

Using Visualisation Techniques to Acquire a Better Understanding of Storytelling for Cultural Heritage

Paulo Carvalho, Olivier Parisot and Thomas Tamisier

Luxembourg Institute of Science and Technology, ERIN Department, 41 rue du Brill, Belvaux, Luxembourg

Keywords: Information Extraction, Data Analysis, Entity Recognition, Information Visualisation, Storytelling.

Abstract: Historical information has an important role regarding cultural heritage. It is used to interpret facts occurred in the past and also to understand the present. Storytelling, when applied in the narrative of true events and resulting from different personal views and anecdotal stories, act as an important source of historical information. In this paper, we discuss the problems we encounter in the field of historical information storytelling and we present a software architecture to facilitate the comprehension of stories. More precisely, the proposed solution helps to analyse a story, examine its composition identifying existing entity classes and computing possible relations with other stories, to finally build a visual representation of these stories.

1 INTRODUCTION

Historical information has always been important for knowledge acquisition in order to understand the present. The higher the amount of information available is, the more we are prepared to make decisions based on the information acquired (Delen and Demirkan, 2013). In the specific field of historical information, the more information we have, the more we are prepared to acquire a better understanding about the reason(s) why things are related with a specific subject. E.g. we now have sufficient information to understand the reasons and events that have directly influenced the evolution of the population in Luxembourg during the XIXth / XXth centuries (Scuto, 1995).

Storytelling, when based in real events, is a source of historical information. By historical information storytelling, we mean people relating personal stories, events or interpretations of facts occurred in the past (Thomson, 2011). A good iconic image representation of storytelling is a circle of people around an old person relating a captivating story (Crawford, 2012). The story necessarily told wheels around a specific subject of interest of the listeners. Such events will normally involve different kind of entities. E.g. people, locations, periods, etc. Because historical information contains information about events occurred in the past of people life, sometimes is defined as digital memories (Garde-Hansen et al., 2009). From these personal events and narratives, valuable and interest-

ing information is acquired. The possibilities of using information extracted from storytelling sources are quite wide. Such information can be used not only to comprehend the present but also to solve problems (McLellan, 2007).

The act of sharing stories is important for the spread of information, to transmit values and to teach beliefs over different generations (Initiative et al., 2007). However, this information diffusion does not happen without a cost or risk.

- Storytelling, when applied to historical facts or true events and facts occurred during periods in the past, is about people relating stories they have lived or witnessed somewhere and during a specific period during their lives. Storytelling is then a source of information for cultural heritage. Since the stories comes from people, each story can have a different interpretation depending on the person who is telling it: each person has its own personality and its own understanding of a specific event or action.
- There are several ways to acquire storytelling stories and personal/historical narratives. Computers have been used to tell stories for several decades (Holmquist et al., 2000) related with ancient vs. recent events. Video and audio files are also another way to extract storytelling information: DST (Digital Storytelling) (Klaebe, 2006). The work of obtaining a coherent narrative and understanding of what has been said using such kind of

sources can be very difficult.

Our work focuses on the first point: how to acquire the correct knowledge about a specific subject/story? How can we deal with different interpretations of a specific event that happened in the past? How can we avoid bad interpretations and how can we analyse simultaneously different stories trying to find some correlation between them?

The subject of our research is based on storytelling and narrative of historical heritage. The scope of this work involves location-based and time-based information. Both combined with other type of information (e.g. Person, Event, Subject, etc.). Such information is complex, first, to identify, and after, to analyse and understand correctly. Navigating and extracting knowledge from it may require data mining and data visualisation techniques. The LOCALE project (Tamisier et al., 2016) (supported by the Luxembourg National Research Fund – FNR) has been designed in order to coincide with the 70th anniversary of the end of World War 2 (WW2). The focus of the project will then be given to facts/events/stories that happened during the period 1945-1975, in the Luxembourg region, to contribute for its cultural heritage preservation. The acquired information will be made freely available for users in general by the Locale project.

First, we built a repository of stories, obtained from different users, who lived after the second world war in Luxembourg (or its region). Thus, the stories were obtained in majority from old people and were about several subjects of interest for the region, mainly occurred after the second world war, e.g. politics, relations with other countries, work, health, family, etc. However, since they are told by different elderly people and happened several decades ago, some conflicts and inconsistencies may exist. However, we do not focus our work on this problem assuming that stories are correct. The idea is then to analyse these stories in order to understand and get knowledge in several fields and answers to certain doubts we have, e.g.:

- How the Luxembourg and neighborhood evolved these last decades?
- What was the effect of the second world war for the country in general and for its population?
- Which were the main events occurred in the region after the fifteen years following this period?
- etc.

Considering the possibility of having a large repository of stories to analyse, fulfilled by different persons and related with various types of events and

subjects, the difficulty to obtain a good understanding of a story can highly increase along the volume of data to analyse. The same subject of discussion have numerous and different interpretations if debated by different persons. We support the idea that the use of Information visualisation techniques can support and help users to obtain a better and more precise idea about the content and meaning of a story. In this article, we will show how such techniques are beneficial for such purpose.

2 STORYTELLING ISSUES

We define in this work Storytelling as being about people telling a narrative about a specific event that happened in the past. The way stories are obtained can be different. E.g. writing a story using a computer to describe it, recording the recital using a voice recorder or a smartphone, a simple conversation we have that we transcribe to a computer after, etc. Based on these different sources, we now have a centralized repository - a database - of stories concerning Luxembourg and its region, happened between 1945 and 1960, and which can be accessed and exploited directly by a computer application.

2.1 Story Structure

Each story can be described as being a set of sentences. Depending on its size and its level of detail, each story is composed by one or more sentences. For example, these two sentences addresses to the same topic but they do not have the same level of information/detail.

- *The National Railway Company of Luxembourg (CFL) is officially established on May 14, 1946.*
- *The National Railway Company of Luxembourg (CFL) is the public railway company of Luxembourg and has been created on 14 May 1946.*

The structure of a story and the way how sentences are organised has a big influence on its interpretation and comprehension (Cirilo and Foss, 1980). Some studies support the idea that the link between recognized terms and entities in the text and predicates has to be made manually (Mulholland et al., 2012). Others sustains that this mapping can be made, at least partially, automatically using Natural Language Processing (NLP) solutions. E.g. named Entity Recognition and classification techniques (ER) (Sekine and Ranchhod, 2009).

2.2 Entity Recognition

In order to be able to interpret a story, we must first have the knowledge to interpret each sentence that compose it. For that, we must recognize specific terms existing in each sentence: a name, a person, a local, a number, a date, etc. A brief example is shown in Section 2.2.1. Entity recognition (ER) can be defined as the task and methods responsible to detect and classify terms existing in a sentence into different classes (e.g. date, location, person, etc.) (Toral and Munoz, 2006). Based on this definition, we assume that ER is a good candidate to support us in the task of understanding stories. The complexity of the data to analyse also have a direct implication over the comprehension of a story. This complexity depends on several factors and increase if one or more of these factors occurs (McCallum and Li, 2003):

- Several languages are used in the text (e.g. an English story with French names).
- The scope of used locations is large (e.g. popular Vs unknown locations - Luxembourg Vs Weiler).
- The volume of organisations is large (e.g. Spuerkess as a company name, OGBL as a union organisation, etc.).
- There is a high variety of Named Entities (e.g. Volkswagen as a brand, Luxembourgish as a nationality, etc.).

The study made by (Nadeau and Sekine, 2007) presents a survey of fifteen years of research made in the ER field (1991-2006). It supports that the choice of a specific ER solution must be made having in mind a clear idea about the features that should be covered. This fact has to be taken into account since it has impact on the computational needs. ER field problems are far from being solved. While an important aspect of ER is the manual annotation of data (supervised), it is also true that some studies shows that other solutions exist and should also be considered, depending on the objective to be accomplished. Indeed, some work has already been made supporting that unsupervised and semi-supervised ER techniques can also lead to encouraging results.

2.2.1 Simple Example

E.g. *The [Alzette] [River] originates in [France].*

The words between square brackets are the terms which must be automatically and correctly recognized in order that we are able to interpret and understand the meaning of the sentence. We have then: a) [Alzette] - a name? b) [River] - a name? c) [France] - a place?

2.2.2 Disambiguation & Multiple Occurrences

Moreover, it may happen that a term appears several times but with different meanings. The need of external knowledge may be necessary to solve this kind of issues because it represents an important barrier to overcome in order to obtain the correct meaning of a sentence (Ratinov and Roth, 2009).

E.g. *The Luxembourg population felt more protected during the war in Luxembourg than in Remich.*

In this example, the sentence has two different occurrences of the term *Luxembourg* but with different meanings: Luxembourg (the country) and Luxembourg (the city). It is mandatory that we are able to obtain first, a correct identification of both terms and finally to obtain the right meaning of each. Otherwise, obtaining the correct understanding of the sentence is compromised.

2.3 Several Sources, Different Interpretations

Depending on the story's source - the person who has told the narrative - different interpretations of the story may happen and distortions can be injected in it (Smith et al., 1983). This leads that the subject discussed in different stories, and told by different persons, can origin a different conclusion. To avoid this, it is important to have the opportunity to view and analyse different stories, coming from different sources and related with the same subject in order to have the means to obtain a more precise evaluation and understanding about a specific event/story.

3 Extracting Knowledge from Stories

Acquiring knowledge from a story can be quite simple. However, depending on its size, it may become a complicated task if the text to be analysed is large. The complexity of its interpretation can also be largely influenced if many stories addressing to the main subject exist (e.g. *Luxembourg-city*, *Luxembourg's airport*, *Luxembourg-country*, etc.). This complexity will even be larger if the stories were told by different persons, using different languages and maybe having different opinions about the subject. To obtain the better and more precise result in terms of a story interpretation, we support the idea that it

is useful to compare different stories being somehow related. This relation can exist due to a period time (e.g. year), a location (e.g. Luxembourg), an event (e.g. flood), etc. Based on this idea, we have worked on different approaches to visualise the information of several stories at once and, simultaneously, to have the ability to view how these stories are linked. Our idea is also to use the advantages provided by Information Visualisation techniques to furnish an efficient platform to analyse and interpret correctly the stories. An important aspect we had to take into account to build our prototype is the type of users who will use the application: users with and without IT expertise and belonging to different generations. The prototype we have developed is introduced in the next sections.

3.1 Architecture

Our prototype is composed by four modules:

- The database where all collected stories are stored.
- The ER (Entity Recognition) module - responsible to detect, identify and classify by classes terms existing in each story.
- The JSON module - builds a file, containing the information computed from the ER module (e.g. Date, Local, Person; Date; Location) and represented in the JSON format.
- The visualisation module - interprets the generated JSON file and displays its information using different types of charts in order to provide an efficient and user-friendly support for stories interpretation and understanding.

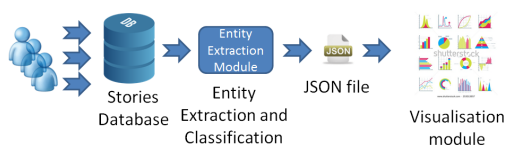


Figure 1: Global architecture with its main components.

3.1.1 The Database

The idea is to gather stories related with Luxembourg and its region, occurred in the period after the second world war. More precisely during the period 1945-1960. We already collected 267 stories just using information obtained from several web portals (e.g. Wikipedia, <http://www.industrie.lu>, <http://www.vdl.lu>). Gathered stories are normally short, relating a specific event occurred in a concrete place during a certain date. We are also gathering stories by visiting retirement homes to interview people interested in participating in our initiative, in

order to record their memories. Another phase we are working on is related with how to process these records with voice recognition software so we can obtain automatically a version under text file formats (Speech to text recognition).

3.1.2 Entity Recognition Module

We have used the Natural Language Understanding API (formerly AlchemyAPI) to extract entities from text¹. Nevertheless, despite the engine being capable of recognize different localities (e.g. Paris, Luxembourg, etc.), it seems to have some limitations regarding smaller Luxemburgish localities like *Sandweiler* or *Kirchberg*. As a result, we have taken the stories, applied Alchemy over them, fixed semi-manually the results and built a JSON file ready to be processed by the next module (Information visualisation module). The idea here is to build first a proof-of-concept to show the different existing possibilities we can use and apply to analyse, understand and interpret stories. Everything related with entity recognition and classification still has to be deeply studied and developed.

3.2 Information Visualisation

In a recent work, various visualisation techniques were applied in order to analyse textual data related to Digital Humanities (Jänicke et al., 2016). We think that the strength and flexibility of Information Visualisation (Ware, 2012; Chen, 2013) is a major asset for our aim. It may be applied to help us to visualise the information related with a specific story but also to get an overview of all the stories related with it and the way they are linked. Some work and studies were already made related with how information visualisation is applied in the field of storytelling and stories analysis. Stories evolution is a topic which has been studied by (Subašić and Berendt, 2008) and where graphs are used to show this evolution. The work made by (Gershon and Page, 2001) is based on how stories can be visually represented. Despite being very interesting and related with our own topic, it does not approach the problematic of visualising relations existing between stories. Finally, (Wojtkowski and Wojtkowski, 2002) supports that information visualisation can be of major importance in the field of storytelling.

This led us to build a visualisation information module to show this information using different forms of representations (charts). Thus, its meaning can be efficiently and easier understood and interpreted, both

¹<https://www.ibm.com/watson/developercloud/natural-language-understanding.html>

by users with or without IT or information visualisation expertise and users belonging to different generations. Indeed, some visualisation techniques are easy to understand in general but others require knowledge in terms of data visualisation (e.g. Scatterplot, Voronoi Diagram, Treemap, etc.).

The visualisation information module is built under a combination of the *Javascript* and *Php* languages and using an external Javascript library (D3JS - <http://www.d3js.org>) which aim is to build the charts needed. It is important to notice that, normally, people born during the first half of 20th century are not very familiar with new technologies and information visualisation techniques. To be prepared for the different user profiles the application must support, we worked on different visualisation charts. In the next section we present some of them, which we think are adapted for a wide range of user profiles.

3.2.1 JSON File

The JSON file, built by the JSON module and using the information generated by the entity recognition module, is the input received and interpreted by the information visualisation module. In order that the information visualisation module is able to display the information related with a story and all the stories having a relation with it, the JSON file is composed by different sections:

- Main story information.
- Entity classes recognized in the story.
- Entity recognition values.
- Stories with common entities with the main stories.

The image 2 shows an example of a JSON file containing these four sections and ready to be processed by the visualisation module.

3.2.2 Visualisation Solutions

In order to achieve our aim, we have worked and tried several information techniques: the *Partition Layout* (Figure 3), the *Collapsible Indented Tree* (Figure 4), the *Dendrogram and Grouped Horizontal Bar Chart* (Figure 5), the *Sankey Diagram* (Figure 6) and the *Zoomable Sunburst* (Figure 7). These solutions were chosen because of their ability to present the relations between stories and because they are also user-friendly and easy to understand. Both charts presents the same information, under a different shape, but organising always the information in four different sections:

- First - the main story section. The history title is shown on the section. If the user moves the mouse over it, the story is shown.
- Second - the different classes of entities recognized in the story (e.g. Local, Period, Person, etc.).
- Third - the different entity values related with each class of entity recognized.
- Fourth - the stories somehow related with the main story. The story title is displayed on the story section. If the user moves the mouse over it, the content of the story is shown.

In the Table 1, we present a brief overview and conclusions for each of the visualisation chart analysed and tested.

The aim of the project is to provide support, for a wide range of user profiles. The application is meant to be used by people without IT knowledge, information visualisation experience or even without any computer usage experience and also by people belonging to different generations: young people to raise their interest over historical events, mid-age people but also old people, born before the WWII, etc. This fact led us to take special care to the types of visualisations to use: they have to be efficient and easy-to-understand. The reason is that, there are many visualisation techniques, which are very efficient but also very complex to interpret by people without any information visualisation background. For this reason, we think that the charts *Collapsible Indented Tree* and the *Dendrogram and Grouped Horizontal Bar Chart* are the most suitable for our needs, based on the analysed solutions. Indeed, the *Partition Layout* has some limitations regarding its capacity to show large amount of information which may limit us if the quantity of stories to analyse is large. On the other hand, since circular charts are often considered due to their efficient behaviour (e.g Pie chart), we needed to analyse a circular chart solution able to display stories information. Despite its popularity, the pie chart would not fit to our objective because it is more designed to display statistical information and distribution values than relations between objects (Spence, 2005). The *Zoomable Sunburst* has this ability. Despite being a different chart than the *Collapsible Indented Tree*, we think that it is more complex to analyse, especially if the analysis is made by old people or users without computer expertise.

3.2.3 Map Chart

Historical narratives are many times related with a specific location. A story occurred in the past has

```

{
  "id": 1,
  "date": "14 mai 1946",
  "name": "Luxembourg",
  "locationName": "Luxembourg",
  "storytext": "La Société nationale des chemins de fer luxembourgeois (CFL) est créée le 14 mai 1946.",
  "longitude": 49.815273,
  "latitude": 6.129582999999999,
  "children": [
    {
      "id": 2, "name": "Lieu", "type": "entity",
      "children": [
        {
          "id": 3, "name": "Luxembourg",
          "children": [
            {
              "id": 3,
              "date": "14 mai 1946",
              "locationName": "Luxembourg",
              "name": "Luxembourg",
              "storytext": "La Société nationale des chemins de fer luxembourgeois (CFL) est l'entreprise ferroviaire",
              "longitude": 49.815273,
              "latitude": 6.129582999999999
            },
            {
              "id": 4,
              "date": "14 mai 1946",
              "name": "Luxembourg",
              "locationName": "Luxembourg",
              "storytext": "La Société nationale des chemins de fer luxembourgeois (CFL) est officiellement créée le",
              "longitude": 49.815273,
              "size": 1000,
              "latitude": 6.129582999999999
            },
            {
              "id": 5,
              "date": "14/7/1946",
              "name": "Luxembourg-Ville",
              "locationName": "Luxembourg-Ville",
              "storytext": "C'est ensuite le départ pour le Grand-Duché où les Churchill sont accueillis par le Princ",
              "longitude": 49.61162100000001,
              "size": 1000,
              "latitude": 6.1319346
            }
          ]
        }
      ]
    }
  ]
}
    
```

Figure 2: JSON partial file example containing the collected stories from Luxembourg and region so far.

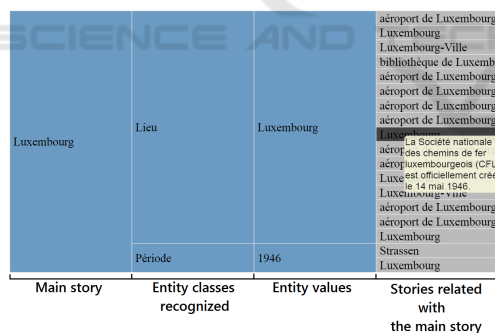


Figure 3: Partition Layout. The information is organised into four different sections to increase chart's understandability and efficiency.

often been experienced in a region, a city, a monument, etc. This led us to sustain that the place where an historical event happened is of high importance to establish the link between stories having some correlation. For this reason, we give importance and work with Map chart visualisations, by using the Google Map API ². The figure 8 shows us the place where the stories presented in the section 3.1.1 occurred.

²<https://developers.google.com/maps/>

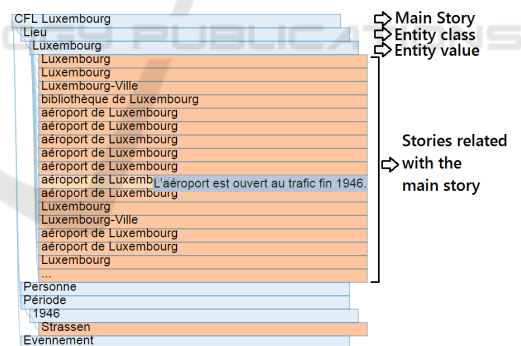


Figure 4: Indented tree chart. The information is organised into four different sections.

4 CONCLUSIONS

In this paper, we started by presenting our definition of storytelling and its relation and importance with historical heritage. The different problems and challenges related with storytelling, especially when the stories are orally told by people, were discussed. A prototype able to interpret visually stories and their respective relations was described. The problems to

Table 1: Charts comparison.

	⊕	⊖
Partition Layout (figure 3)	Efficiency; Easy to understand; Sober	Design not attractive; Scalability
Collapsible Indented Tree (figure 4)	Efficiency; Easy to understand; Sober; Scalability (due to the possibility to collapse sections)	Not the sexiest chart
Dendrogram and Grouped Horizontal Bar Chart (figure 5)	Efficiency; Easy to understand; Sober; Scalability (due to the possibility to collapse sections); Colors helping to evaluate the proximity between stories	
Sankey diagram (figure 6)	Efficiency; Easy to understand; Sober; Scalability (due to the possibility to collapse sections); Design more attractive than other solutions	May be complex to understand
Zoomable Sunburst (figure 7)	More sophisticated	Scalability; More complex to interpret; Not fit for some types of user profiles

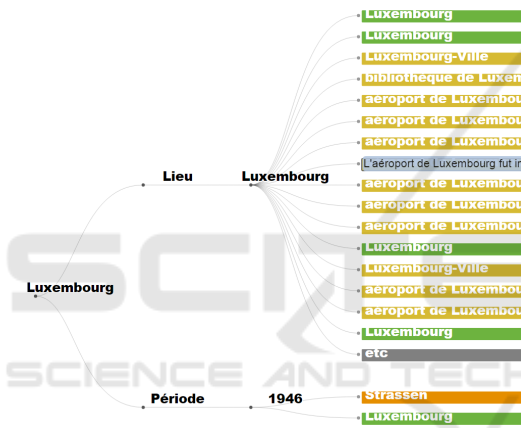


Figure 5: Dendrogram and Grouped Horizontal Bar Chart. The information is organised into four different sections to increase chart's understandability and efficiency.

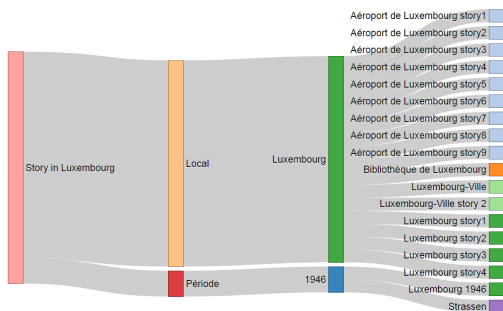


Figure 6: Sankey diagram. The information is organised into four different sections and uses colors to enhance its understandability and efficiency.

be overtaken were discussed. After having converted the stories from their source (e.g. audio) to text, other issues must be solved. It is first necessary to interpret and comprehend the meaning of each sentence composing a story. Entity recognition methods are used

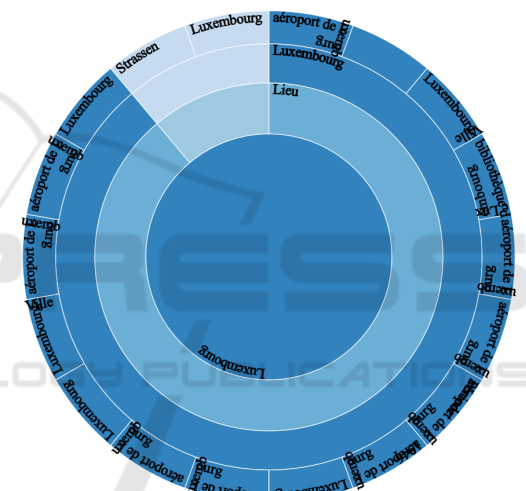


Figure 7: Zoomable Sunburst Chart - the complexity to understand information in this type of chart is higher than other solutions presented previously.

to extract the meaning of each word. Entities like a place, a person, a date and so on, should be recognized. We show that information visualisation is a valid solution to understand stories.

Due to the public the solution is designed for - people belonging to different generations and people with(out) IT expertise - not every visualisation solution fits to our scope. The provided solution must be well studied taking into account this important problematic. We have identified some visualisation charts able to display the relations between stories. These are easy to interpret and should be usable by a wide group of users, with different type of profiles and expertise. However, there is still work to do. A user testing phase should be done to support and evaluate our work. The reason we did not execute such phase

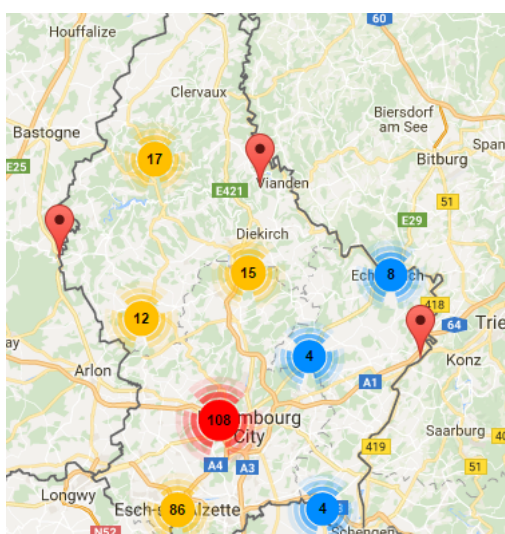


Figure 8: Map chart with the stories locations. To summarize the information, clustering of points is applied.

yet is because we still have to focus and invest time in the entity recognition module in order that entities are automatically and correctly recognized.

ACKNOWLEDGEMENTS

This work was realized in the context of the LOCALE research project (Tamisier et al., 2016), funded by a Core grant from the Luxembourg National Research Fund.

REFERENCES

Chen, C. (2013). *Information visualisation and virtual environments*. Springer Science & Business Media.

Cirilo, R. K. and Foss, D. J. (1980). Text structure and reading time for sentences. *Journal of Verbal Learning and Verbal Behavior*, 19(1):96–109.

Crawford, C. (2012). *Chris Crawford on interactive storytelling*. New Riders.

Delen, D. and Demirkan, H. (2013). Data, information and analytics as services.

Garde-Hansen, J., Hoskins, A., and Reading, A. (2009). *Save as... digital memories*. Springer.

Gershon, N. and Page, W. (2001). What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37.

Holmquist, L. E., Helander, M., and Dixon, S. (2000). Every object tells a story: Physical interfaces for digital storytelling. In *Proceedings of the NordiCHI*.

Initiative, E. L. et al. (2007). 7 things you should know about digital storytelling. *Boulder, CO: EDUCAUSE*. Retrieved on February, 8:2010.

Jänicke, S., Franzini, G., Cheema, M., and Scheuermann, G. (2016). Visual text analysis in digital humanities. In *Computer Graphics Forum*. Wiley Online Library.

Klaebe, H. G. (2006). The problems and possibilities of using digital storytelling in public history projects.

McCallum, A. and Li, W. (2003). Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 188–191.

McLellan, H. (2007). Digital storytelling in higher education. *Journal of Computing in Higher Education*, 19(1):65–79.

Mulholland, P., Wolff, A., and Collins, T. (2012). Curate and storyspace: an ontology and web-based environment for describing curatorial narratives. In *Extended Semantic Web Conference*, pages 748–762. Springer.

Nadeau, D. and Sekine, S. (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1):3–26.

Ratinov, L. and Roth, D. (2009). Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 147–155.

Scuto, D. (1995). Emigration et immigration au Luxembourg aux XIXe et XXe siècles. *Itinéraires croisés. Luxembourgeois à l'étranger, étrangers au Luxembourg*, pages 24–28.

Sekine, S. and Ranchhod, E. (2009). *Named entities: recognition, classification and use*, volume 19. John Benjamins Publishing.

Smith, S. W., Rebok, G. W., Smith, W. R., Hall, S. E., and Alvin, M. (1983). Adult age differences in the use of story structure in delayed free recall. *Experimental Aging Research*, 9(3):191–195.

Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics*, 30(4):353–368.

Subašić, I. and Berendt, B. (2008). Web mining for understanding stories through graph visualisation. In *IEEE ICDM'08*, pages 570–579.

Tamisier, T., McCall, R., Gheorghe, G., and Pinheiro, P. (2016). *Visual Analytics for Interacting on Cultural Heritage*, pages 296–299. Springer.

Thomson, A. (2011). Memory and remembering in oral history. *The Oxford handbook of oral history*, pages 77–95.

Toral, A. and Munoz, R. (2006). A proposal to automatically build and maintain gazetteers for named entity recognition by using wikipedia. In *EACL*, pages 56–61.

Ware, C. (2012). *Information visualization: perception for design*. Elsevier, Third edition.

Wojtkowski, W. and Wojtkowski, W. G. (2002). Storytelling: its role in information visualization. In *European Systems Science Congress*, volume 5. Citeseer.