Topics Compass: Uncovering Trending Topics for Optimised Media Content Publication

Lyndon J.B. Nixon

MODUL Technology GmbH, Vienna, Austria, nixon@modultech.eu

Arno Scharl

webLyzard technology gmbh, Vienna, Austria, scharl@weblyzard.com

Rasa Bocyte

Netherlands Institute for Sound and Vision, Hilversum, the Netherlands, rbocyte@beeldengeluid.nl

ABSTRACT

This demo presents the *Topics Compass*, a Web based dashboard for media organizations to uncover trending topics in online discourse. By focusing future content publications on these trending topics, they can optimize the reach and engagement with their content on digital channels.

KEYWORDS

Predictive Analytics; Data Analysis; Data Visualisation; Content Success Metrics; TV Intelligence

ACM CLASSIFICATION

G.3

1 Introduction: ReTV Project Goals

The ReTV Project (www.retv-project.eu) is an EU Horizon 2020-funded research initiative whose goal is to enable media organizations including broadcasters and media archives to optimize the publication of their media content across digital channels. Through analysis of the success of past content publication, we are building cross-channel prediction models to anticipate which (type of) content will potentially be most successful on which channel at what time in the future. This can inform organizational decisions regarding which content to publish as part of an optimized content publication strategy. This includes the creation of content summaries for different channels (e.g. social media video is generally shortened to the key segments to highlight to a user of that channel) as well as the recommendation of when and where to publish those summaries to optimize reach and engagement with the audience.

To provide these innovations to media organizations, a number of use case applications have been developed. A Topics Compass allows the exploration of trending topics among an audience and the predicted future topics of interest; the Content Wizard uses the content publication recommendations derived from those predictions and provides organizations a tool to select and summarize relevant content for the optimal publication. This demo submission will focus on the Topics Compass.

2 Scientific and Engineering Concept: Trans-Vector Platform

The fundamental engineering concept to realize this goal is an integrated platform known as the Trans-Vector Platform (TVP) [1]. Its architecture follows a service-oriented model to combine Web services from different project partners into an end-to-end workflow. Persistent and scalable data stores are hosted for both binary data (local recordings of TV programming for analysis) and metadata (TV content annotations created by the analysis). The core concept behind the TVP is that monitoring and analyzing online content about TV programming may be used to better understand TV audiences and that understanding can be used by AI-driven prediction algorithms so that we can recommend the best content to publish on a particular channel at a particular time to maximize the content's reach among the audience (Fig. 1, left).

3 Novelty of the Approach: A TV Intelligence Dashboard

The novelty of our approach lies in extracting content and associated success metrics from many different sources combined with the use of domain knowledge in the annotation of that content (using NER/NEL [3]) to identify and track distinct, disambiguated topics, and the prediction of success of future content publication based on its topics, using a combination of statistical methods, event knowledge and topical trends. As shown in Fig. 1 (right), the sources of data for this approach are twofold: a Metadata Search Index hosted by project partner webLyzard stores in scalable Elastic indexes metadata automatically extracted from Websites and social media posts collected daily based on their relevance to TV content (broadcasters, programs). Our available data analysis components allow us to extract from the online textual content keywords and entities using state of the art NLP and NER approaches [4].
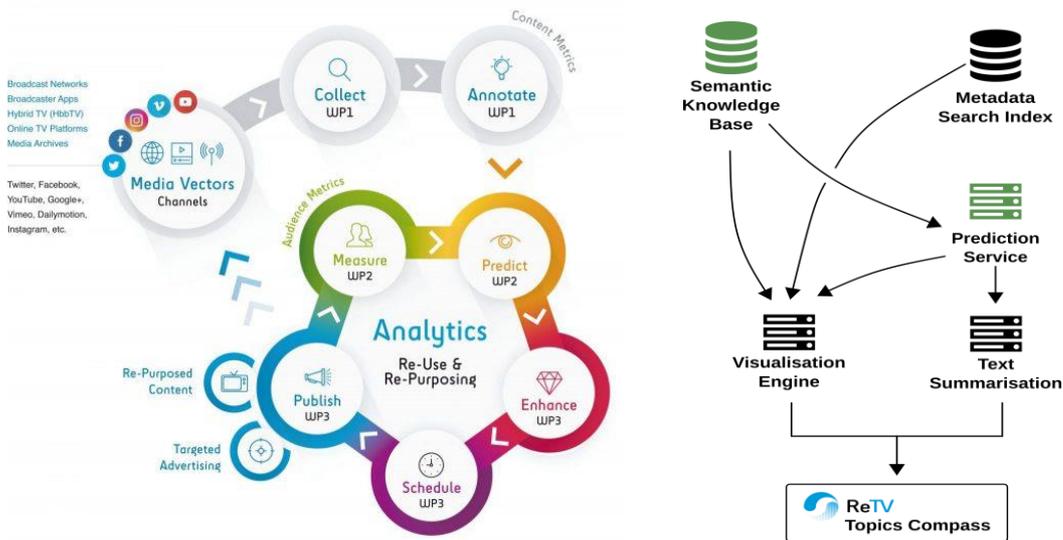
Figure 1. Novel ReTV approach to develop TV Intelligence (left);
TVP sub-architecture for the Topics Compass ([2]; right)

Using aggregation functions in Elastic, we can detect at scale emerging topics from the daily ingestion of ca. 50 000 news articles, 25 000 social media posts and 800 TV/radio Web articles. This is supported by a Semantic Knowledge Base which provides structured domain knowledge about persons, organizations, locations and events which form part of those topics, allowing concept disambiguation as well as advanced reasoning over the annotated data. Quantitative metrics extracted from the ingested data ordered by time – topic frequency (a measure of popularity), sentiment (a measure of desirability) or semantic disagreement (a measure of controversy) – are used by a Prediction Service to develop quantitative predictions for the future, i.e. how popular, desirable or controversial a topic is expected to be at a future time. The service implements a hybrid prediction model [5], combining the quantitative predictions (using regression models and machine learning) with future event knowledge (from the Semantic Knowledge Base) and future trending topic detection (from extracting future dates from content and associating content keywords with that future date). As opposed to Business Intelligence, or even Media Intelligence, this is something we call "TV Intelligence". Just as with classical BI tools, a dashboard represents an online user interface that presents the collected and analyzed data for exploration and understanding. The metadata, entity knowledge from the Semantic Knowledge Base and the predicted values for topics can all be fed into a Visualisation Engine which produces data visualisations for embedding in a TV Intelligence dashboard interface, including trend charts, keyword graphs, tag clouds and geomaps. Fig. 1 (right) illustrates this TVP sub-architecture which enables one use case application - *Topics Compass* – which will be the subject of our demo (see Section 4).

While media organizations may already use social media or Web monitoring tools to extract insights about audience reaction to their media content, the novelty of our approach is that we measure success metrics from content across various types of media vectors (channels such as the Web, social media and broadcast TV) and use these metrics to predict the future success of content published on any channel.

Knowing which topics to focus on for optimal publication, we have additional functionality to enhance content items so that they are summarized to those segments related to those topics, schedule their publication to the optimal time and publish to the optimal channel (these functionalities are combined in a separate tool called the Content Wizard).

4 Demo: Topics Compass

Our demo will show the ReTV-developed Web dashboard that provides a means to browse and explore the TV Intelligence insights extracted by the Trans Vector Platform, including through data visualizations, which we call the *Topics Compass*. We have two instances of the Topics Compass initialized: (i) one for the German broadcaster Rundfunk Berlin Brandenburg (RBB) analyzing German news and social media data for news stories about the Berlin and Brandenburg region, global news and discussions around RBB's TV and radio shows and with an UI in the German language and (ii) one for the Dutch media archive the Netherlands Institute for Sound and Vision (NISV) analyzing Dutch and English news and social media data for topics around Dutch television and European culture and history, with an UI in English. The Topics Compass allows a user to explore an overview of the main (general) topics of discussion in news, (TV related) social media or (TV related) Websites among the local or international audience. A topic configuration tool allows us to define and track topics in terms of changes in popularity among the audience: both general topics like "cultural heritage" which may be used to identify specific popularity trends (a museum, an artist, etc.) and specific events like "Eurovision Song Contest" which may be used to follow what the audience is focusing on over time (e.g. the varying popularity of different countries' entries). For example, NISV uses the Topics Compass to decide what topic would interest their audiences on social media channels based on tracking the 'cultural heritage' topic and then finding a video in their archive which relates to that topic.
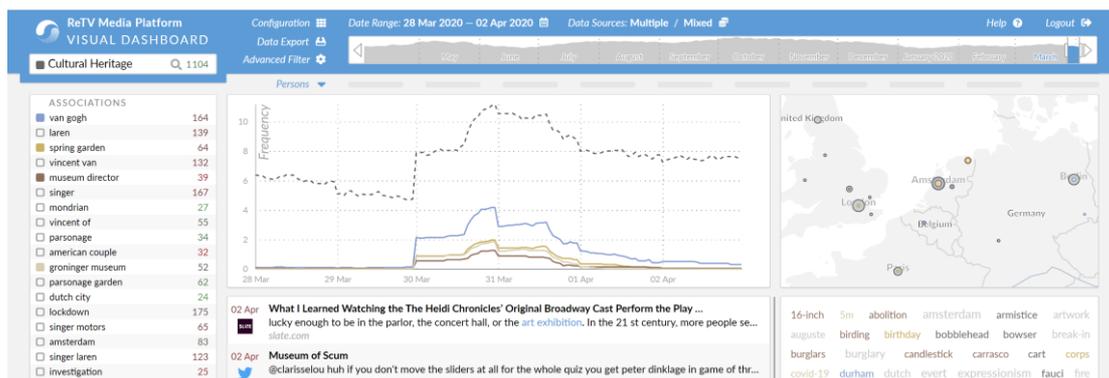
Fig. 2. Screenshot of the Topics Compass monitoring cultural heritage topics

Fig. 2 shows part of the Topics Compass dashboard. The user at NISV is monitoring online Web and social media content for cultural heritage topics. A jump in discussion about a cultural heritage topic is detected on the March 30, and the matching documents show this is because news broke of a Van Gogh painting being stolen from the Singer Laren museum. The top associations with cultural heritage right now among the audience is therefore "Van Gogh", "Laren" and other related keywords. EUScreen might post to their channels a video of a visit to the Singer Laren museum or of a Van Gogh painting, since these are the topics their audience is more likely to pay attention to.

The Topics Compass also visualizes the predicted future popularity of topics using a combination of machine learning regression models, knowledge of future events and topic prediction through future date extraction from documents (see [5] for last evaluations of the accuracy of the predictive approach). The user can access this "prediction mode" to explore which topics are predicted to be more popular on selected future dates (or date ranges). This can be used to guide a media organization's content publication strategy, planning the publication of content related to a certain topic at the best time and channel in which it is expected to achieve maximal reach to the intended audience. Fig. 3 shows the current prediction of topics for June 2020 based on the international news sources. The top predicted associations (topics that will be in the attention of the global audience at that time) are 'democrats' and 'China'. Early

June we can see a peak on the keyword 'democrats' which we can find is related to the closing Democratic caucuses to choose their candidate for the US election at that time. We also learn that 'China' becomes more significant in later June due to the expectation among journalists that the effects of the Chinese trade tariffs may become decisive in the US election as the challenger to Donald Trump is known.

Fig 3. Predicted topics for June 2020.

For the conference demo, we will publish a public instance of the Topics Compass online using an English language user interface and offering analysis of news, social media and TV Websites in English, French, German and Dutch. We will define a number of topics in advance which relate to general topics of interest (like cultural heritage) and specific events (originally, we had the Tokyo Olympics for example, now we are mainly tracking the public discussion regarding COVID-19 / coronavirus). On-site we will guide visitors through the Topics Compass interface and demonstrate to them the exploration of trending topics over both the recent past online content as well as in the prediction mode. If we move to an online session, we will make ourselves available via a videoconferencing platform to explain to visitors the Topics Compass and either show them our use of the interface via screen sharing or guide them to use it by themselves when they screen share with us.

ACKNOWLEDGMENTS

REFERENCES

[1] Lyndon Nixon, Miggi Zwicklbauer, Lizzy Komen and Basil Philipp, "The Trans-Vector Platform for optimised Re-purposing and Re-publication of TV Content". Proceedings of the DataTV workshop, at ACM TVX 2019, Manchester, UK, June 2019.

[2] ReTV Deliverable 4.2 "Trans-Vector Platform, Technology Roadmap and Revised Prototype", available from https://retv-project.eu/deliverables/

[3] Weichselbraun, A., Brasoveanu, A., Kuntschik, P. and Nixon, L. (2019): "Improving Named Entity Linking Corpora Quality". Poster at RANLP 2019, Varna, Bulgaria, September 2019.

[4] ReTV Deliverable 1.2 "Data Ingestion, Analysis and Annotation (v2)", available from https://retv-project.eu/deliverables/

[5] ReTV Deliverable 2.2 "Metrics-based Success Factors and Predictive Analytics", available from https://retv-project.eu/deliverables/