

Storybook: A Tool for the Semi-automatic Creation of Book Trailers

Eleonora Bernasconi¹, Miguel Ceriani², Francesca De Luzi¹, Francesco Sapiro¹ and Massimo Mecella¹

¹Sapienza Università di Roma, Department of Computer, Control, and Management Engineering Antonio Ruberti (DIAG) Via Ariosto, 25, 00185 Rome, Italy

²Università degli Studi di Bari Aldo Moro, Department of Computer Science Via Edoardo Orabona, 4, 70125 Bari, Italy

Abstract

Multimedia storytelling is an effective and engaging method to convey information in multiple domains. Specifically, book trailers –video advertisements for books– positively influence the desire to learn and the motivation to read. This work describes the design and implementation of Storybook, a tool for the semi-automatic creation of book trailers aiming to support storytelling for digital libraries. Storybook supports an expert by gathering relevant crowd-sourced multimedia content, which, arranged as stories, can be used to showcase a book in the form of video clips. Crucially, the expert controls how the content is finally combined and edited rather than offering a fully automated process. In an early informal evaluation, experts consider the method favourably.

Keywords

storytelling, knowledge extraction, digital library

1. Introduction

Storytelling is the act of narrating, using the principles of narratology in the audiovisual or literary field to communicate and facilitate learning and understanding. For example, in school books (to make a concept simple), a story with characters is used; similarly, in language courses, even for adults, the contents are organized with characters who show an aspect of the language through dialogue or a text. This methodology is an experiential resource, as it promotes a generative development between experience, its observation and the resulting insights.

Numerous researches [1, 2, 3, 4] show that a book trailer fosters the desire to learn and the level of motivation to read. From publishers' point of view, the usage of promotional trailers is a response to a changing market with a high focus on digital and visual media. The goal of a digital presentation of a book is nevertheless broader than selling it and includes providing helpful information to the potential future reader.

The design, implementation, and early evaluation of Storybook are described in this work.

IRCDL 2022: 18th Italian Research Conference on Digital Libraries, February 24–25, 2022, Padova, Italy

✉ bernasconi@diag.uniroma1.it (E. Bernasconi); miguel.ceriani@uniba.it (M. Ceriani); deluzi@diag.uniroma1.it (F. D. Luzi); sapiro@diag.uniroma1.it (F. Sapiro); mecella@diag.uniroma1.it (M. Mecella)

>ID 0000-0003-3142-3084 (E. Bernasconi); 0000-0002-5074-2112 (M. Ceriani); 0000-0002-9896-2528 (F. D. Luzi); 0000-0001-9544-2108 (F. Sapiro); 0000-0002-9730-8882 (M. Mecella)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Storybook is a software tool to support the creation of book trailers by collecting and organizing relevant video content. The system users retain control on how to edit and compose the final content. The proposed process aims to semi-automatically build digital trailers that allow the viewer, generically interested in a specialized topic but not expert, to appreciate better the topic, both for their own cultural/professional enrichment and a possible purchase.

The remaining sections are organized as follows. Section 2 presents related work about data-driven storytelling systems. Section 3 describes the pipeline of the proposed system and the technologies used. Section 4 reports an early evaluation. Finally, Section 5 discusses future research directions.

2. Related work

This section briefly surveys relevant literature for relevant works, starting from data-driven storytelling systems. Storytelling and narrative visualization techniques are two lines of research that intertwine and have led to new methods for enhancing data understanding, information dissemination and displaying information [5, 6].

Users usually encounter difficulties creating a data-driven story due to technical barriers, which motivate the design and development of various automated and simplified generation tools. The creation of a data-based narrative can be divided into two branches [7]. The first focuses on authoring systems to facilitate the design process. In this case, it should be noted that users already have a deep understanding of the data and what they want to present. For instance: Ellipsis [8], InfoNice [9] and ChartAccent [10] enable dynamic annotations to support data storytelling allowing the user to integrate visualizations into an illustrative story directly through direct manipulation and reuse of existing story elements; Narvis [11] is a tool to extract the combination of visual elements of visualization and organize them as a presentation; DataClips [12] is designed to help users generate data videos.

The second branch of data narration research, which is one of our interests, seeks to break down technical barriers in creating data-based stories, focuses more on automated systems to save users' efforts that take data as inputs and generate organizations of history components for users. For instance:

- VoxPopuli [13] is developed to generate video documentaries based on interviews about controversial topics. Via a Web interface, the user selects one of the possible topics and a point of view, and the engine assembles video material from the repository to satisfy the user request.
- FotoInmotion [14] is designed to produce automatic videos based on a still image. The tool considers semantic information to produce storytelling videos, focusing on the relevant features of the input image.
- NewsViews [15] is developed to create interactive, annotated maps from news articles. The NewsViews's maps support trend identification and data comparisons relevant to a given news article.

Summarizing, existing storytelling tools focus on how to tell a story but rarely base the story on original data. Moreover, these tools assume that the story content is created manually,

resulting in inefficiency. In contrast, Storybook supports automatic story generation and flexible story editing features, which ensure quality, reduce the barrier and improve the efficiency of visual storytelling.

3. Pipeline and discussion

A pipeline to support making book trailers has been designed and developed. The process is schematized in the Algorithm 1. The following subsections introduce the deployed pipeline's details and discussion and motivations of choices.

Algorithm 1: The algorithm

```

Input :Data about a book: content (PDF) and metadata (title and topic).
Input :A set of videoServices, web services offering video search capability.
Input :Set of keywords whitelist to add to the search.
Input :Set of keywords blacklist to exclude from the search.
Input :Parameters k (number of videos), l (number of named entities extracted), m (number of video results), o (number of labels from image classification).
Output: A set of k videos relevant to the book content.

1 entities  $\leftarrow$  top l named entities found in content (by named-entity recognition)
2 bookKeywords  $\leftarrow \{title\} \cup \{topic\} \times (\{topic\} \times (entities \cup whitelist))$ 
3 videosFound  $\leftarrow \{\}$ 
4 for service  $\in$  videoServices do
5   for kwd  $\in$  bookKeywords do
6     results  $\leftarrow$  first m results from service searching kwd while excluding blacklist
7     videosFound  $\leftarrow$  videosFound  $\cup$  results
8   end for
9 end for
10 images  $\leftarrow$  k largest images of content
11 selectedVideos  $\leftarrow \{\}$ 
12 for image  $\in$  images do
13   imageLabels  $\leftarrow$  top o labels assigned to image by image classification
14   rank videosFound by overlap of their tags with imageLabels
15   videoForImage  $\leftarrow$  video with maximum overlap
16   selectedVideos  $\leftarrow$  selectedVideos  $\cup \{videoForImage\}$ 
17 end for
18 return selectedVideos

```

The process is designed to work on books with meaningful visual content in images extracted from the PDF. The idea is to find videos relevant to the domain and subject of the book that can be associated with a set of images found in the book. The underlying intuition elicited working with content experts is that the visual content included in the book, like images, is relevant to identify appropriate video content to create a book trailer.

The algorithm has four main parameters: *k*, number of images extracted from the book, it is also the number of videos provided as output; *l*, number of named entities extracted from the book; *m*, number of video results for each keyword and each video search service; *o*, number of labels gathered from the classification of each image. We assume that all the considered books have at least *k* images.

3.1. Retrieving data for the book

The first important step in the pipeline is to retrieve/extract metadata related to the book itself (see input metadata and line 1 of Algorithm 1).

We already have implemented an automatic semantic extractor of information from a digital library called *Arca* [16, 17, 18] that produces a knowledge graph (KG) [19, 20] with all extracted information and integrated with KGs of the web, so we decided to use that one. However, it can be easily replaced or extended with other semantic enrichment tools as long as they process and extract relevant information from the book. In particular, we perform a SPARQL query¹ to the KG and retrieve existing and extracted metadata for a specific book, including the top l concepts found as named entities in the book.

It is essential to notice that the whole script can run in batch mode. A web-server has been built to manage the execution of the whole Storybook process, while also providing metadata related to the book before, during, and after the process.

3.2. Web Crawling

We consider the book metadata gathered in the previous step during this step, and we run a web crawl looking for relevant videos (lines 2–9 of Algorithm 1). After experimenting with different sources of video content, we decided to focus on a single source, YouTube². YouTube is currently the most extensive database of videos in the world.

In the interface for Storybook, the curator can also specify a keywords' whitelist and blacklist. In this way, the curator can try to steer the research of the crawler in different directions. The crawler uses the whitelist to extend the top concepts retrieved in the previous steps, extending the different keywords and searches that the crawler will perform. Instead, the blacklist is used to filter out more videos by excluding certain keywords. For example, depending on the specific search, sometimes it happens to find videos that explain a particular concept with slides (not helpful in building a trailer). Using the blacklist, the curator can cancel out some of these videos.

It is crucial to note that some retrieved videos are too long to suit our purpose. Even if a long video contains a shot valuable for the trailer, searching that shot in hours of videos would be difficult for the curator. So, we decided to cap the length of the retrieved videos. By default, the limit is set to be 480 seconds, but the curator can easily change this parameter in the application's settings.

We also encountered another problem in searching in such a vast and mixed database. A lot of the videos uploaded on Youtube have licenses that prohibit reuse; hence they cannot be used for our purpose of building a trailer. Therefore, we implemented another filter for retrieving only videos with the Youtube Creative Common license that allows for commercial use. Nevertheless, we can search and retrieve many videos that fit within the search parameters. The gathered video content and related metadata will be used in the last step to compose the draft of a trailer for the book.

¹<https://www.w3.org/TR/rdf-sparql-query/>

²<https://www.youtube.com/>

3.3. Image Extraction and Processing

Another fundamental step in the pipeline of Storybook is to extract images from the *PDF* files of the books (line 10 of Algorithm 1).

Difficulties arise from the fact that many book publishers compress the *PDF*, hence losing the information related to the images. Thus, it is necessary for the algorithm first to detect whether the *PDF* file was compressed or not. This detection can be done by looking into the metadata and searching for the compressor's sign (often, it leaves a signature). As a result, we can separate the compressed *PDF* to pass through an *OCR* process that includes rebuilding the images as well; the one we found most efficient was the *Adobe Acrobat* one³.

Subsequently, whether the *PDF* was not compressed or had been rebuilt from the *OCR*, the next step was to understand the color space in which the image is stored inside the *PDF*. In some rare cases, it is not possible to rebuild the color space, so it is impossible without analyzing the image to know if a 0 means a white or black pixel⁴. In these cases, we extract both versions of the image.

Finally, we select the k best images of the book. After trying different methods, the easiest one that yielded good results was sorting the images by their dimensions. We observed that book publishers tend to let essential images take up more space on the page (thus being more significant in size once extracted).

3.4. Image Classification

An image classification process is run on each of the best k images identified at the previous step (line 13 of Algorithm 1). In order to speed up the development, instead of implementing a custom image classification, we went for using an external service, *Google Vision AI*⁵. For each image, we retrieve up to o labels, along with their confidence level. In the next step, this information is used when generating the suggestions for the trailer.

An interesting piece of information is the *SafeSearch* detection: how likely the image contains *adult*, *spoof*, *medical*, *violent* or *racy* themes⁶. We do not use this information directly; however, it is available for the curator inside a comprehensive JSON file generated at the end of the process.

3.5. Trailer Composition

The last step of the process consists in filtering videos (lines 14–16 of Algorithm 1) and organising them in a draft of the final trailer (line 18 of Algorithm 1).

The algorithm assigns each video a multidimensional score (one dimension for each image). The score increases when there is a match between the image's labels and the video's metadata, such as the description. The confidence of the image-keywords association gives a score's weight (as assigned during the classification step). Once each video has a score, the algorithm matches the highest score per single image, associates that video with the specific image, and

³<https://www.adobe.com/acrobat.html>

⁴The opposite applies to 255.

⁵<https://cloud.google.com/vision/>

⁶<https://cloud.google.com/vision/docs/detecting-safe-search>

discards all the others. Afterwards, the trailer is generated by interleaving the extracted images, and their correspondent retrieved videos. This draft of the trailer is generated for compatibility reasons as an annotated PowerPoint presentation, thus allowing the curator to manipulate the content as they see fit before the actual video creation. Furthermore, all the data generated along the pipeline is packaged inside a JSON file that the curator can access to check the intermediates results and other details on the process.

4. Early evaluation

The system was designed and implemented based on requirements iteratively refined with a team of curators of a publishing house. In respect to the algorithm parameters, satisfactory results were reached by choosing $l = k = o = 10$ and $m = 12$.

Furthermore, the curators gave some general feedback on the system paradigm. They identified several perceived strengths and potential of the system:

- the transformation of the images of books into information nodes;
- the control of the automatic creation process of a book trailer, through accessible configuration parameters;
- the free access and management of the information output generated by Storybook.

They also expressed some concerns for perceived weaknesses: the scarce availability of videos of niche topics on the web and the prolonged extraction process duration for searches involving longer crawler queries.

5. Conclusions and future work

Storybook is a tool for the semi-automatic generation of book trailers. It has been designed with experts in the field, implemented, and tested on the book catalog of a publishing house. From a high-level perspective, the main novelty of the approach is that it supports the creation process while not replacing the human contribution and final choice. An early evaluation shows that expert users appreciate this aspect and react positively to how relevant videos are collected. A potential limiting factor has been found in the limited availability of online videos on niche topics, significantly restricting the search to videos with licenses that allow unrestricted reuse. For the future, we have planned a formal evaluation of the system with a larger sample of users and the integration of Storybook into the Arca system mentioned in section 3.1.

Acknowledgments

This work has been partly supported by projects ARCA (POR FESR Lazio 2014–2020 - Avviso pubblico “Creatività 2020”, domanda prot. n. A0128-2017-17189), STORYBOOK (POR FESR Lazio 2014-2020 - Avviso Pubblico “Progetti di Innovazione Digitale”, domanda prot. n. A0349-2020-34437), and PON R&I 2014–2020 – Attraction and International Mobility (AIM) – project n. 1852414 of the Italian Ministry of University and Research.

References

- [1] N. I. Nikonova, S. Y. Zalutskaya, Modern communication technologies in education: book trailer, *Revista Tempos e Espaços em Educação* 14 (2021). URL: <http://dx.doi.org/10.20952/revtee.v14i33.15256>. doi:10.20952/revtee.v14i33.15256.
- [2] O. V. Barnych, K. V. Naidon, O. B. Mamchych, Book trailer as a means of motivating younger schoolchildren to read, *Pedagogical sciences reality and perspectives* 1 (2021) 192–196. URL: <http://dx.doi.org/10.31392/npu-nc.series5.2021.79.1.41>. doi:10.31392/npu-nc.series5.2021.79.1.41.
- [3] G. Jiménez-Marín, R. E. Zambrano, The book trailer as a publishing house promotional tool, *Advances in Business Strategy and Competitive Advantage* (2020) 147–160. URL: <http://dx.doi.org/10.4018/978-1-7998-3119-8.ch011>. doi:10.4018/978-1-7998-3119-8.ch011.
- [4] D. Luchev, D. Paneva-Marinova, M. Dimova, Digital storytelling and digital book trailer applications for educational purposes in bulgaria, *INTED2019 Proceedings* (2019). URL: <http://dx.doi.org/10.21125/inted.2019.0211>. doi:10.21125/inted.2019.0211.
- [5] E. Segel, J. Heer, Narrative visualization: Telling stories with data, *IEEE Transactions on Visualization and Computer Graphics* 16 (2010) 1139–1148. URL: <http://dx.doi.org/10.1109/tvcg.2010.179>. doi:10.1109/tvcg.2010.179, 432 cites:.
- [6] C. Tong, R. Roberts, R. Borgo, S. Walton, R. Laramee, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo, X. Ma, Storytelling and visualization: An extended survey, *Information* 9 (2018) 65–65. URL: <http://dx.doi.org/10.3390/info9030065>. doi:10.3390/info9030065, 25 cites:.
- [7] Y. Wang, Z. Sun, H. Zhang, W. Cui, K. Xu, X. Ma, D. Zhang, Datashot: Automatic generation of fact sheets from tabular data, *IEEE transactions on visualization and computer graphics* 26 (2019) 895–905.
- [8] A. Satyanarayan, J. Heer, Authoring narrative visualizations with ellipsis, *Computer Graphics Forum* 33 (2014) 361–370. URL: <http://dx.doi.org/10.1111/cgf.12392>. doi:10.1111/cgf.12392, 42 cites:.
- [9] Y. Wang, H. Zhang, H. Huang, X. Chen, Q. Yin, Z. Hou, D. Zhang, Q. Luo, H. Qu, Infonice, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018). URL: <http://dx.doi.org/10.1145/3173574.3173909>. doi:10.1145/3173574.3173909, 19 cites:.
- [10] D. Ren, M. Brehmer, B. Lee, T. Hollerer, E. K. Choe, Chartaccent: Annotation for data-driven storytelling, *2017 IEEE Pacific Visualization Symposium (PacificVis)* (2017). URL: <http://dx.doi.org/10.1109/pacificvis.2017.8031599>. doi:10.1109/pacificvis.2017.8031599, 25 cites:.
- [11] Q. Wang, Z. Li, S. Fu, W. Cui, H. Qu, Narvis: Authoring narrative slideshows for introducing data visualization designs, *IEEE Transactions on Visualization and Computer Graphics* 25 (2019) 779–788. URL: <http://dx.doi.org/10.1109/tvcg.2018.2865232>. doi:10.1109/tvcg.2018.2865232, 7 cites:.
- [12] F. Amini, N. H. Riche, B. Lee, A. Monroy-Hernandez, P. Irani, Authoring data-driven videos with dataclips, *IEEE Transactions on Visualization and Computer Graphics* 23 (2017) 501–510. URL: <http://dx.doi.org/10.1109/tvcg.2016.2598647>. doi:10.1109/tvcg.2016.2598647, 36 cites:.

- [13] S. Bocconi, F. Nack, L. Hardman, Vox populi: a tool for automatically generating video documentaries, in: Proceedings of the sixteenth ACM conference on Hypertext and hypermedia, 2005, pp. 292–294.
- [14] I. Teixeira, P. Viana, M. Andrade, P. Carvalho, L. Vilaça, J. Pinto, T. Costa, S. Rapanakis, P. Jonker, Creating automatic storytelling videos from still images: a semantically-aware approach, 2021.
- [15] T. Gao, J. R. Hullman, E. Adar, B. Hecht, N. Diakopoulos, Newsviews: an automated pipeline for creating custom geovisualizations for news, in: Proceedings of the SIGCHI conference on human factors in computing systems, 2014, pp. 3005–3014.
- [16] E. Bernasconi, M. Ceriani, M. Mecella, T. Catarci, M. C. Capanna, C. D. Fazio, R. Marcucci, E. Pender, F. M. Petriccione, Arca. semantic exploration of a bookstore, Proceedings of the International Conference on Advanced Visual Interfaces (2020). URL: <http://dx.doi.org/10.1145/3399715.3399939>. doi:10.1145/3399715.3399939.
- [17] M. Ceriani, E. Bernasconi, M. Mecella, A streamlined pipeline to enable the semantic exploration of a bookstore, in: Digital Libraries: The Era of Big Data and Data Science - 16th Italian Research Conference on Digital Libraries, IRCDL 2020, Bari, Italy, January 30-31, 2020, Proceedings, volume 1177 of *Communications in Computer and Information Science*, Springer, 2020, pp. 75–81.
- [18] E. Bernasconi, M. Ceriani, M. Mecella, T. Catarci, M. C. Capanna, C. Di Fazio, R. Marcucci, E. Pender, F. M. Petriccione, Arca. semantic exploration of a bookstore, in: Proceedings of the International Conference on Advanced Visual Interfaces, AVI '20, Association for Computing Machinery, New York, NY, USA, 2020. URL: <https://doi.org/10.1145/3399715.3399939>. doi:10.1145/3399715.3399939.
- [19] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, ..., A survey on knowledge graph-based recommender systems, ... on Knowledge and ... (2020). URL: <https://ieeexplore.ieee.org/abstract/document/9216015/>.
- [20] X. Zou, A survey on application of knowledge graph, Journal of Physics: Conference Series (2020). URL: <https://iopscience.iop.org/article/10.1088/1742-6596/1487/1/012016/meta>. doi:10.1088/1742-6596/1487/1/012016.