The Paper or the Video: Why Choose?

Hugo Mougard, Matthieu Riou, Colin de la Higuera, Solen Quiniou, Olivier Aubert
Laboratoire LINA UMR CNRS 6241
Université de Nantes, France
{hugo.mougard, cdlh, solen.quiniou, olivier.aubert}@univ-nantes.fr, matthieu.riou@etu.univ-nantes.fr

ABSTRACT

This paper investigates the possibilities offered by the more and more common availability of scientific video material. In particular it investigates how to best study research results by combining recorded talks and their corresponding scientific articles.

To do so, it outlines desired properties of an interesting e-research system based on cognitive considerations and considers related issues. This design work is completed by the introduction of two prototypes.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**];
I.7.2 [**Document Preparation**]: Multi/mixed media

Keywords

hypermedia, e-research, multimodal alignment

1. INTRODUCTION

Teachers and researchers in education science have investigated online education for a long time. The growing popularity of Massively Open Online Courses (MOOCs) has brought new issues in this field: new needs, new demands, new questions. In addition to this important change of scale, the overall improvement in broadband availability has turned videos, that were before hard to use online, into primary study material. Yet MOOCs and other online learning sites, as well as the sites offering filmed tutorials given at conferences and summer schools usually all employ videos separately from the rest of the material with little consideration for a harmonious combination with the rest of the learning material.

Furthermore, more and more video based research material has been developed. The pioneering work done by `videolectures.net` to film conferences and events, the possibility of reducing the cost of production and edition of video content (eg `http://opencast.org/matterhorn/`), the growing interest by the institutions themselves to keep a visual record of the events they have promoted, all this has contributed to build a large but unorganized collection of video presentations of research material.

In many cases, the videos correspond to talks given at conferences, talks for which an article has been written and is published in the proceedings of the conference. This represents a great opportunity for e-research. We note that if traditionally studying a research article has been done through its article alone, it is mainly because it used to be the only material available. As detailed above, a growing corpus of recorded talks and their corresponding articles is available. This brings us to the following question: "Is working on the article still the best way to study a research result? Or should researchers listen to the talk?". In this paper, we investigate the following proposal "the best way would be to be able to do both in a shared environment, with the possibility of switching effortlessly from one mode (reading) to the other (viewing)".

Hypermedia e-learning is a well established research field [2] which gained renewed attention with the change of scale induced by the arrival and growing popularity of MOOCs [8]. This growth has been enabled by a range of factors, beside broadband availability. We hypothesize that the most important ones are (1) the availability of good audio transcriptions—an important stepping stone towards correct handling of multimedia contents; (2) the ongoing effort to better handle multimedia by the hypertext, hypermedia and signal processing communities; (3) the appearance of a strong movement in favour of open data and open education resources.

Hypermedia e-research is closely related to hypermedia e-learning. Therefore—considering the favourable context—it is sensible for researchers to focus on both domains.

The following sections present work that has been done in the past concerning these questions and then focus on design considerations and related challenges. We propose in Section 4 an innovative prototype that allows a combined usage of a scientific paper and its filmed conference presentation. Finally, we link the extensions of our work with a semantic driven approach.

2. RELATED WORK

Previous work in domains related to proper handling of multimedia academic communication can take many aspects. We will focus mainly on alignment and cognitive issues.

Aligning documents of different modalities is vital to be able to create an interesting experience for scientists that

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.
WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742017.

want to learn from both an article and its related resources (video presentation, slides, etc). However the technical capability of synchronizing different media is not sufficient. It is also necessary to take into account the cognitive specificities of each medium to combine them properly.

2.1 Multimodal alignment

This work treats multimodal alignment as a natural language processing task. It is also a core task in semantic web and future work will benefit from investigating both domains jointly.

Alignment between talk transcription and text can be seen as a task combining two different subtasks: monolingual alignment and multimedia information retrieval.

2.1.1 Monolingual alignment

In 2003, Barzilay and Elhadad introduced a method to learn the alignment between two documents at a macro level (thematic units), through clustering similarity analysis, and then at a micro level (sentences) [1], thanks to a dynamic programming procedure. In 2006, Nelken and Shieber showed that the use of a different learning process (logistic regression on TF-IDFs) completed by a global alignment procedure gave better result while being globally simpler to implement [7]. Both approaches require annotated data that we do not have at our disposal. We plan to adapt them to learn in a reinforcement learning fashion based on user behaviour.

2.1.2 Multimedia Information Retrieval

As soon as 2001, yearly campaigns have been organized to investigate information retrieval in a multimedia context [12]. In recent years, we can mention `MediaEval` and more specifically its Search & Hyperlinking task [3] as an interesting source of research on the subject. This work is important both to propose a system (to correctly segment and link videos) and evaluate it (standard IR metrics lack in efficiency in a multimodal context).

2.2 Cognitive experience

In a scholar context, it is possible to use different media either separately or jointly [9], taking the form of hypermedia documents. Through hypermedia, in our scholar context, we want to play on the strengths and weaknesses of each involved media to maximize learning efficiency. Hypermedias where video takes a pivotal role have shown their value in learning contexts [2, 5] and we would like to build upon this experience.

3. AMBITIONS, GOALS AND RESEARCH CHALLENGES

The goal we pursue is to develop a system which can (1) automatically take as input a bunch of research papers and their corresponding videotaped and transcribed conference presentations; (2) build the necessary alignments between these materials; and (3) propose to a user the possibility of viewing the video and reading the paper "at the same time". The user should be able to interrupt the viewing at any moment and find herself on the correct page and paragraph in the article. Conversely she may click on a paragraph to be able to get the video to play from the best corresponding moment.

3.1 Research activity potential

Presentation videos bring interesting benefits to paper study, relatively to scientific articles:

- the author's bias is easier to discern (why is the work important);
- complex figures may be explained incrementally, reducing greatly the complexity of the learning effort;
- the vocabulary might be simpler, which is interesting for out-of-speciality learners.

Presentation videos also have their drawbacks compared to scientific articles:

- the content is usually harder to skim through. Usually, even for limited information needs, there is no option but to watch the entire video;
- they usually lack formal definitions;
- they usually lack references.

Once those characteristics are defined, the goal is to come up with a way to study scientific papers that minimizes the drawbacks of each medium and maximizes their benefits.

3.2 Use-cases definition & interface design

The first challenge to overcome is the definition of use cases for hypermedia e-research. From classical online tutorials to speedy evaluations of the interest of articles, many possibilities arise and each requires thoughtful consideration to correctly exploit the specificities of the hypermedia material. Once the use-cases of hypermedia e-research are properly defined, the focus can shift to the design of one or several interfaces that address them.

A first analysis of this question is that one can imagine to favour one medium (and center the interface around it) or to combine everything in a more leveled way. The following sections detail our thoughts on the possibilities offered by each option.

3.2.1 A hypervideo model

Among hypermedia documents, video content can be used as atomic, sequential and not easily navigable clips mainly used as support to give a better idea of a concept [6]. However, video integration can also be carried out more deeply, dynamically creating non-linear or user-defined navigation paths within the videos, qualifying such documents as hypervideos, where a hypervideo can be considered as an interactive video-centered hypermedia document, that brings additional capabilities and improved interactivity to videos [11].

3.2.2 Video enriched by the article

This is the closest option to what is currently used in the vast majority of MOOCs and `VideoLectures.net`. Video only courses have some shortcomings. Here we show how to address some of them by using textual content:

Unskimmability The structure of an article provides valuable anchors that can be used to skim through the video. The article can also be used to replace the video content when the part is too lengthy for the user to go through. It is also easier to summarize a text than a video.

Lack of formal definitions When a user decides to "go through the maths", video support is often not enough. The article can complement some parts nicely with precise definitions.

Lack of references References from the article can either be used directly or lead to interesting hyperlinking, for example to other video presentations of referenced articles (and more precisely the parts that are relevant to the current object of study).

Lack of details This drawback is two-fold: (1) the lack of details about what is told during the presentation is addressable similarly to the lack of formal definitions (2) the parts of the article that did not make it to the video presentations can also be very valuable to the user and be retrieved by looking for the non aligned part of the different media.

3.2.3 Article enriched by video

The other way to present multimedia material would be to consider the scientific article as the main object of study and the presentation video as complementary material. What follows is a list of shortcomings of a study based on an article alone and some ways to address them using a talk:

Figure explanation The dynamic nature of a videoed talk is often better suited to explain complex figures than a scientific article. Integrating a clip of the author explaining some of the hardest parts of the concept at play in the article figures is very helpful to the quick understanding of the scientific material.

Author bias Videos of talks are more likely to be subjectively biased than scientific articles. It's therefore interesting to use them to understand what is the author bias on his work.

Global vision The talk will often contain more context and global vision regarding the motivations and strategy behind a scientific work. This can be valuable when the article alone doesn't answer all the information need of the user.

Synthesis Quite often, time restrictions of conference presentations lead authors to shorten and summarize their work. The video can therefore be used to get the underlying idea behind a lengthy section of an article.

3.2.4 Intertwined article and video

Instead of focusing on a particular medium to expose scientific material and enrich it with complementary material, it is possible to use all the resources in a specific narrative sequence to guide the user through the scientific content. For example, it would be possible to let the speaker introduce the context, then give an overview of the state of the work from the article, come back to the video for a dynamic explanation of a summarizing figure, etc. This is an option to keep in mind for the development of e-research that requires a high level of content understanding.

3.3 Algorithmic issues

Another big part of the challenges to consider is the definition of the objects and algorithms that allow the proper manipulation of the multimedia material.

Conceptually, the essential building block is made of *typed alignments* between the different resources, so that their relations (thematic, didactic, etc) can be easily exploited in hypermedia systems. Those alignments need to satisfy local but also global constraints. For example, relying on local proximity to align an article and its related talk can lead to jumps from one end of the article to another during the video viewing. That may prevent the user from getting any useful information out of the additional resource.

These typed alignment questions have important overlaps with work in the semantic web, hypertext and hypermedia communities where they are the very core of the domain. Progress over those issues will require cooperation with those communities.

3.4 Evaluation

The quality of a hypermedia application is hard to measure. It is highly subjective and depends on the information need of the user, on her learning habits and on her level of interest.

A first step is to conduct an intrinsic evaluation of the system to evaluate. To achieve that, the hypermedia application can be understood as an answer to an information need, with queries going to and from both the article and the talk. With this vision in mind, it is possible to use multimedia information retrieval metrics [4] to assess the quality of a system. However, just as for search diversification, recommendation systems or machine translation, an extrinsic evaluation is required to properly measure the strengths and weaknesses of the system.

A way to conduct this extrinsic evaluation is to compare two systems through A/B testing. However it has been noted in the past that easily obtainable metrics don't lead to interesting evaluations [10] and we hypothesize that these difficulties will also apply for hypermedia scientific communication. Clever metrics following [10] are therefore necessary.

4. CURRENT PROTOTYPE

We have developed two prototypes[1] to investigate the issue of how to best study scientific papers. In both cases, we are working on a restricted set of videos and with preliminar algorithms building the alignments between the paragraphs of the research paper and the scenes in the video. These are currently built by using information retrieval techniques.

The first prototype, shown in Figure 1, highlights article paragraphs related to the video at a given moment, and allows bidirectional navigation.

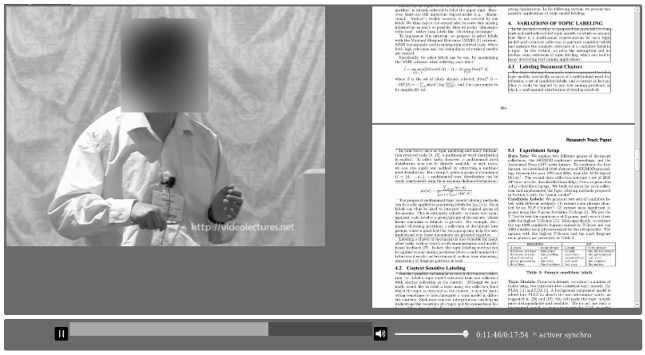

Figure 1: Joint navigation prototype screenshot.

[1] `http://alignement.comin-ocw.org/`

The second prototype shown in Figure 2 aligns a scientific paper and its video counterpart based on the TF-IDF cosine similarity between a transcription segment and an article paragraph: this results in the presentation of a weighted bipartite graph to the researchers so that they can analyze the quality of the alignment and of the computed weights.

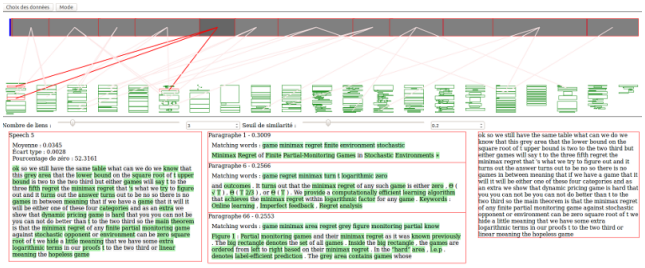

Figure 2: Alignment prototype screenshot.

4.1 Technical choices

The following section gives a very broad overview of the system. The code is available on github[2]. All the tools and technologies used are open-source and allow for easy substitutions and iterations.

Server side, the joint navigation prototype makes use of `Poppler`[3] to spatially delimit paragraphs in an article and the alignment prototype uses a `NLTK`[4] back-end that is responsible for the actual alignment algorithm.

Client side, the joint navigation prototype only makes use of standard web techniques and alignments are presented using the Javascript library `d3.js`[5].

4.2 Usability choices

Usability design is a work in progress. At this point we focused the interface strong points around fast prototyping and development iterations. While the prototypes are already great tools to investigate hypermedia coordination and use—for *developers* and *researchers*—they are not yet ready for *users*.

5. CONCLUSION

The prototype we propose allows to envisage hypermedia scientific communication with a different perspective. A number of research challenges, concerning automatic transcription of videos, machine learning, natural language processing, computer-human interfaces, cognition have to be dealt with if we want to address more complex questions:

- The correct alignment should clearly be guided by intention: why has the user stopped the video? because she was bored? because she couldn't understand one of the last words or concepts? Depending on the question, the text fragment to which she should be addressed would be different.
- The case of aligning the talk with the paper, when they are comparable, should only be considered as the first interesting challenge. A better one would be to get the answer to a question written by one author in a video of another.

[2] `https://github.com/Matthieu-Riou/Multimodal-Alignment`
[3] `http://poppler.freedesktop.org/`
[4] `http://www.nltk.org/`
[5] `http://d3js.org/`

Acknowledgement

The authors acknowledge support from the National Research Agency in the 'Investments for the Future' program under reference ANR-JO-LABX-07-0J (COCo project), and the help of colleagues from JSI, Ljubljana with the Videolectures material.

6. REFERENCES

[1] R. Barzilay and N. Elhadad. Sentence alignment for monolingual comparable corpora. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 25–32. Association for Computational Linguistics, 2003.

[2] T. Chambel, C. Zahn, and M. Finke. Hypervideo design and support for contextualized learning. In *IEEE International Conference on Advanced Learning Technologies*, pages 345–349, 2004.

[3] M. Eskevich, G. J. Jones, R. Aly, R. J. Ordelman, S. Chen, D. Nadeem, C. Guinaudeau, G. Gravier, P. Sébillot, T. De Nies, et al. Multimedia information seeking through search and hyperlinking. In *Proceedings of the 3rd ACM conference on International Conference on Multimedia Retrieval*, pages 287–294. ACM, 2013.

[4] M. Eskevich, W. Magdy, and G. J. Jones. New metrics for meaningful evaluation of informally structured speech retrieval. In *Advances in Information Retrieval*, pages 170–181. Springer, 2012.

[5] J. J. Leggett and F. M. Shipman. Directions for hypertext research: Exploring the design space for interactive scholarly communication. In *Proceedings of the fifteenth ACM Conference on Hypertext and Hypermedia*, pages 2–11, 2004.

[6] T. Navarrete and J. Blat. VideoGIS: Segmenting and indexing video based on geographic information. In *5th AGILE Conference on Geographic Information Science*, pages 1–9, 2002.

[7] R. Nelken and S. M. Shieber. Towards robust context-sensitive sentence alignment for monolingual corpora. In *In Proc. EACL*. Association for Computational Linguistics, 2006.

[8] L. Pappano. The year of the mooc. *The New York Times*, 2(12):2012, 2012.

[9] H. E. Pence. Teaching with transmedia. *Journal of Educational Technology Systems*, 40(2):131–140, 2011.

[10] F. Radlinski, M. Kurup, and T. Joachims. How does clickthrough data reflect retrieval quality? In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 43–52. ACM, 2008.

[11] M. Sadallah, O. Aubert, and Y. Prié. CHM: an Annotation- and Component-based Hypervideo Model for the Web. *Multimedia Tools and Applications*, Oct. 2012.

[12] A. F. Smeaton, P. Over, and R. Taban. The TREC-2001 video track report. *Proceedings of TREC-2001*, 2001.