Report on the 1st International Workshop on Recent Trends in News Information Retrieval (NewsIR16)

Miguel Martinez[1]　　Udo Kruschwitz[2]　　Gabriella Kazai[3]

Frank Hopfgartner[4]　　David Corney[1]　　Ricardo Campos[5]

Dyaa Albakour[1]

[1] Signal Media, UK

{*firstname.lastname*}*@signalmedia.co*

[2] University of Essex, UK

udo@essex.ac.uk

[3] Lumi, UK

gabs@lumi.do

[4] University of Glasgow, UK

frank.hopfgartner@glasgow.ac.uk

[5] Polytechnic Institute of Tomar / LIAAD-INESC TEC, Portugal

ricardo.campos@ipt.pt

Abstract

The news industry has gone through seismic shifts in the past decade with digital content and social media completely redefining how people consume news. Readers check for accurate fresh news from multiple sources throughout the day using dedicated apps or social media on their smartphones and tablets. At the same time, news publishers rely more and more on social networks and citizen journalism as a frontline to breaking news. In this new era of fast-flowing instant news delivery and consumption, publishers and aggregators have to overcome a great number of challenges. These include the verification or assessment of a source's reliability; the integration of news with other sources of information; real-time processing of both news content and social streams in multiple languages, in different formats and in high volumes; deduplication; entity detection and disambiguation; automatic summarization; and news recommendation. Although Information Retrieval (IR) applied to news has been a popular research area for decades, fresh approaches are needed due to the changing type and volume of media content available and the way people consume this content. Hence, the first international workshop on recent trends in News Information Retrieval (NewsIR) was held in conjunction with ECIR 2016. As part of the workshop, we released a new dataset consisting of one million news articles to the research community. The workshop was very

well attended with around 70 registered participants. We received a healthy number of 19 submissions in total of which 12 were accepted for presentation. In addition to that, we were pleased to have two keynote talks by well-known experts in the field - on with an industry background (Jochen Leidner) and one from academia (Julio Gonzalo). The workshop also included a breakout session to discuss ideas for a future data challenge in news IR and closed with a focused panel discussion to reflect on the day. Throughout the day the workshop stimulated discussions around new and powerful uses of IR applied to news sources and the intersection of multiple IR tasks to solve real user problems. In particular, several ideas were presented on solving complex information needs for media monitoring, event detection and summarisation. Moreover, and going forward, the workshop concluded with a long list of suggestions for shared tasks, and dataset requirements.

1   Introduction

News from mainstream media outlets is often one of the most relevant, and influential sources of information. This ranges from the influence that newspapers may have on elections to the reputational damage that a negative article in a well-known magazine can cause to a brand. The process of consuming news itself is constantly changing. We receive a continuous influx of news information from different sources (e.g., traditional newspapers, blogs and social media) and this has had a massive impact on the nature of information systems. Some of the current challenges we are facing are the integration of news data with other sources of information such as social media [1]; real-time analytics [2]; processing text in multiple languages; automatic temporal summarization [3]; and scalable processing of millions of articles on a daily basis.

Following discussions at ECIR 2015 we created a forum[1] to discover if there was enough interest within the IR community for a workshop focusing on online media, and news data in particular. Around 40 members joined the forum straightaway and several fruitful discussions started. This was a clear indication for the strong interest in the community for organizing such a workshop. Furthermore, the discussion in the forum illustrated the diversity of topics that this workshop aimed to explore, which include among others: traditional and social media integration, event and anomaly detection, credibility, bias and polarity, temporal summarisation, and data visualization.

Overall, the workshop aimed to bring together scientists conducting relevant research in the field of news and information retrieval to facilitate discussion and debate about the problems they face and the solutions they are exploring. Moroever, to stimulate workshop participation (and more generally to provide a useful resource for researchers in the area), we have released a new dataset of one million recent news articles from a wide range of sources: The Signal Media One-Million News Articles Dataset (Signal-1M)[2] [4]. In contrast to many existing collections (such as Reuters-21578 and Reuters RCV1), the dataset includes news articles from a wide range of sources including global, national and local newspapers, along with magazines and blogs. This dataset is released under the standard Creative Commons licence[3] to encourage re-use in diverse non-commercial research projects. Furthermore, in the call for papers, we encouraged submissions of experimental results on our new dataset.

---

[1]https://groups.google.com/forum/#!forum/news-ir
[2]http://research.signalmedia.co/newsir16/signal-dataset.html
[3]https://creativecommons.org/

We believe that one or more shared tasks or challenges will emerge and that, with suitable refinement, these may form the basis of future workshops.

2 Workshop Programme

Each submitted paper was reviewed by at least three members of the programme committee. This was followed by a discussion period among the workshop organisers as a result of which decisions about acceptance/rejection were made. The Workshop programme included two keynote talks representing both academia and industry, full paper presentations and a poster session, finishing with three breakout groups followed by a panel. The full workshop proceedings were published in the CEUR Workshop Proceedings[4] [5].

The one-day workshop was very well attended with around 70 registered attendees and was structured as four sessions: two in the morning and two in the afternoon. In the first session, the workshop was kicked off with an introduction from Miguel Martinez providing a motivation and a summary of the overall objectives for holding the workshop. The introduction also included a summary of participation figures, which reflect the success of the workshop where a total of 19 submissions were received resulting in 12 accepted papers. Following this, the first keynote talk was presented by an industry representative, Jochen Leidner from Thomson Reuters. After the keynote, authors of full papers presented short talks. These talks were short as the authors had the opportunity to discuss their work in detail with attendees during a poster session in which authors of all papers (long and short) were able to present posters and/or demos. Indeed, the poster session was a success as can be inferred from the crowded corridors in the conference venue and the informal feedback we got from participants. The poster session marked the end of the second session and the first half of the day. The third session started an academic keynote by Julio Gonzalo from UNED. After that, Dyaa Albakour (Signal Media) presented the Signal-1M dataset which was released as part of the workshop to foster participation from the community and to provide a basis for a data challenge track in news IR. This paved the way for the breakout group discussion where each group was asked to discuss challenges and requirements for a modern data challenge track in news IR. A representative from each group presented their findings to the wider audience. Finally, the workshop was closed with a panel discussion that reflected on the day and the various research challenges pointed out during the workshop.

In the following, we detail the workshop's activities outlined above. To give the reader a flavour of the workshop's presentations and discussions, in Figure 1, we show a word cloud generated from the tweets posted on the workshop day mentioning the account @*newsir*16. We discuss these activities under the various themes that were covered by the talks and the keynotes. In particular, to structure the discussion, we identify three main themes which have no clear boundaries and definitely overlap. These themes are *Media Monitoring & Insights*, *News Events* and *Analysis & Visualisation*.

2.1 Theme 1: Media Monitoring and Insights

This theme is a broad one and it refers to a wide range of IR tasks and applications where the aim is to provide professional users with timely access to information of their interest. This ranges from real-time filtering of news articles for a certain brand or certain topics, to

---

[4]http://ceur-ws.org/Vol-1568/

Figure 1: A word cloud generated from the contents of the @*newsir*16 tweets

more complex tasks such as providing market intelligence tools by mining large amounts of news data from main media outlets, blogs and social media.

The keynote by Jochen Leidner [6] presented views of research in this area and the challenges that the IR community faces in such contexts. These views are supported by the various real-world use cases presented in the keynote. He started with the early days innovations in Thomson Reuters where the founders used pigeons to "be quicker than the news" that were being distributed by ship. Going several decades ahead, Thomson Reuters is now investing largely in the domain of media monitoring and insights. An interesting example of this is detecting the side effects of "popular" medicines by exploiting social media (i.e. Twitter). One of the challenges for doing this is the sparsity of the tweets. For a total of 500M tweets that are generated in a given day, 721 of them mention one of the 2,200 more commonly bought drugs. Given these tweets, it is still possible to find correlations and unfold patterns of unknown side effects that should be documented. Also, in some cases, such side effects could potentially be used to cure or address other illnesses. Another example is building a risk profile for a company from public media. This requires a combination of heuristics, machine learning and human intervention to understand situations that will potentially jeopardise some companies. The take-away message from this talk is that the future of media monitoring is transitioning from "delivering news to delivering actionable intelligence", which surfaced the aforementioned tasks.

The second keynote in the afternoon session by Julio Gonzalo [7] also touched on the specific topic of reputation management in social media and the lessons learnt in tackling reputation management within the RepLab Evaluation framework. A typical task for reputation management is sentiment analysis of news articles and social media posts. Julio highlighted the challenges of performing sentiment analysis in this context. This includes differentiating between opinions and polar facts. For instance "Tesco is firing 15,000 people because their shares are dropping" would be a very negative story about Tesco, but an objective one nevertheless (assuming it is a true story). In addition, sentiment analysis for reputation management should take perspective into account. For example, "Antena 3 buys la sexta" is positive news for investors but may be negative for citizens. Finally, Julio

reminded us that sentiment polarity is not always a measure of reputation.

Media monitoring applications have made their way into some of the workshop papers. For example, Igor Brigadir presented a paper on detecting attention dominating moments in online media [8], i.e. moments in which people talk about the same thing with regards to a specific topic. The paper identifies these moments by looking at points of time where the diversity of documents drops significantly. The paper shows examples of applying this approach to news articles and blogs in the Signal-1M data set [4], along with a parallel Twitter stream in September 2015. The examples show how this approach can detect attention moments that are related to real-world events documented in the Wikipedia events portal. While this paper considers diversity to detect attention, the paper by Prati *et al.* [9] looked at a very related concept to diversity which is novelty. In particular, the authors applied a recent model for novelty dynamics to analyse news coverage. In their analysis, their main variable to define the concept of novelty is the lead newsmaker. Their analysis uncovers interesting insights for different media sectors and topics. For example, their analysis shows that newspapers have the largest novelty in introducing new lead newsmakers, while crime is the topic with the most novelty.

Finally, Suzan Verberne presented a user study [10] on traditional media monitoring, where professional users use very long and complex Boolean queries with keywords to specify their information needs. In particular, the user study investigated the usefulness of query suggestions in this context to aid the professional users in refining their queries. High precision is not an issue for these professional users and they are generally satisfied with results without any relevance ranking. However, for generating relevant query suggestions, the relevance ranking becomes important as the quality of the query suggestions is dependent on a good ranking of documents and users would abandon non-relevant query suggestions.

2.2   Theme 2: News Events

Detecting, summarising and filtering events in news has been a popular research topic in IR and it has definitely dominated some of the talks presented and the dicussions in the workshop. To this end, the paper presented by Andrey Kutuzov and Elizaveta Kuzmenko [11] proposes yet another application of neural word embeddings [12] to detect real-world events in the news. Their approach involves learning word embeddings on a large reference corpus and then successively updating the model from daily news using the Signal-1M dataset [4], as well as a Russian news collection. For a given entity, they can estimate the association shift (semantic shift) over time where any significant shift may be triggered by an occurrence of an event. Their analysis on country names shows the promise of the approach, where they were able, for example, to automatically detect a shift of associations for Chile to earthquakes on the day when an earthquake took place there. Similarly, the poster presented by Basile *et al.* [13] proposes an alternative approach for finding temporal shifts in word meanings as results of real-world events. In particular, the authors apply Temporal Random Indexing to build several geometrical spaces of word meanings called Distributional Semantic Models. They also used the Signal-1M dataset [4] to validate their approach. Their approach was able to detect the shift in the contextual use of the word 'scandal' during the third week of September 2015. In particular, and as a result of breaking news on the Volkswagen emission scandal, words like 'volkswagen', 'automaker' and 'diesel' became direct neighbours of the word 'scandal'.

While the previous papers focused on detecting events, Michael Bender from Thomson

Reuters presented work [14] on clustering news articles around the sub-topics of an event. The seed clusters for an event are identified from editorially supplied topic labels called "sluglines". Their approach involves fuzzy deduplication of articles to identify highly similar documents and then clustering them using similarity based on text, entities and topics. Since ground truth data of ideal clusters is hard to obtain, the authors opted for a qualitative evaluation of this approach with human assessments on the cohesiveness and the accuracy of the clusters using a 5-point Likert scale.

Finally, Gregor Leban presented a paper [15] on the Event Registry system[5]. The motivation behind developing this system is trying to move away from traditional news consumption platforms where users are overwhelmed with duplicates of popular and dominating stories. This would also prevent them from discovering the "long tail" which may be of interest. Their approach is based on semantic annotations on the document level, e.g. entity linking, clustering news articles, cross-lingual clustering and event construction (driven by the clustering step). Using these components, they are able to show events rather than articles to provide a more general understanding.

2.3   Theme 3: Analysis and Visualisation

There were several papers and posters that exploited, analysed or visualised news for a variety of tasks. An excellent example on how news can be exploited for other IR tasks was presented by Stefano Mizzaro [16]. Stefano and his co-authors propose to categorise tweets with collection enrichment using new sets of words, extracted from news on webpages of the same temporal context. They tested three different features of news namely volume, variety and freshness. The experiments confirmed the importance of all of these features. The paper was not just well received but also won the *Best Paper Award*. Svitlana Vakulenko presented a demo [17] where tweets are used in the opposite direction. In particular, the authors proposed to build a bipartite graph with two types of nodes, namely tweets and news, and two types of edges: (i) explicit edges from tweets to news articles they cite, and (ii) implicit edges infered from a similarity between news articles using a bag-of-relations representation constructed with NLP techniques (since news sources do not cite each other). In practice these implicit edges serve as network diffusion models to uncover the latent relations between news articles which can be useful to journalists for example to understand the flow of a story.

As discussed in Section 1, the Signal-1M dataset was released to foster research in news IR. Indeed, several papers in the workshop used the dataset to validate and evaluate the proposed approaches (c.f. Section 2.1 and Section 2.2). Dyaa Albakour from Signal Media presented the dataset [4] outlining the motivations for releasing it in the first place. Traditional news datasets are often drawn from one source and have become outdated. This dataset aims to address this gap by covering a large number of sources representing main media outlets, less popular media sources and even blogs. Dyaa also presented some insights from an initial analysis of the dataset. For example, by computing similarity between all possible pairs of articles, we can observe a substantial ratio of duplicates (on average 2.2 duplicates per article). Moreover, an analysis of the daily and hourly distribution of articles uncovers publishing patterns across different media types and in different regions (e.g. peaks during early morning of certain timezones). This was echoed in the presentation of Sergio Nunes [18] who demonstrated the MediaViz Project[6], which provides a visualisation tool to explore the dataset. In

---

[5]http://eventregistry.org
[6]http://irlab.fe.up.pt/p/mediaviz/newsir/

particular, the tool plots graphs of volume and allows the user to explore the diversity over time as estimated by a combined distance metric between all n-gram daily vectors.

2.4   Breakout Groups

For the breakout groups, to initiate the discussion, we raised hypothetical questions to the participants. The questions were:

- If we organise this workshop next year, and we want to introduce a shared task (challenge) to work on, what will the task(s) be?

- What datasets do we need to evaluate it?

Answering these question will help us in understanding the priorities of the community around news IR in addition to building datasets and evaluation frameworks to foster research in this area. Participants formed three groups and each group discussed the questions and reported their findings to the workshop. We summarise their findings in the following:

- Although the Signal-1M dataset [4] was well received by the participants the groups pointed out some of its limitations and suggested some additions. Firstly, the one-month range chosen for it (September 2015) makes it difficult to be used for tasks that require mining temporal patterns over longer time spans, such as monthly or quarterly patterns and therefore it was suggested that a longer time span is required. Other suggestions include the integration with other information sources like multimedia content from articles or more multilingual documents. Another very common request was aligning a Twitter dataset over the same period to have a unified collection. This will be very valuable for a multitude of different tasks, such as reputation management. The Twitter collection presented in the workshop by Brigadir *et al.* [8] is a good candidate. Finally, one group suggested to align the dataset with popular knowledge bases and public ontologies.

- For the shared task, a lot of different ideas were brought forward by the groups. Some of the most common were temporal summarisation, entity linking and disambiguation, diversity and sentiment analysis. In general, there was a tendency that the task should be a complex one, as opposed to adhoc retrieval, and should replicate the jobs of professional users such as journalists. However, one group suggested that in addition to a complex task, and in order to encourage more participation, a future data challenge track should also include a simple one that involves labeling some data. Other emerging problems that were mentioned several times in the breakout discussions include media bias detection, the evolution and the verification of news. The evaluation of the potential tasks surfaced the discussions and were brought forward to the panel.

2.5   Panel

For the panel, we invited our two keynote speakers, Jochen Leidner and Julio Gonzalo to be joined by Gabriella Kazai and Stefano Mizzaro. This panel represents both the industry world, as well as the academic one, with expertise in a multitude of fields in IR and NLP being represented by the experience and knowledge of the panelists. The questions raised to the panelists stemmed from the various talks and the discussions during the day. In particular, the panelists discussed the following topics:

- **Evaluation: Are me measuring the wrong thing?** Using the right evaluation metric is a common challenge in IR tasks. Moving towards more complex tasks (e.g. reputation management) implies that we should focus on user satisfaction when designing the evaluation process. Stefano and Julio also mentioned that IR has a multitude of evaluation metrics already. We probably do not need new ones, but we need to use a multiple of them to provide better explanations. Jochen suggested to involve and invite journalists to the next NewsIR event to help us better understand the user model when coming up with an evaluation metric.

- **Social media, News and Blogs. Are they different at all?** The media interaction landscape has dramatically changed in the last decade or so. Traditional news outlets, blogs, social media and personal journalism are now intrinsically linked and the boundaries between them are becoming thinner as time passes. Therefore, it is challenging to classify some of the internet sources as news or blogs. The panelists agreed that the space is getting more complex with clear dependencies between some of the different media. For example, it is not uncommon that a tweet causes someone to create a short blogpost, that in return causes someone to verify a story and then write a short piece in a local newspaper that will eventually be picked up by a major worldwide publication. This symbiosis should be acknowledged and addressed by the community. On the other side, while some blogs could be seen as an authority in a topic, a large portion of the social media space could be considered of low journalistic quality, with low (or none) verification of news and a high likelihood of spreading rumors. This highlights the importance of automatically distinguishing between pieces of real news, pseudo-news, rumours and clickbait.

- **How important is Trust and how can we help measuring it?** Trust is a critical factor in news that has been the foundation of real journalism since its inception. However, the explosion of new sources of information has increased the difficulty of knowing what sources are legitimate or trustworthy without doing detailed research. A very important question then is what do we mean by 'trustworthy'. We live in a world where given a world event, we can find several points of views, that may contradict each other. What source do we then label as 'trustworthy'? We can indeed compute who are the main influencers in a given space, but we face a more difficult challenge when we try to measure if the same person is also an expert in the field. Alternatively, some of us believe that the first step is to show the differences in opinion or point of view, educating the user and allowing her to see how the sources differ from each other or even summarising each point of view.

3   Conclusions

Overall, the workshop was a great success according to various indicators: the large number of submissions, the quality of the papers, the diversity of topics, the large number of attendees and the high level of interactions in the various discussions throughout the workshop. This emphasises the interest of the community in revisiting news IR. It reassures us that there is a wide scope for future work on improving current models to better serve more complex information needs or indeed improving the evaluation approaches to better reflect the user model. Building on this, we aim to follow it up by another edition in 2017, either in ECIR 2017, in a bigger conference such as SIGIR 2017, or as part of an evaluation framework such as CLEF.

Acknowledgments

References

[1] Gianmarco De Francisci Morales, Aristides Gionis, and Claudio Lucchese. From chatter to headlines: Harnessing the real-time web for personalized news recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 153–162, New York, NY, USA, 2012. ACM.

[2] Michael Mathioudakis and Nick Koudas. Twittermonitor: Trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, SIGMOD '10, pages 1155–1158, New York, NY, USA, 2010. ACM.

[3] Javed A Aslam, Fernando Diaz, Matthew Ekstrand-Abueg, Virgi Pavlu, and Tetsuya Sakai. TREC 2013 Temporal Summarization. In *Proceedings of the Twenty-Second Text REtrieval Conference (TREC)*, 2013.

[4] David Corney, Dyaa Albakour, Miguel Martinez-Alvarez, and Samir Moussa. What do a million news articles look like? In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 42–47, 2016.

[5] Miguel Martinez-Alvarez, Udo Kruschwitz, Gabriella Kazai, Frank Hopfgartner, David Corney, Ricardo Campos, and Dyaa Albakour, editors. *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, volume 1568 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.

[6] Jochen L. Leidner. Recent advances in information access at thomson reuters r&d: News and beyond. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, page 1, 2016.

[7] Julio Gonzalo. Monitoring reputation in the wild online west. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, page 2, 2016.

[8] Igor Brigadir, Derek Greene, and Pádraig Cunningham. Detecting attention dominating moments across media types. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 15–20, 2016.

[9] Ronaldo Cristiano Prati and Walter Teixeira Lima Junior. An analysis of novelty dynamics in news media coverage. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 9–14, 2016.

---

[7]http://research.signalmedia.co/newsir16/organisation.html

[10] Suzan Verberne, Thymen Wabeke, and Rianne Kaptein. Boolean queries for news monitoring: Suggesting new query terms to expert users. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 3–8, 2016.

[11] Andrey Kutuzov and Elizaveta Kuzmenko. Cross-lingual trends detection for named entities in news texts with dynamic neural embedding models. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 27–32, 2016.

[12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[13] Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. Temporal random indexing: a tool for analysing word meaning variations in news. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 39–41, 2016.

[14] Jack G. Conrad and Michael Bender. Semi-supervised events clustering in news retrieval. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 21–26, 2016.

[15] Gregor Leban, Blaz Fortuna, and Marko Grobelnik. Using news articles for real-time cross-lingual event detection and filtering. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 33–38, 2016.

[16] Marco Pavan, Stefano Mizzaro, Matteo Bernardon, and Ivan Scagnetto. Exploiting news to categorize tweets: Quantifying the impact of different news collections. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 54–59, 2016.

[17] Svitlana Vakulenko, Max C. Göbel, Arno Scharl, and Lyndon J. B. Nixon. Visualising the propagation of news on the web. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 60–62, 2016.

[18] Tiago Devezas, José Luís Devezas, and Sérgio Nunes. Exploring a large news collection using visualization tools. In *Proceedings of the First International Workshop on Recent Trends in News Information Retrieval co-located with 38th European Conference on Information Retrieval (ECIR 2016), Padua, Italy, March 20, 2016*, pages 48–53, 2016.