# A Visual Analytics Approach for Semantic Multi-Video Annotation

Markus John[*] Kuno Kurzhals[†] Steffen Koch[‡] Daniel Weiskopf[§]

**ABSTRACT**

The annotation of video material plays an important role in many Digital Humanities research fields including arts, political sciences, and cultural and historical studies. The annotations are typically assigned manually and convey rich semantics in accordance with the respective research question. In this work, we present the concept of a visual analytics approach that enables researchers to annotate multiple video sources in parallel. It combines methods from the fields of natural language processing and computer vision to support the manual annotation process with automatically extracted low-level characteristics. The benefits of our approach are twofold. With the extracted annotations and their visual mapping onto a suitable overview visualization, we support scholars in finding the relevant sections for their high-level annotations on the one hand, and on the other hand, we offer an environment that lets them compare and analyze such annotations in several videos at once. Our concept can be flexibly extended with additional processing methods to simplify annotation tasks further.

**Index Terms:** Visual analytics, movie analysis, digital humanities

## 1 INTRODUCTION

In recent years, the number of new approaches to support movie analysis in the Digital Humanities has grown rapidly [3, 12, 23]. Most of them support analysis tasks with visualization, providing summarizations and overviews to better understand and communicate research results. However, the research on movie content, such as the analysis of plots and entities, is still a time-consuming and complex task that requires semantically rich annotations. For example, humanities scholars may want to investigate the relationship between characters and their evolution during the plot as well as specific events or places where characters interact with each other [12]. The common workflow to analyze movie content is watching the movie and taking notes.

Automatic approaches can support tasks such as finding relevant events, topics, characters, and places, which can then be summarized and visualized to simplify exploration and analysis. These approaches provide visual abstractions of video content and facilitate 'distant viewing' similar to the distant reading idea of Moretti [21]. Visual abstractions and summarizations can convey useful information and assist in exploring research questions as well as in verifying hypotheses and forming new research ideas [17]. For browsing and searching video content in large movie repositories, many retrieval approaches have been introduced [13, 34]. There are also quite a number of approaches to summarize the movie content including approaches that use storyboards [2, 9], plot view visualizations [16], or video skims [27] to provide an overview of the relevant content. Furthermore, several approaches that combine natural language processing (NLP) and visualization techniques have been developed for text summarization, extraction of the characters and places, and the interactive exploration of them. El-Assady et al. [7] or Stasko et al. [29] present examples of such approaches that automatically extract entities from text, enabling users to perform search queries and to explore their relationships visually. We argue that a combination of NLP, video processing, and visualization supports the annotation process and allows us to generate visual abstractions for analyzing video content.

However, a semantic gap remains which Smeulders et al. [26] described as "[...] the lack of coincidence between the information that can be extracted from the visual data and the interpretation that the same data have for a user in a given situation". Hence, typical high-level semantic annotation of video content as required for research in Digital Humanities cannot be acquired without human input, yet. Visual analytics aims at bridging this gap by combining automatic processing and interactive visualization to provide support for human experts in exploring, analyzing, understanding and finally annotating the data. The automatically extracted low-level characteristics offer the framework and starting points for human hypotheses building and annotation tasks.

We present the concept of a visual analytics approach for multi-video annotation derived from the "visual movie analytics" technique [18]. This previous work describes an interactive annotation approach for single movies in combination with semantic information from subtitles and movie scripts. However, the approach is limited to movies for which all three data sources are available. In this work, we discuss a more general approach that primarily focuses on video content and subtitles, broadening the applicability not only to movies, but to all video material containing subtitles. Furthermore, we focus on how to extend the concept for analyzing multiple video sources in parallel. This supports additional research tasks, as a topic, trend, or event sequence is often investigated in more than one video source. We developed the ideas and the concept in cooperation with domain experts from social sciences in a formative process. In addition, we implemented a prototypical approach of this concept as a basis for a further iterative development and evaluation by the domain experts. The primary goal is to enhance the support for analysis tasks that are part of the research efforts of them. In the following, we present their needs and real world research questions.

## 2 RESEARCH QUESTIONS

Four aspects can serve as the building blocks to summarize and analyze video content: *who ($W_1$)*, *what ($W_2$)*, *where ($W_3$)*, and *when ($W_4$)* [4, 19]. Those aspects do have relevance for the research questions from our domain experts. The social science scholars are particular focus on conflict research. Cinematographic works about political conflicts are an interesting data source for them for two reasons. These videos can be research objects on there own and can serve as training data set for inexperienced students to learn how to deal with complex annotation tasks. For example, they examine cinematographic works about Islamist terrorism and want to compare the Israeli and Palestinian perspective on this topic. The scholars are particular interested in sections of the video where conflicts that taken place and need to find out which characters were involved in them. It is therefore interesting to see: *Who ($W_1$)* was involved in the conflict? *What ($W_2$)* happened in this conflict? *Where ($W_3$)* and *when ($W_4$)* did the conflict take place?

[*]e-mail:markus.john@vis.uni-stuttgart.de
[†]e-mail:kuno.kurzhals@visus.uni-stuttgart.de
[‡]e-mail:steffen.koch@vis.uni-stuttgart.de
[§]e-mail:daniel.weiskopf@visus.uni-stuttgart.de

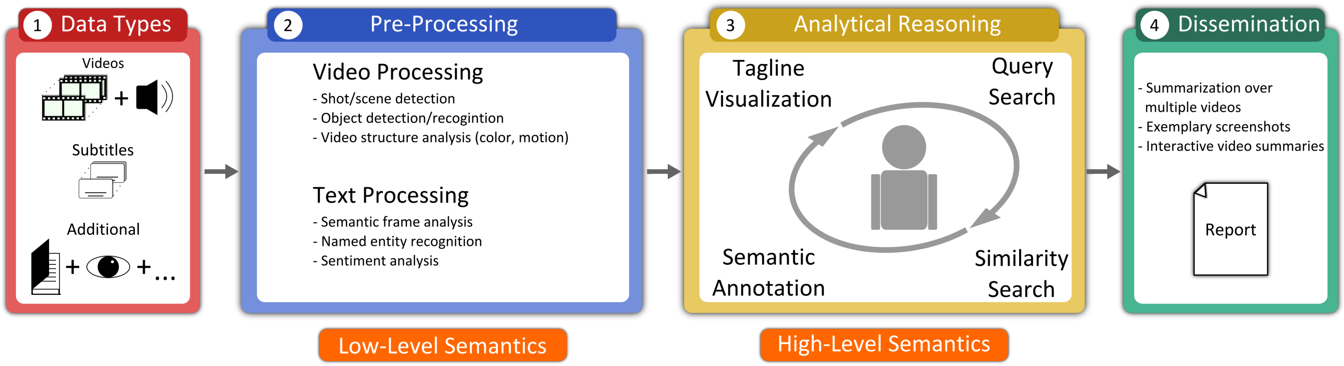



Figure 1: Proposed, generalized approach for multi-video analysis: pre-processing of the input data mainly focuses on the analysis of video and subtitle data. However, in the future additional data sources could also be processed. Analytical reasoning is an iterative process, focused on the human expert. We support the reasoning process by providing tagline visualizations for all videos as an overview, possibilities to carry out direct and similarity search on the processed data, and simple semantic annotation of video sections. Dissemination of extracted knowledge in an appropriate report form concludes the analysis.

However, the domain experts are also interested in further research questions such as: Are the female leads in romantic comedies much more emancipated today? If so, in which respects and in which not? Furthermore, they are interested into how the camaraderie was represented in older and newer movies about the Second World War or what kind of political understanding is expressed in series like "House of Cards"?

Our cooperators want to find scenes in which certain characters occur/co-occur or scenes with a similar meaning or plot. Furthermore, the stylistic means of shot compositions, such as camera angles and color palettes, can be an important part of the analysis process.

A big challenge for these tasks is to provide a good overview of the analysis, since our experts are interested in a detailed and compact representation of the results. This dissemination should be in an appropriate report form and include a summarization over multiple videos, example screenshots, and an overview of the annotated and extracted insights.

## 3 Conceptual Visual Analytics Approach

A fundamental idea of visual analytics is the combination of automatic data processing and interactive visual representation of the results for reasoning [31]. Figure 1 depicts our proposed approach for semi-automatic analysis and annotation of multi-video content. To this point, we discuss the conceptual integration of potential algorithms for video and text processing and how their results can be incorporated into an interactive visualization interface for annotation purposes. A prototypical implementation of our concept containing some of the discussed aspects is presented in Section 4. Furthermore, we distinguish between automatically extracted low-level characteristics (low-level semantics) and high-level annotations from our experts that are based on their knowledge and interpretations (high-level semantics).

### 3.1 Data Types

The current focus of our work is on videos and their according subtitles. As it is possible to create captions for any video (e.g., on YouTube), subtitle information is not only available from feature films. Subtitles are valuable for semantic content analysis since they constitute a form of (unstructured) explicit semantic abstraction. In contrast to the picture information from the video, this data is likely to produce more robust extraction results. Regarding the video, individual frames, motion, and audio play an important role in various analysis tasks and automatic preprocessing can be applied to provide enriched data for visualization as discussed in the next subsection. Additional sources could also provide further data to enrich specific analysis tasks. Examples could be the integration of eye tracking data to investigate visual attention, movie scripts [18], or original literature sources a movie is based on.

### 3.2 Pre-Processing

Video Processing The video can be processed with a wide range of available computer vision algorithms. Video-specific features can be investigated, e.g., to analyze the camera motion in a scene. Especially for the questions *who* ($W_1$) and *when* ($W_4$), established methods for person/object detection (*there is a person*) [33] and recognition (*this is person X*) [32] could be applied to hint the annotator where a person or object appears in the inspected videos. Hence, video processing can provide additional information that can be integrated with the results of text processing algorithms, for example, to align a person's name recognized in the text with the corresponding figure in the video. An analysis of the audio channel can be considered for highlighting musical themes or sounds (e.g., SoundRiver [15]). We suggest applying automatic techniques when possible, as they can provide semantic characteristics on a low-level (e.g., for questions *who* and *when*). The human expert can then summarize such low-level annotations into high-level semantic interpretations.

Text Processing To tackle the aforementioned research questions (Section 2), NLP methods can be applied to subtitles to detect characters, places, and to derive the relations between them automatically. For the questions *who* ($W_1$), *where* ($W_3$), and *when* ($W_4$) Named Entity Recognition (NER) methods are available, such as Stanford CoreNLP [28] or OpenNLP [1], which extract the entities automatically and allow one to indicate the relations between them across the plot [29]. To further inspect the relation between entities, weighting schemes, such as term *term frequency - inverse document frequency* (tf-idf) or $G^2$ [24], could be used to find out which terms or topics describe the relation between entities (*What* ($W_2$)). Another option is to use semantic role labeling approaches [10, 25], which typically rely on the output of part-of-speech (POS) taggers and the concept of *semantic frames* [8]. POS taggers process a text and identify words as nouns, verbs, adjectives, adverbs, etc., whereas a *semantic frame* is defined as a coherent structure of concepts and is invoked by respective target words in a sentence. The different semantic frame role representations express the abstract role that arguments of a predicate can take in the event [22]. For example, the SEMAFOR parser [5], combine both approaches and facilitate the automatic labeling of semantic roles. For finding scenes with characters stealing a treasure, for example, we would be interested in

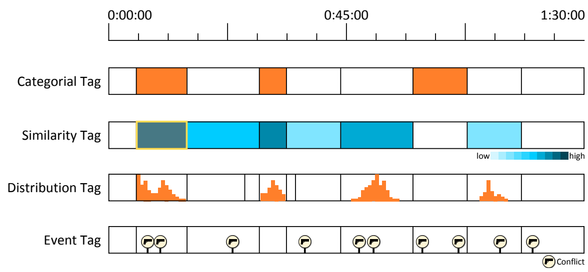



Figure 2: The tagline visualization offers multiple encodings of relevant data. Annotations can characterized through categorical, similarity, distribution, and event tags.

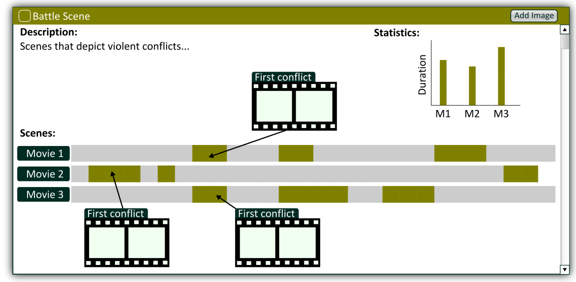



Figure 3: Illustration of a possible summarization report. Annotated scenes from multiple videos can be compared directly, including corresponding taglines, representative thumbnails, and statistics.

the theft semantic frame *What* $(W_2)$. The semantic frame is invoked by a variety of related terms such as thief, snitch, or pilfer. That way, users can find specific events faster and compare them.

### 3.3 Analytical Reasoning

To support an efficient, analytical assessment of the processed data, we propose a visualization of data elements and annotations by separate timelines, i.e. taglines (Figure 2). By applying specific queries or similarity searches, relevant scenes for a topic can be identified and annotated by a high-level, semantic annotation. This process can be repeated iteratively, including former annotations to create a higher abstraction level of the investigated topic.

Tagline Visualization For the depiction of search results and annotations, we propose a simple timeline visualization for two main reasons: (1) most researchers are familiar with them and (2) visual scalability. A timeline visualization is easy to interpret, as it is established in everyday life, e.g., in form of schedules. Especially with respect to the comparison of multiple video sources, timelines can be compressed spatially more easily than comparable visualizations. Additionally, multiple visual encodings can be applied to a simple timeline, covering a wide range of possible data features. Figure 2 shows a set of four visual encodings, suitable to cover numerous analysis tasks. Categorical tags can depict simple characteristics such as the occurrence of a person in a scene. Similarity tags depict the accordance of scenes with a selected one. Distribution tags depict quantities that may change over time, for example, the magnitude of motion over time. Event tags mark specific points in time when something happened (e.g., the beginning of a shooting). For the comparison of multiple video sources, corresponding taglines for each videos can be stacked on top of each other to provide a compact overview of all annotated scenes (see Figure 4).

Query and Similarity Search Initially, the preprocessed data can be searched for specific criteria or similar scenes can be identified by provided similarity measures. For example, a simple keyword search can ease the annotator's work by emphasizing timespans of potential interest. To identify two similar scenes, we could apply the *tf-idf* or $G^2$ measures which take into account the term distributions of the different scenes. This works very well for scenarios in which users are searching for similarly phrased text passages. However, if analysts want to find similar scenes that are not characterized through similar subtitles, the comparisons of word distributions are not helpful. To address this issue, we offer a search mechanisms based on *semantic frames*. In this case, we measure their overlap with the Jaccard coefficient for a pair of scenes to identify similar semantic scenes. The analyst can summarize these search results by assigning a new label to all relevant scenes, as an example, all scenes that contain a verbal conflict.

Semantic Annotation We propose a hierarchical label structure for semantic annotation, providing a flexible degree of abstraction. As an example, we can label the occurrence of individual persons in scenes. On the next abstraction level, we annotate individual groups they belong to. One abstraction level higher, we can annotate if this group is acting as protagonist or antagonist. With this annotation structure, individual research questions can be answered on the appropriate semantic abstraction level.

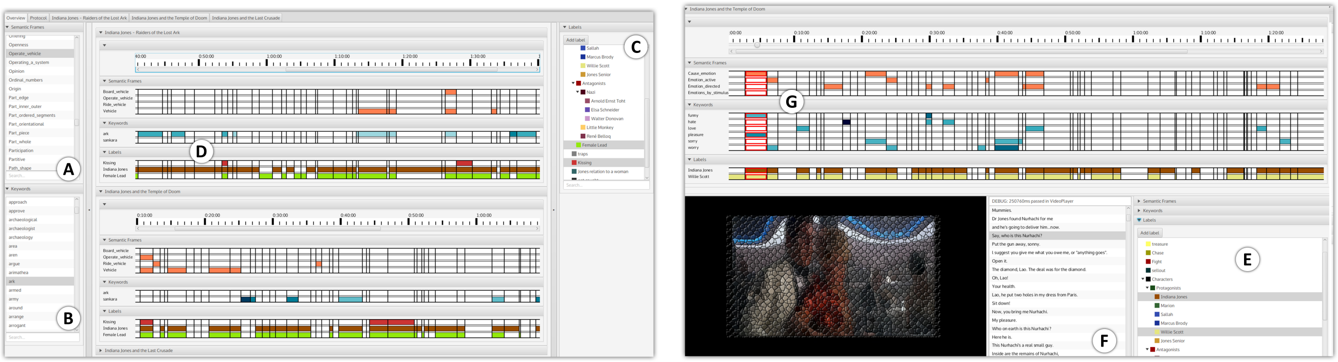
### 3.4 Dissemination

Assuming that semantic annotations on a video database will increase over time, many labels from other research questions might include information that is not necessary for a current analysis. Hence, it is important to be able to reduce and summarize the data to the currently relevant labels. Especially for communicating the results of an analysis, a summarizing report is essential. Figure 3 depicts an example of such a report. In this example, a summarization of three movies of all scenes with the label "Conflict Scene" is displayed. The report can contain only the relevant taglines, focusing on specific research questions and reducing visual overload. Label description text should be incorporated, as well as a dynamic option to add one or multiple thumbnails from the videos to the corresponding scenes. For example, thumbnails could help to illustrate when the first conflict in each movie occurred and what it looked like. Since quantitative data is available, either by the count of labeled scenes, or measured from other features, descriptive and inferential statistics could be integrated in the report. The example in Figure 3 shows the total duration of depicted conflict scenes for each movie. It would be possible to add specific measures, suitable to support research hypotheses. Such a report can be extended by all labels necessary to summarize the analysis results. Additionally, other summarizing visualizations such as theme rivers could be integrated in the report.

## 4 Example

We implemented a prototypical system to showcase our proposed approach for the annotation and comparison of multiple videos (Figure 4). It supports an overview mode (Figure 4a) that provides a summary over the loaded movies and their annotations. Furthermore, a detail view (Figure 4b) offers specific information for individual videos, including a video player and the subtitles (Figure 4 F). To this point, we included categorical tags and similarity tags to investigate the videos (Figure 4 D and G). Subtitles are processed by the *tf-idf* weighting scheme (Figure 4 B), and semantic frames identification (Figure 4 A). In addition, we present an overview of the hierarchical label structure (see Figure 4 C and E), which can be easily edited by the users. Further implementations of the discussed tagline visualizations are planned, as well as video and text processing algorithms to ease the analysis process.

We demonstrate the capabilities of the application by analyzing the first two Indiana Jones movies, namely *Raiders of the Lost Ark (1981) and Indiana Jones and the Temple of Doom (1984)*. We want




(a) Overview for multiple movies consists of A semantic frames, B keywords, C hierarchical label structure, and D categorical and similarity tags.

(b) Detail view for individual movies consists of E hierarchical label structure, F video player and the subtitles, and G categorical and similarity tags.

Figure 4: Prototypical implementation for multi-video annotation consists of a overview and detail view.

to find out how the different female protagonists have interacted with Indiana Jones, and analyze the character of the female lead and her relationship to Indiana Jones. Subsequently, we compare the respective annotations in each movie. After the two movies have been preprocessed, an overview page is shown as depicted in Figure 4a.

### The Character of the Female Lead and Her Relationship to Indiana Jones

In a first step, we go through the movie and annotate all the occurrences of Indiana Jones and female characters, as well as specific places (*who* ($W_1$) and *where* ($W_3$)). This way, we can easily identify the scenes where both, Indiana Jones and the respective woman, co-occur (*when* ($W_4$)). In order to get more information, we start a keyword search to get an overview of when "Indiana Jones" is mentioned together with terms like "kissing" or "love" in the movie (see Figure 4a D) (*what* ($W_2$)) and *when* ($W_4$)). To find similar scenes across all movies, we select one of the scenes and perform *semantic frame* similarity search (*what* ($W_2$)). With the aid of a highlighting, we can easily recognize the relevant movie scenes in the overview and *when* ($W_4$) they occurred.

Next, we switch to the detail view to analyze the occurrences in more detail (Figure 4b). While analyzing some of the occurrences with the video player and the subtitles, we find out that "Marion Ravenwood", the female lead, often occurs without Indiana Jones and vice versa in movie *Raiders of the Lost Ark*. In all of the analyzed scenes, she is portrayed as a strong and independent woman (*what* ($W_2$)). In the following, we switch to movie *Indiana Jones and the Temple of Doom* in the detail view since we are interested in examining the next female lead "Willie Scott". With the help of the taglines, we identify that she never appears without Indiana Jones (*where* ($W_3$) and *when* ($W_4$)). After watching the movie scenes and reading some of the subtitles, we discover that many of her early interactions with Jones involve complaints about the circumstances. In addition, we have have the assumption that she is a very emotional person. To verify this assumption, we activate different semantic frames and keywords, which could describe a emotional person, and we find out that terms indicating emotionality co-occur often with her and Indiana Jones (see Figure 4b) G (*what* ($W_2$)).

The example shows that our approach facilitates experts in finding and analyzing movie scenes faster. However, we could further support the analysis through visualizations or automatic methods.

## 5 Conclusion

We presented a general approach for the semantic annotation and dissemination of time spans in videos with subtitles. The implemented prototype, which comprises many of the previously discussed methods, provides first insights and serve as basis for further discussion with our domain experts. For the specific research questions in Digital Humanities, the implementation of corresponding feature detection algorithms (e.g., faces, sentiment analysis, sounds) could be considered according to the task. With the proposed tagline visualization concept, a wide range of such features could be represented in a compact, interpretable way. However, for the future there are open questions and challenges, which we would like to address:

An important, yet untackled point is the detection of scenes in a video. Numerous algorithms to summarize shots into scenes have been developed [6]. However, the semantic interpretation of a scene can be ambiguous. As an example, different shots of similar visual features could be comprised into a scene. But it would also be possible that the same video shots depict a conversation considering two different topics. In such a case, separating the shots in two scenes might be reasonable, too. In unedited video content (e.g., smartphone videos), the detection of scenes can be even more complicated, because no shots are available to summarize. Hence, it is necessary to incorporate the human user into the process. Different algorithms could provide suggestions about initial scene separations that can be selected and modified accordingly.

Another challenge is to provide further visualizations, such as a plot view [20] to get a rough idea of the storyline, which can support users by their analysis. Tanahashi and Ma [30] introduce design considerations, which are based on annotated book information, for generating storyline visualizations automatically. Since our annotations are very similar, we could use their approach to provide such a visualization. Furthermore, we want to offer a network visualization [11] that represents connections between characters and their evolution during the plot.

Furthermore, established annotation methods from computer vision tasks could provide means for detailed search queries. To this point, we considered only the annotation of time spans. If we incorporate the annotation of specific image regions, combined with computer vision algorithms, it would be possible to search for specific objects in multiple videos, maybe in combination with other labels that were annotated before.

A further possibility might be to extent the approach with additional descriptive features, such as audio [14], or relational information in order to provide a more flexible analysis. The open question here is whether these features can be displayed in our tagline visualization or if we need new representations.

We plan to further develop our approach in close cooperation with the domain experts. In a formative process, we can tailor required features and specific visualizations for their needs.

## Acknowledgments

We would like to thank Cathleen Kantner from the University of Stuttgart for her constructive feedback on our approach. In addition, we thank our students Clemens Lieb, Heiko Roggenbuck, Marco Radic, und Verena Schütz who implemented the prototypical approach in a student project. This work was funded by the German Federal Ministry of Education and Research (BMBF) as of the Center for Reflected Text Analysis CRETA at University of Stuttgart.

## References

[1] Apache Foundation. Apache opennlp. `http://opennlp.apache.org`, 2014. version 1.5.3.

[2] J. Boreczky, A. Girgensohn, G. Golovchinsky, and S. Uchihashi. An interactive comic book presentation for exploring video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pp. 185–192, 2000.

[3] M. Burghardt, M. Kao, and C. Wolff. Beyond shot lengths–using language data and color information as additional parameters for quantitative movie analysis. In *Proceedings of the Digital Humanities*, DH '16, 2016.

[4] B. W. Chen, J. C. Wang, and J. F. Wang. A novel video summarization based on mining the story-structure and semantic relations among concept entities. *IEEE Transactions on Multimedia*, 11(2):295–312, 2009.

[5] D. Das, D. Chen, A. F. Martins, N. Schneider, and N. A. Smith. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56, 2014.

[6] M. Del Fabro and L. Böszörmenyi. State-of-the-art and future challenges in video scene detection: a survey. *Multimedia Systems*, 19(5):427–454, 2013.

[7] M. El-Assady, R. Sevastjanova, B. Gipp, D. A. Keim, and C. Collins. NERE*x*: Named-Entity Relationship Exploration in Multi-Party Conversations. *Computer Graphics Forum*, 36(3):213–225, 2017.

[8] C. Fillmore. Frame semantics. *Linguistics in the Morning Calm*, pp. 111–137, 1982.

[9] M. Furini, F. Geraci, M. Montangero, and M. Pellegrini. Visto: Visual storyboard for web video browsing. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, CIVR '07, pp. 635–642, 2007.

[10] D. Gildea and D. Jurafsky. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288, 2002.

[11] J. Heer and D. Boyd. Vizster: visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 32–39, Oct 2005.

[12] E. Hoyt, K. Ponto, and C. Roy. Visualizing and analyzing the hollywood screenplay with scripthreads. *Digital Humanities Quarterly*, 8(4), 2014.

[13] W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank. A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(6):797–819, 2011.

[14] Y. Hu, J. S. Ren, J. Dai, C. Yuan, L. Xu, and W. Wang. Deep multimodal speaker naming. In *Proceedings of the 23rd ACM international conference on Multimedia*, pp. 1107–1110. ACM, 2015.

[15] H. Jänicke, R. Borgo, J. S. D. Mason, and M. Chen. Soundriver: Semantically-rich sound illustration. *Computer Graphics Forum*, 29(2):357–366, 2010.

[16] M. John, S. Lohmann, S. Koch, M. Wörner, and T. Ertl. Visual analytics for narrative text - visualizing characters and their relationships as extracted from novels. In *Proceedings of the 7th International Conference on Information Visualization Theory and Applications*, IVAPP '16, pp. 27–38. Scitepress, 2016.

[17] S. Koch, M. John, M. Worner, A. Muller, and T. Ertl. Varifocalreader in-depth visual analysis of large text documents. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1723–1732, Dec 2014.

[18] K. Kurzhals, M. John, F. Heimerl, P. Kuznecov, and D. Weiskopf. Visual movie analytics. *IEEE Transactions on Multimedia*, 18(11):2149–2160, 2016.

[19] A. Liu and Z. Yang. Watching, thinking, reacting: A human-centered framework for movie content analysis. *International Journal of Digital Content Technology and its Applications*, 4(5):23–37, 2010.

[20] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2436–2445, Dec 2013.

[21] F. Moretti. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.

[22] M. Palmer, D. Gildea, and N. Xue. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103, 2010.

[23] J. Pause and N. Walkowski. The colorized dead: Computerunterstützte analysen der farblichkeit von filmen in den digital humanities am beispiel von zombiefilmen. In *Proceedings of the Digital Humanities im Deutschsprachigen Raum*, DH '16, 2016.

[24] P. Rayson and R. Garside. Comparing corpora using frequency profiling. In *Proc. Workshop on Comparing Corpora*, pp. 1–6, 2000.

[25] D. Shen and M. Lapata. Using semantic roles to improve question answering. In *Emnlp-conll*, pp. 12–21, 2007.

[26] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[27] M. A. Smith and T. Kanade. Video skimming and characterization through the combination of image and language understanding techniques. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 775–781, 1997.

[28] Stanford CoreNLP. A suite of core nlp tools. `http://nlp.stanford.edu/software/corenlp.shtml`, 2014. version 3.3.1.

[29] J. Stasko, C. Görg, and Z. Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[30] Y. Tanahashi and K.-L. Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, 2012.

[31] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006.

[32] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

[33] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[34] A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 11(1):81–93, 1999.