to the MT engine. Approximately 30% of the words remained untranslated but an analysis showed these to be mainly named entities (people, locations) and abbreviations. The documents were then processed by the event detection system.

### 3.5 Generic Event and MAE Detection

We implemented a state-of-the-art event extraction system based on (Yang and Mitchell, 2016) to pinpoint words or phrases in a sentence that refer to events involving participants and locations, affected by other events and spatio-temporal aspects. The system is trained on the ACE 2005 data (Doddington et al., 2004), consisting of 529 documents from a variety of sources (newswire reports, blogs, discussion forums). We apply the tool to extract generic events in an ACE 2005 test set (30 news documents consisting of 672 sentences with 4,184 entity mentions and 438 triggers) and to detect MAEs in the Mendelsohn letters.

After processing the Mendelsohn letters, the English data set consisting of 295 documents and 7,899 sentences yielded 1,600 event triggers. The German (translated into English, see Section 3.4) data set consisting of 2,450 documents and 76,350 sentences yielded 6,950 event triggers. For MAE detection, the most relevant event type is the ACE “Transport” event. According to the ACE guidelines<sup>3</sup> a transport event occurs whenever an entity (person, vehicle, weapon) is moved from one place (GPE, facility, location) to another; a Transport Event contains seven slots (agent, entity, vehicle, price, origin, destination, time). Circa 45% and 40% of the labelled events in the English and German Mendelsohn letters respectively are Transport events. After detection, the events are passed to the next step in the workflow.

(Yang and Mitchell, 2016) decompose the learning problem into three subproblems: learning within-event structures, learning event-event relations, and learning for entity extraction. These learned models are then integrated into a single model that performs joint inference of all event triggers, semantic roles for events, and also entities across the whole document. With a precision of 82.4, recall of 79.2 and F-score of 80.8

<sup>3</sup>ACE English Events guidelines, <https://www ldc.upenn.edu/sites/www ldc.upenn.edu/files/english-events-guidelines-v5.4.3.pdf>

we achieve comparable results to those reported by (Yang and Mitchell, 2016). As there is no gold standard available for the Mendelsohn data set, we manually evaluated a small subset and discovered that several events could not be detected due to data formatting issues and the fact that the system is trained on news documents from the early 21st century. After normalising the statistics and comparing with the ACE 2005 test data, we found that the Mendelsohn data set (out-of-domain) yielded 5 times and 7 times less events in the English and German letters than in the ACE 2005 test data.

### 3.6 Mode of Transportation

In the MAE six-tuple,  $m$  refers to the mode of travel, e. g., plane, train, car etc. An obvious approach is to look for linguistic cues, i. e., for corresponding nouns in sentences like “Tomorrow I’ll go to New York by *train*”. Often, the event’s trigger verb provides the mode (“I’m *flying* to Los Angeles tonight.”). For these two sets of cues, we can rely on a set of rules to cover all means of transportation. If there is no linguistic evidence available, we can attempt to deduce the mode. As we retrieve a URI for locations we can also retrieve related geographical location information using SPARQL. Using latitude and longitude of the origin and destination, we can calculate the distance using Vincenty’s formulae.<sup>4</sup> From the distance, we attempt to deduce the mode using a set of threshold values. For short trips (from San Francisco to Palo Alto, say), typically the train, bus or car is used, but not a plane. We can also divide the time difference between departure and arrival and deduce the mode. For distances of more than 5,000km and a time of less than 10 hours, a plane is likely. For trips of more than 3,000km spanning different continents and taking more than a week, a cruise ship is more likely. Based on this approach we can identify 369 modes of transportation in the (English) Mendelsohn letters and 5,152 in the Obama corpus (Section 4).

### 3.7 Instantiation of MAE Six-Tuples

The following approach iterates over all documents. First, temporal expressions (Section 3.1), geolocations (Section 3.2), participants (Section 3.3) and trigger elements are annotated; we use two types of trigger elements, a motion-type

<sup>4</sup>[https://en.wikipedia.org/wiki/Vincenty%27s\\_formulae](https://en.wikipedia.org/wiki/Vincenty%27s_formulae)

verb class and a list of modes of transport (Section 3.6).<sup>5</sup> Afterwards, event detection is performed (Section 3.5). Finally, we filter for MAEs. This algorithm operates on the sentence level; for this we segment the letters into individual sentences. A rule set determines if an event is a MAE:

- 1) If a general candidate event does not contain a trigger element it is deleted.
- 2) If the event does not contain a participant, location, or temporal expression, we include, in the six-tuple, the author, location, or date – as noted by the author in the letter head – as  $P$ ,  $L_O$  or  $t_d$  of the MAE candidate.
- 3) We generate all combinations of MAE candidate six-tuples by filling the six-tuple with the available entities. Every candidate receives a score that is computed as a weighted linear combination of the existence of the six-tuple components:

$$\begin{aligned}
 sc_{MAE} = & w_P * sc_P + \\
 & w_{L_O} * sc_{L_O} + \\
 & w_{L_D} * sc_{L_D} + \\
 & w_{t_d} * sc_{t_d} + \\
 & w_{t_a} * sc_{t_a} + \\
 & w_m * sc_m
 \end{aligned} \tag{1}$$

where  $sc_i$  is the score of the  $i^{th}$  feature (in this case these scores are always 1),  $w_i$  is the weight of the  $i^{th}$  feature and  $\sum_i w_i = 1$ .

- 4) The MAE candidates with a score greater than a certain threshold  $th$  are processed further.

For the evaluation we use a quantitative and a qualitative measurement: the number of MAEs annotated and a manual evaluation of the MAEs of some randomly selected documents. We apply five different approaches to generate MAE candidates: (**A1**) using all entities available in a candidate event; (**A2**) like A1 but also including the metadata of the letters as entities (author, location, date); (**A3**) using all entities available in a candidate event but avoiding similar locations for  $L_O$  and  $L_D$  as well as similar dates for  $t_d$  and  $t_a$ ; (**A4**) like A3 but also including the metadata of the letters as entities; (**A5**) like A3 but only including the MAEs that appear in sentences that also include a trigger element. The number of MAE candidates in the Mendelsohn letters are shown in Table 3.

The approaches that include the metadata of the letters generate much more MAE candidates. This is to be expected because the inclusion of the

	th=0	th=.25	th=.5	th=.75	th=1
A1	591	328	98	0	0
A2	6386	4831	3554	736	0
A3	563	253	54	0	0
A4	5640	3166	1260	53	0
A5	116	60	11	0	0

Table 3: Generating MAE candidates

metadata makes three entities (person, date, location) available in each candidate. We tried the approaches including the letters’ metadata because the author often uses “I” instead of her/his name, of course, which is why the author is often not included as an extracted entity. All candidates that do not make sense have to be filtered in a post-processing step. We tried to determine the best threshold value by using five values between 0 and 1. The respective score is directly related to the amount of features they are composed of: the higher the number of included features, the higher the score. This is why the different thresholds can be seen as a “proof” of the number of MAE candidates including the needed amount of information.

We also performed a qualitative evaluation selecting randomly 10 MAE candidates. While 9 out of 10 inspected candidates were extracted correctly and refer to proper MAEs, the instantiation of the six-tuples (esp.  $t_d$  and  $t_a$ ) needs further improvement: 5 correct departure times, 1 correct arrival time, 3 correct persons, 8 correct origins, 1 correct destination and 0 correct transportation modes.

George Downs, Santa Cruz, Carmel, [], [], [].

*[George Downs will pick us up tomorrow at 9.30 a.m. and we intend to drive skyline to Los Gatos to see Kate Ostwald for a moment and then via Santa Cruz [sic] to Carmel!]*

A general problem is the huge number of MAE candidates, much higher than the actual number of genuine complete MAEs, due to the combination of all possible entities existing in an event. Some common errors appear in many MAEs candidates. Sometimes, regarding departure and arrival time, the current time (i. e., execution time/date) is used, because the date is underspecified and the anchor year “now” is used. In some cases the arrival times are *before* the departure time, which can be taken care of easily by making the instantiation algorithm time-aware. In some cases we had false MAE positives due to misinterpreted triggers

<sup>5</sup>The verb cues are based on (Levin, 1993), Chapter 51.

such as, for example, “Drive” (referring to street names). These errors are more common on MAE candidates with higher scores because they contain more features, even if some of them are incorrect. Some MAE candidates with a lower score have better features, or a higher number of correct features. In the following we present two MAE candidates that are correct MAEs with less features. With the inclusion of additional metadata from the letters the results could be improved considerably because in both cases the subject “I” was not identified as an entity and, thus, not included in the six-tuple (see above with regard to the incorrectly identified arrival and departure times).

[], [], Cleveland, Sat May 06 12:00:00 CEST 2017, [], [].

[*My discussion here will, I hope, be finished before I leave for Cleveland tomorrow night.*]

[], [], New York, , [], [].

[*I left Sunday – soon after the pleasant meeting – for New York.*]

We also performed more traditional relation extraction experiments by using the Stanford Dependency Parser to extract relation triples (subject, verb, object) to collect the information for filling the six-tuple slots. In the dependency graph we extract sub-root level nodes (typically verbs) that connect two noun phrases or other candidates. The extracted relations are then filtered for motion-type verbs (Levin, 1993). Typically, the subject would be the  $P$  in the MAE, and the object of the relation any of the other slots ( $L_O$ ,  $L_D$ ,  $t_d$ ,  $t_a$ ,  $m$ ). Applying this approach on the English subset of the Mendelsohn collection resulted in only 10 triples that met the criteria of having a movement action at the core of the relation.<sup>6</sup>

### 3.8 Semantic Storytelling Dashboard

To get a better understanding of the data set, the analysis results, the extracted MAEs and to prepare attaching the Semantic Storytelling backend to the authoring environment (see Figure 1), we implemented an experimental dashboard (Figure 2). The upper left window shows a list of the documents in the data set; extracted MAE candidates, visualised in the map, can be filtered by document. The bottom left window shows the list of annotated named entities. The map visualises the locations involved in the MAE candidates with

<sup>6</sup>This approach performed better on the Obama news corpus (Section 4).

highlighted annotations. The slider below the map can be used to filter MAEs by time. The windows on the right hand side show all location names, temporal expressions and modes of transportation. Additional details and case studies can be found in (Rehm et al., 2017; Schneider et al., 2017)

## 4 MAE Detection in News Data

Our primary data set in this paper is the Mendelsohn collection but we also see multiple application scenarios for the news domain – the Semantic Storytelling backend and authoring environment are meant to be applied to arbitrary data sets after all. We performed an initial evaluation of our system applied to a data set that consists of news articles on the multiple trips of Barack Obama.<sup>7</sup> The corpus contains 487 files with 24,387 sentences and 897,630 tokens. We annotated 17,241 persons, 21,569 locations, 19,572 temporal expressions, 5,104 transport modes and 3,537 trigger verbs. The event extraction system annotated 61,718 entity mentions and 6,752 event triggers, 31% of which were “Transport” events. We found that in-domain data (the Obama data set) produced three times more event triggers than out-of-domain data (Mendelsohn letters). We plan to close this gap through domain adaptation.

For the evaluation we applied three of the approaches mentioned in Section 3.7: (A1) using all entities available in a sentence; (A3) using all entities in a sentence but avoiding similar locations in  $L_O$  and  $L_D$  and similar dates in  $t_d$  and  $t_a$ ; and (A5) the same as A3 but only including the MAEs that appear in sentences that also include a trigger element (see Table 4).

	th=0	th=.25	th=.5	th=.75	th=1
A1	13030	9700	5314	0	0
A3	7841	4511	2784	0	0
A5	2545	1768	1328	0	0

Table 4: Generating MAE candidates (Obama)

While, in our manual evaluation, many MAE candidates turned out to be genuine MAEs, we also found instances of false positives, which contained information extracted from non-article contents such as, for example, the imprint and copy-

<sup>7</sup>Based on a list of links to news articles in [https://en.wikipedia.org/wiki/List\\_of\\_international\\_presidential\\_trips\\_made\\_by\\_Barack\\_Obama](https://en.wikipedia.org/wiki/List_of_international_presidential_trips_made_by_Barack_Obama)

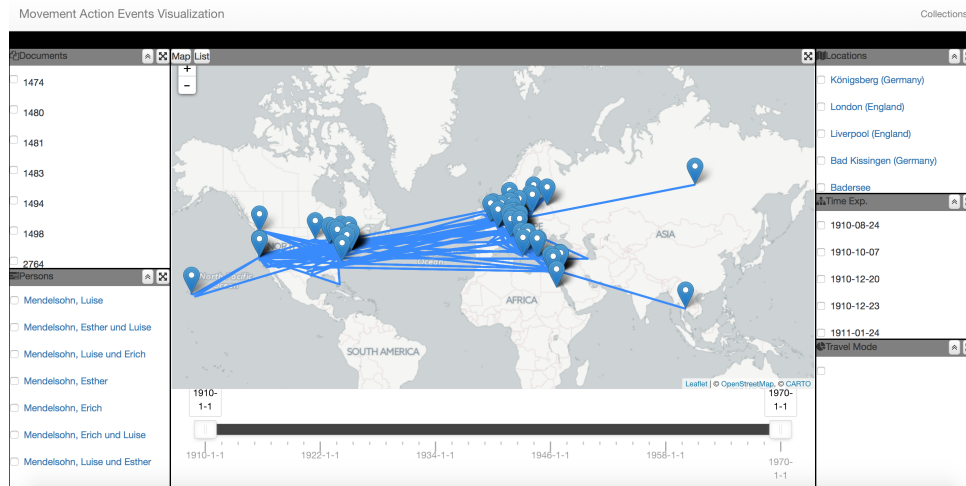


Figure 2: The storytelling dashboard showing Movement Action Events annotations

right information; we tried to remove all HTML boilerplates and templates using a dedicated tool but in some instances these pieces of text were kept. Sometimes, organisations were incorrectly annotated as person entities, which lead to several incorrect MAEs. In some cases the locations used for the six-tuple were too generic (e. g., continent names). Nevertheless, many candidates are genuine MAEs, for example:

Obama, Brasilia, Rio de Janeiro, [], [], [].

[*Mr Obama arrived in Rio de Janeiro after a day of talks in the capital, Brasilia, with Ms Rousseff and business leaders.*]

## 5 Related Work

Most approaches in the event detection literature are machine learning-based and adhere to a modular approach (Ahn, 2006), i. e., they use the output from constituency and dependency parsers, named entity recognisers, coreference resolution systems, and part-of-speech taggers to build classifiers for subtasks of trigger labelling and argument labelling. However, recently, state-of-the-art results have been achieved by joint entity and event extraction systems (Yang and Mitchell, 2016; Li et al., 2013), i. e., approaches which compute joint inference in one combined model to minimise the errors introduced by sub-modules.

Several approaches are related to our Semantic Storytelling concept, all of them concentrating on their own objectives and providing solutions for their respective challenges. A few systems focus on providing content for entertainment purposes (Wood, 2008). Other researchers focus on specific domains, for example, storytelling in gaming

(Gervás, 2013), for recipes (Cimiano et al., 2013; Dale, 1989) or for weather reports (Belz, 2008; Goldberg et al., 1994; Reiter et al., 2005; Turner et al., 2006), requiring knowledge about characters, actions, locations, events, or objects that exist in this particular domain (Gervás et al., 2005; Riedl and Young, 2010; Turner, 2014). The most closely related approach is the one developed by (Poulakos et al., 2015), which presents “an accessible graphical platform for content creators and even end users to create their own story worlds, populate it with smart characters and objects, and define narrative events that can be used by existing tools for automated narrative synthesis”.

## 6 Summary and Future Work

We present an approach at identifying a specific class of events, movement action events, in the Mendelsohn data set. The goal is to expose these and other semantic analysis results through the Semantic Storytelling backend to an authoring environment that curators can use to produce new pieces of content based on this data collection. The authoring environment can provide recommendations, ideas, suggestions or potential story paths to the human expert, in this case, with the goal of producing a travelogue, i. e., a vivid description of the multiple trips and journeys undertaken by the Mendelsohns.

The evaluations show that the task of processing the Mendelsohn data set to identify MAEs is an ambitious challenge. This is especially due to the rather old-fashioned, highly abbreviated, partially poetic, spoken-style language employed and



also due to the fact that most actual MAE mentions are contained only implicitly, making their automatic extraction difficult. Initial results from applying our system to the Obama corpus are more promising as MAEs are contained in news articles in a more explicit way. We assume that our approach can be applied to contemporary news documents more effectively than to personal letters that are, partially, almost 100 years old and belong to a genre and register that is notoriously difficult to process automatically.

In terms of future work, we will connect the storytelling backend to the authoring environment and we will integrate additional components to arrive at an integrated working prototype.

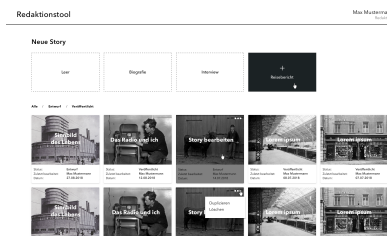
## Acknowledgments

The project “Digitale Kuratierungstechnologien” (DKT) is supported by the German Federal Ministry of Education and Research (BMBF), “Unternehmen Region”, instrument Wachstumskern-Potenzial (no. 03WKP45). More information: <http://www.digitale-kuratierung.de>. The authors would also like to thank the anonymous reviewers for their valuable comments.

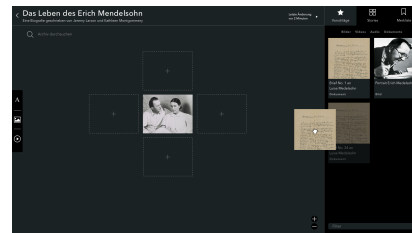
## References

- David Ahn. 2006. The stages of event extraction. In *Proc. of the Workshop on Annotating and Reasoning About Time and Events (ARTE 06)*. ACL, Stroudsburg, PA, USA, pages 1–8.
- Apache Software Foundation. 2016. [Apache OpenNLP](http://opennlp.apache.org). <http://opennlp.apache.org>.
- Anja Belz. 2008. Automatic Generation of Weather Forecast Texts Using Comprehensive Probabilistic Generation-space Models. *Nat. Lang. Eng.* 14(4):431–455.
- Andreas Bienert and Wim de Wit, editors. 2014. *EMA – Erich Mendelsohn Archiv. Der Briefwechsel von Erich und Luise Mendelsohn 1910-1953*. Kunstbibliothek – Staatliche Museen zu Berlin and The Getty Research Institute, Los Angeles. With contributions from Regina Stephan and Moritz Wullen, Version March 2014. <http://ema.smb.museum>.
- Peter Bourgonje, Julian Moreno-Schneider, Jan Nehring, Georg Rehm, Felix Sasaki, and Ankit Srivastava. 2016a. Towards a Platform for Curation Technologies: Enriching Text Collections with a Semantic-Web Layer. In H. Sack, G. Rizzo, N. Steinmetz, D. Mladeni, S. Auer, and C. Lange, editors, *The Semantic Web*. Springer, number 9989 in Lecture Notes in Computer Science, pages 65–68. ESWC 2016 Satellite Events. Heraklion, Crete, Greece, May 29 – June 2, 2016 Revised Selected Papers.
- Peter Bourgonje, Julin Moreno Schneider, Georg Rehm, and Felix Sasaki. 2016b. Processing Document Collections to Automatically Extract Linked Data: Semantic Storytelling Technologies for Smart Curation Workflows. In A. Gangemi and C. Gardent, editors, *Proc. of the 2nd Int.*

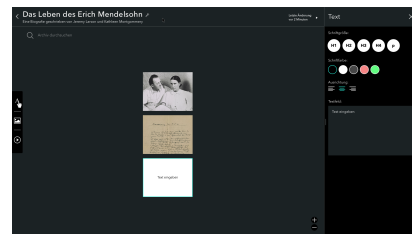
*Create new story from a document collection:*



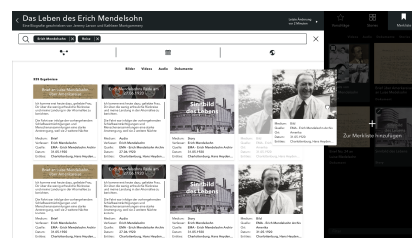
*Dragging and dropping content into the story:*



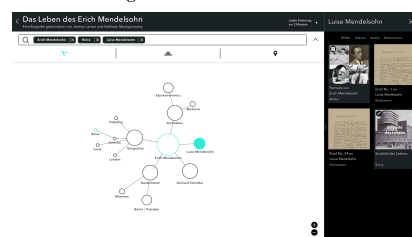
*Annotating and arranging the story:*



*Searching content pieces:*



*Examining relations between entities:*



*The final story, ready to be deployed:*

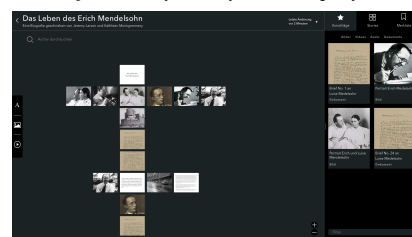


Figure 3: The smart authoring environment

- Workshop on Natural Language Generation and the Semantic Web (WebNLG 2016). ACL, Edinburgh, UK, pages 13–16.
- Tommaso Caselli and Piek Vossen. 2016. The Storyline Annotation and Representation Scheme (StaR): A Proposal. In T. Caselli, B. Miller, M. van Erp, P. Vossen, and D. Caswell, editors, *Proc. of the 2nd Workshop on Computing News Storylines*. Austin, Texas, pages 67–71.
- Philipp Cimiano, Janna Lüker, David Nagel, and Christina Unger. 2013. Exploiting Ontology Lexica for Generating Natural Language Texts from RDF Data. In *Proc. of the 14th European Workshop on Natural Language Generation*. ACL, Sofia, Bulgaria, pages 10–19.
- Robert Dale. 1989. Cooking Up Referring Expressions. In *Proc. of the 27th Annual Meeting of the ACL*. Stroudsburg, PA, USA, ACL '89, pages 68–75.
- George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ace) program – tasks, data, and evaluation. In *Proc. of the Fourth Int. Conf. on Language Resources and Evaluation (LREC 2004)*. ELRA, Lisbon, Portugal.
- Pablo Gervás. 2013. Stories from Games: Content and Focalization Selection in Narrative Composition. In *I Spanish Symposium on Entertainment Computing*. Universidad Complutense de Madrid, Madrid, Spain.
- Pablo Gervás, Belén Díaz-Agudo, Federico Peinado, and Raquel Hervás. 2005. Story Plot Generation based on CBR. In A. Macintosh, R. Ellis, and T. Allen, editors, *Applications and Innovations in Intelligent Systems XII: Proc. of AI-2004, the 24th SGAI Int. Conf. on Innovative Techniques and Applications of Artificial Intelligence*. Springer, London, pages 33–46.
- Eli Goldberg, Norbert Driedger, and Richard I. Kittredge. 1994. Using Natural-Language Processing to Produce Weather Forecasts. *IEEE Expert: Intelligent Systems and Their Applications* 9(2):45–53.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-burch, Richard Zens, Marcello Federico, Nicola Bertoldi, Chris Dyer, Brooke Cowan, Wade Shen, Christine Moran, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. ACL, Prague, Czech Republic, pages 177–180.
- Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago Press, University.
- Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the ACL (Volume 1: Long Papers)*. Association for Computational Linguistics, Sofia, Bulgaria, pages 73–82.
- Joel Nothman, Nicky Ringland, Will Radford, Tara Murphy, and James R. Curran. 2012. Learning multilingual named entity recognition from Wikipedia. *Artificial Intelligence* 194:151–175.
- Yanwei Pang, Xin Lu, Yuan Yuan, and Xuelong Li. 2011. Travelogue enriching and scenic spot overview based on textual and visual topic models. *International Journal of Pattern Recognition and Artificial Intelligence* 25(3).
- Steven Poulakos, Mubbasir Kapadia, Andrea Schüpfer, Fabio Zünd, Robert Sumner, and Markus Gross. 2015. Towards an Accessible Interface for Story World Building. In *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. pages 42–48.
- Georg Rehm, Jing He, Julian Moreno Schneider, Jan Nehring, and Joachim Quantz. 2017. Designing User Interfaces for Curation Technologies. In *19th Int. Conf. on Human-Computer Interaction – HCI Int. 2017*. In print, Vancouver, Canada.
- Ehud Reiter, Somayajulu Sripada, Jim Hunter, and Ian Davy. 2005. Choosing words in computer-generated weather forecasts. *Artificial Intelligence* 167:137–169.
- Mark Owen Riedl and Robert Michael Young. 2010. Narrative Planning: Balancing Plot and Character. *J. Artif. Int. Res.* 39(1):217–268.
- Julian Moreno Schneider, Peter Bourgonje, Jan Nehring, Georg Rehm, Felix Sasaki, and Ankit Srivastava. 2016. Towards Semantic Story Telling with Digital Curation Technologies. In L. Birnbaum, O. Popescuc, and C. Strapparava, editors, *Proc. of NLP meets Journalism – IJCAI-16 Workshop (NLP MJ 2016)*. New York.
- Julian Moreno Schneider, Peter Bourgonje, and Georg Rehm. 2017. Towards User Interfaces for Semantic Storytelling. In *19th Int. Conf. on Human-Computer Interaction – HCI Int. 2017*. In print, Vancouver, Canada.
- Jannik Strötgen and Michael Gertz. 2010. Heildetime: High quality rule-based extraction and normalization of temporal expressions. In *Proc. of the Int. Workshop on Semantic Evaluation*. ACL, Stroudsburg, PA, USA, SemEval '10, pages 321–324.
- Jannik Strötgen and Michael Gertz. 2011. Wikiwarsde: A german corpus of narratives annotated with temporal expressions. In *Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2011)*. Hamburg, Germany, pages 129–134.
- Ross Turner, Somayajulu Sripada, Ehud Reiter, and Ian P. Davy. 2006. Generating Spatio-temporal Descriptions in Pollen Forecasts. In *Proc. of the 11th Conf. of the European Chapter of the ACL: Posters & Demonstrations*. ACL, Stroudsburg, PA, USA, EAACL '06, pages 163–166.
- S.R. Turner. 2014. *The Creative Process: A Computer Model of Storytelling and Creativity*. Taylor & Francis.
- Mark D. Wood. 2008. Exploiting Semantics for Personalized Story Creation. In *Proc. of the 2008 IEEE Int. Conf. on Semantic Computing*. IEEE Computer Society, Washington, DC, USA, ICSC '08, pages 402–409.
- Bishan Yang and Tom Mitchell. 2016. Joint extraction of events and entities within a document context. In *Proc. of the 2016 Conf. of the North American Chapter of the ACL: Human Language Technologies*. ACL, pages 289–299.
- Mao Ye, Rong Xiao, Wang-Chien Lee, and Xing Xie. 2011. On theme location discovery for travelogue services. In *Proc. of the 34th Int. ACM SIGIR Conf. on Research and development in IR*. ACM, Beijing, China, pages 465–474.