Linking Twitter Archives with Television Archives

Zeynep Pehlivan

Abstract Social media data has already established itself as an important data source for researchers working in a number of different domains. It has also attracted the attention of archiving institutions, many of which have already extended their crawling processes to capture at least some forms of social media data. However, far too little attention has been paid to providing access to this data, which has generally been collected using application programming interfaces (APIs). There is a growing need to contextualize the data gathered from APIs, so that researchers can make informed decisions about how to analyse it and to develop efficient ways of providing access to it. This chapter will discuss one possible means of providing enhanced access: a new interface developed at the Institut national de l'audiovisuel (INA) that links Twitter and television archives to recreate the phenomenon of the "second screen", or more precisely the experience of "social television". The phrase "second screen" describes the increasingly ubiquitous activity of using a second computing device (commonly a mobile phone or tablet) while watching television. If the second device is used to comment on, like or retweet television-related content via social media, this results in the so-called social television. The analysis of this activity, and this data, offers a promising new avenue of research for scholars, especially those based on digital humanities. To the best of our knowledge, the work that will be discussed here is the first attempt at considering how to recreate best the experience of "social television" using archived data.

1 Introduction

Social media has helped to transform people from content readers to content publishers, thereby massively extending the range and number of people whose voices may be captured in web archives. It is an important data source for

Z. Pehlivan (✉)
Institut National de l'Audiovisuel, Bry-sur-Marne, France
e-mail: zpehlivan@ina.fr

D. Gomes et al. (eds.), *The Past Web*,
https://doi.org/10.1007/978-3-030-63291-5_10

researchers in a variety of fields, such as sociology, data science or history. Twitter attracts more attention than any other platform because of its open and public nature, exemplified by its ease of use and ready access to its data. It is also often a place of first response to global news events, a locus of information and misinformation. Consequently, Twitter has become something of a "must archive" source for web archivists. A number of archival institutions have extended their crawling processes to cover social media, and particularly Twitter, but many do not yet make much of this material publicly available.

When the Web is archived, it is generally something approaching a canonical form that is captured, the view that is seen by most users through a PC.[1] Twitter, by contrast, does not really have a canonical form: people can read tweets via an app, through Twitter's own webpages, on another page displaying feeds, etc. This is one of the reasons why application programming interfaces (APIs) are used to archive Twitter data, as explained in this chapter. In contrast to webpage crawlers, the Twitter API also provides a centralized and convenient access point for anyone wishing to collect data. All of this means that when archiving Twitter, the canonical form is considered to be the tweet data itself, stripped of much of its context. For this reason, we need different approaches to re-contextualize Twitter (and indeed other forms of social media) than we do when archiving the Web. There are two main modes of presentation currently being used for archived Twitter data: full-text search and data mining, including aggregations of top hashtags, top emojis, top users or word clouds, etc., and representation of the "second screen", which involves bringing together televisual and Twitter archives to recreate the phenomenon of "social television". It is this latter approach to archiving born-digital data, which is hugely promising from the perspective of digital ethnography, that will be discussed here.

For several years now, the audiovisual media landscape has been expanding beyond a strictly broadcast dimension. The Web has provided new opportunities for airing programmes and for enriching, editorializing or repurposing them. A new ecosystem has thus been framed that creates communities online and enables audience interaction. The affordances of social networks, such as hashtags or mentions, notably increase media exposure and ensure—for better and sometimes for worse—online virality. A "second screen", born from this new ecosystem, involves the use of a computing device (commonly a mobile phone or tablet) to provide an enhanced viewing experience for content on another device, such as a television. The use of a second screen to augment television consumption is now a common practice. Audiences started to interact on social media in connection with programmes or events that mattered to them, and this gave birth to the new phenomenon of "social television". Twitter has played a crucial role in facilitating these interactions. The analysis of "social television", through the combination of social media, web and televisual archives, offers a promising new way for

[1]There is, of course, no single view of a webpage, given the different browsers and devices that are commonly used, but the archived Web is most often presented as if seen on a standard PC.

researchers to explore not just the content of born-digital data but the contexts within which it has been used.

A new interface developed by INA (France's National Audiovisual Institute) offers a different kind of access to online archives through the simultaneous presentation of social media and televisual archives. This interface allows researchers in the social sciences to recreate the "social television" experience with archived data. This chapter also discusses the issues that arrive when television and Twitter archives are brought together and studied in this way. The next section introduces in more detail the concept of the "second screen" and social television. Then, the new interface and dataset are presented, and related issues are discussed. Finally, the conclusion gives a brief summary and defines areas for further work.

2 The "Second Screen"

In recent years, the concept of watching television has changed significantly, following the huge growth in web technologies and the prevalence of mobile devices. Together, these have created the popular phenomenon of the "second screen". Although, for a long time, the television ecosystem consisted of a single box with a screen, which was capable of receiving broadcast signals and turning them into pictures and sounds, the concept of the second screen is not particularly new. In the USA in the 1950s, the CBS children's show "Winky Dink and You" (Prichett and Wyckoff 1953), created by Harry Prichett and Ed Wyckoff, was the first interactive television show with a second screen. The show introduced a kit that included a "magic drawing screen", a large piece of plastic stuck to the television screen via static electricity. Audience members could "help" the characters in the programme by drawing on this screen.

Television consumption has often been accompanied by other activities, but their scope and nature have changed over time with the introduction of new technologies. Although the "second screen" lacks an agreed conceptual definition, it refers to a technical device (mostly Internet-enabled) with a screen that is used simultaneously with a television. This device might be a smartphone, a computer or a laptop, but it does not include another TV screen, calculators, books, radios, etc. There is a clear distinction between the "second screen" and using two screens in parallel (e.g. watching television while shopping online) (Strippel 2017). Several studies have identified that the activities on the second screen should be directly related to the television programme (Filho and Santos 2015; Zúñiga et al. 2015). A good example of a content-dependent second screen is HBO's "Game of Thrones" app (Silva et al. 2015) that gives users details (on a device like a smartphone) related to the character appearing on the first—television—screen.

A recent report from Nielsen (2018) demonstrates how digital devices have affected the ways in which we consume and interact with media today. According to this report, 71% of second-screen users deploy their device to look up something

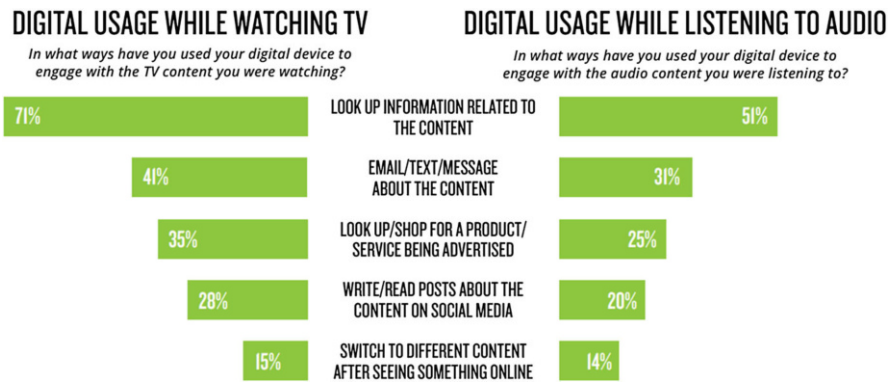

Fig. 1 From Nielsen Total Audience Report: Q2 2018. Copyright 2018 The Nielsen Company (US), LLC. All Rights Reserved

related to the television content they are watching, while 41% said that they have texted, emailed or messaged someone about that content as shown in Fig. 1.

The combination of the second screen and social media has led to a new phenomenon, "social television". The terms "social television" and "second screen" are often used interchangeably, but they represent two different experiences. While some researchers define social television as a special form of second-screen interaction (Wolk 2015), others (Montpetit and Medard 2012; Bellman et al. 2017; Hu et al. 2014) consider it to be more than a second-screen experience, which allows people to communicate and interact with other audiences, content producers, performers, advertisers, etc., around television content. Among all social media platforms, Twitter plays a special role in relation to social television. Twitter serves as a perfect backchannel for television (Bruns and Burgess 2011; D'heer et al. 2015; Harrington et al. 2012). It does not replace the existing media channels but complements them, providing its users with alternative opportunities to contribute more actively to a wider media sphere (Harrington et al. 2012). According to Albarran et al. (2018), there are three main elements that have contributed to the success of Twitter in the sphere of social television: flexibility, which allows the circulation of information in different formats (image, text or video); the hashtag convention and the visibility that this brings even to users with limited numbers of followers (except for private accounts); and the succinct and convenient nature of tweeting, which supports the expression of opinions in a concise, quick and timely manner.

This interaction is not one sided. Viewers may tweet about a programme, but television producers and advertisers can also encourage audiences to interact by sharing hashtags or tweeting during the programme. Some highly interactive shows, for example, news and current affairs programmes, display these tweets live on air. Thus, "Twitter becomes not only a backchannel but part of the show itself" (Harrington et al. 2012). Discussions related to television programmes regularly

feature as "trending topics" on Twitter. Guerrero-Pico (2017) studies how TV fans have made strategic use of Twitter as a tool for activism to launch "Save Our Show" campaigns and highlights a shift towards a collaborative relationship between activist fans and producers in the context of these campaigns. To measure this change in viewing practice, Nielsen and Twitter cooperated to launch the first large-scale survey of social television in 2013. Since January 2016, Twitter television ratings have become "Social Content Ratings", including other platforms like Facebook.

Television sets today are often controlled through an advanced interface connected to the Web. Manufacturers like Samsung, LG and Toshiba have developed their own software to create "smart", web-enabled televisions. Their aim is to draw second-screen activity to the first screen by providing similar options. The most common approach is to have a home screen with a grid of installed apps and widgets for weather, calendars, social media, etc. The social television experience offered here allows users to see their activities from different social media platforms and to interact (e.g. writing a tweet) without needing another device. Most of the work in this field focuses on the development of content-dependent apps. For example, an app called Snapscreen (Lösch et al. 2016) detects both live and recorded television content that is being watched and then offers content-related information and functionality. A tool that recreates the social television experience using archived data offers researchers an invaluable insight into the ways in which social media and television have converged and the behaviour and motivations of the users who engage with the second screen. Capturing Twitter data in isolation allows many forms of analysis, for example, network and linguistic study, but it cannot support different kinds of ethnographic analysis. The experience of the user is missing without the wider context that archives are beginning to provide.

3 A Tool for Researchers

The social television phenomenon has mainly attracted the attention of researchers in the social studies (De Michele et al. 2019; Ceron and Splendore 2019; Delgado et al. 2018; Shokrpour and Darnell 2017; Pfeffel et al. 2016; Gorkovenko and Taylor 2016) but also of those working in different domains, from neuroscience (Seixas et al. 2018) to computer science (Katz et al. 2018). A tool that recreates social television activities using archived data will help social scientists to explore this phenomenon in detail and open up new avenues of research. It allows researchers without any background in programming to handle archived Twitter data, television archives and data statistics in combination, to study the effects of social television (Bruns and Burgess 2011; Katz et al. 2018). In this section, we give a detailed description of the data harvested by INA, present the user interface and explain the issues that have been encountered in developing a new research tool.

3.1 Dataset: Twitter and Television Archives

France's National Audiovisual Institute (INA), founded in 1975, is the world's leading resource for digitized audiovisual content. Its remit is to collect, preserve and share France's audiovisual heritage. From 1995, the programs of the seven national Hertzian channels were preserved as part of the "legal deposit of radio and television", introduced on 20 June 1992, and entrusted to INA. The programmes collected under this remit have gradually been extended to include cable and satellite channels, the free service known as TNT and regional channels. At the time of writing, 103 television channels and 66 radio channels are recorded every day. INA's audiovisual collections include more than 40 million records.

Legal deposit in France was extended to cover the public Web in 2006, and the DLWeb (legal deposit web) team at INA is responsible for the collection and preservation of French websites related to media and audiovisual content. Tweets have also been archived under legal deposit since 2014. Using Twitter, APIs more than 13,000 user timelines and 600 hashtags are archived every day. This collection exceeded one million tweets in June 2019, and an average of 300,000 new tweets are added every day. This is the data with which INA has been working, so that it can be made available to researchers in new configurations.

3.2 User Interface

One such configuration is offered by the social television interface that has been developed. As shown in Fig. 2, the interface is split into four areas. First, there is a search facility that allows users to search for a television or radio programme from the archives by selecting a channel, date, hour and duration. It also contains a full-text search for tweets archived during the programme. Search keywords can be hashtags, mentions related to the programme or a complex query based on metadata (e.g. hashtags: "cashinvestigation" user_mentions.screen_name: "EliseLucet"). Second, there is a programme area that displays the chosen multimedia in a player with "classic" controls (backwards, forwards, volume, etc.). Third, a timeline area shows the number of tweets per minute during a selected programme. It also displays a red vertical line that is synchronized with the programme area and can be used to rewind or forward-wind the programme. As shown in Fig. 3, different keywords can be compared, and significant moments in the show, as reflected in social media activity, can be detected easily. Finally, a tweets area has two different tabs. In the first, activated by default, there is a list of tweets generated by a query formulated in the search area. Each tweet is displayed in a Twitter-like layout, with date, user's screen name and user's profile image. Retweets and like counts are not displayed as only the moment of capture is represented. Users can read all of the tweets using the scroll bar on the left and click on a tweet to move the player to the moment that it was published during the programme. This scroll bar is only visible

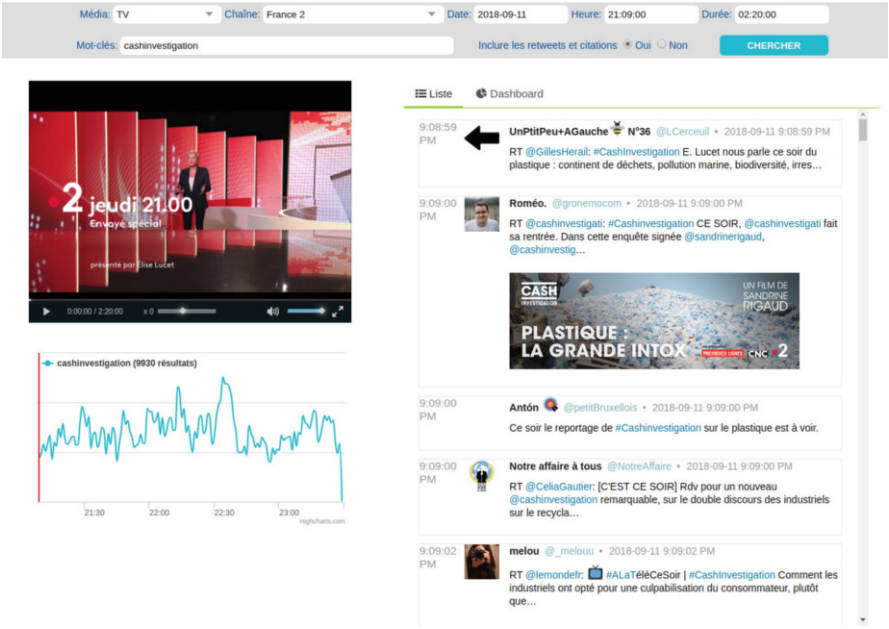

Fig. 2 Overall view of the user interface

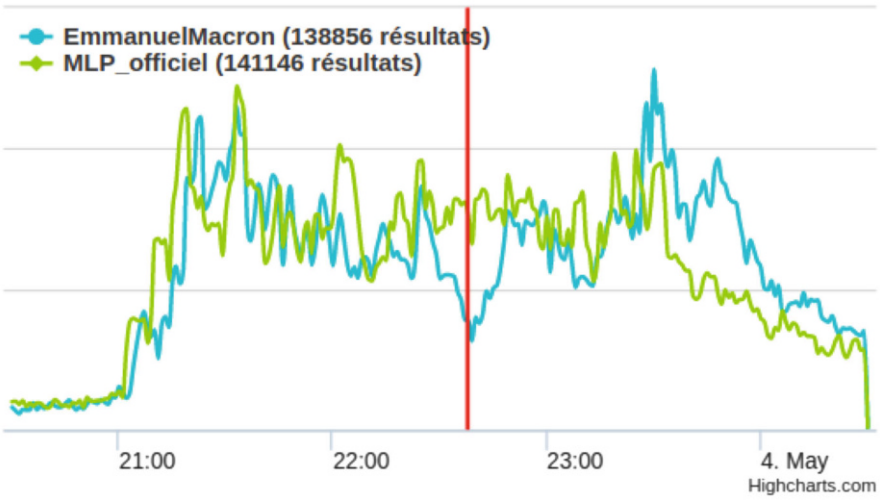

Fig. 3 Timeline for the 3 May 2017 debate for the presidential elections in France showing results for the query "EmmanuelMacron MLP_officiel" that contains official Twitter account screen names for Emmanuel Macron and Marine Le Pen

when the player is turned off. When the player is turned on, the scroll bar disappears and synchronized streaming begins. This gives researchers the opportunity to study social television behaviour in real time using archived data. The second tab, the dashboard, contains statistics for the listed tweets. Different categories are applied to cover different user needs, such as the distribution of tweets/retweets/quotes, languages, geo-localization and top entities (e.g. hashtags, mentions, users, etc.), as shown in Fig. 4. Any other type of aggregation based on Twitter metadata can easily be added to this dashboard. A demonstration video of the tool, which offers a uniquely rich user experience for researchers, can be found at https://youtu.be/XMQF9rQUIyc.

3.3 Challenges

The tool developed by INA allows researchers in the social sciences to analyse the phenomenon of social television, as it plays out on Twitter. It marks the first attempt to visualize audiovisual and Twitter archives in combination, and like any prototype, it has limitations.

The most significant of these lies in the quantity of tweets related to a particular television programme. A regular television show in France, which does not create a particular buzz, generates around 70 tweets (including retweets and quotes) per minute, that is, 1.16 tweets per second. In a case like this, it is easy to display and follow activity and information flows on the player. However, special events both generate more second-screen activity than this and attract more attention from researchers subsequently. This is true of political elections, for example. The debate held on 3 May 2017 during the French presidential elections, between Emmanuel Macron and Marine Le Pen, lasted for 3.5 h and was watched by nearly 16.5 million people. The debate generated more than 3.1 million tweets, that is, 246 tweets per second on average. In this case, it is impossible to generate the synchronization between player and tweets in a human-readable way. As a workaround for this problem, a maximum of 10 tweets per second is displayed in play mode during the synchronization. Users who would like to study such important moments in detail have to pause the player in order to see all published tweets. These challenges of scale and velocity are common to all archives that are concerned with capturing and archiving the Web and social media, and no doubt new approaches and solutions will be found in the future.

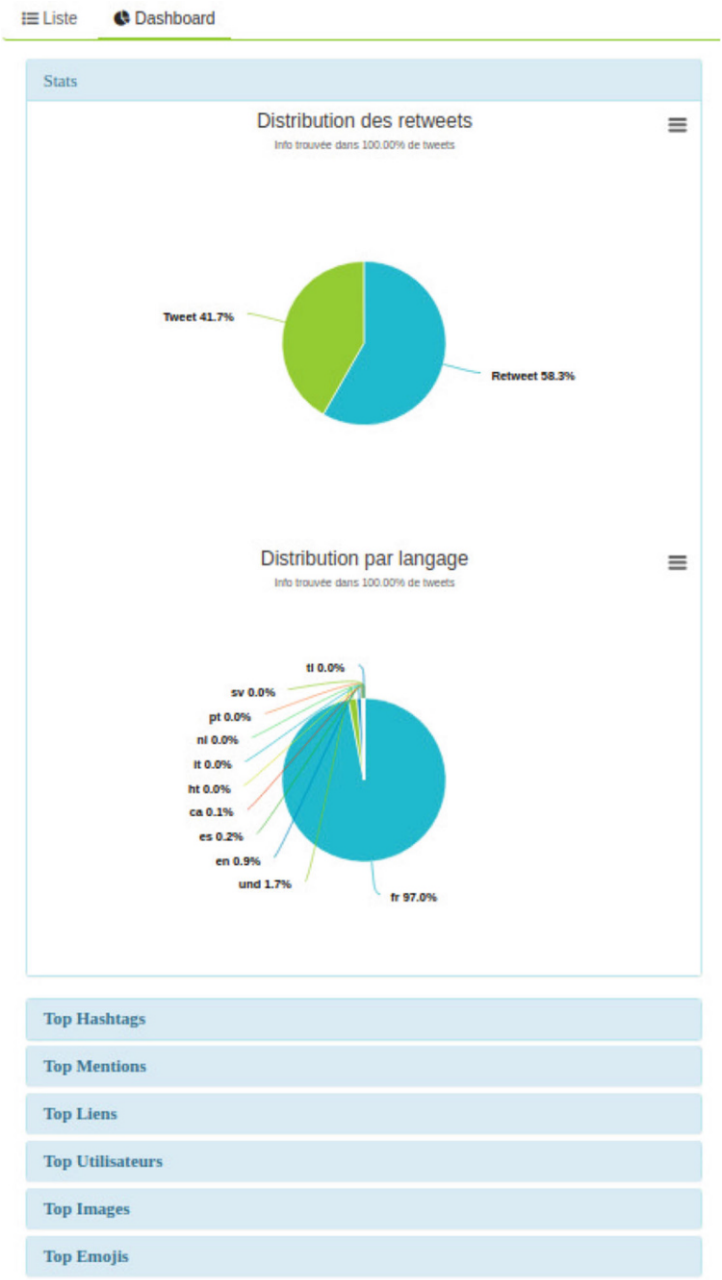

Fig. 4 Dashboard in Tweets Area with different aggregations

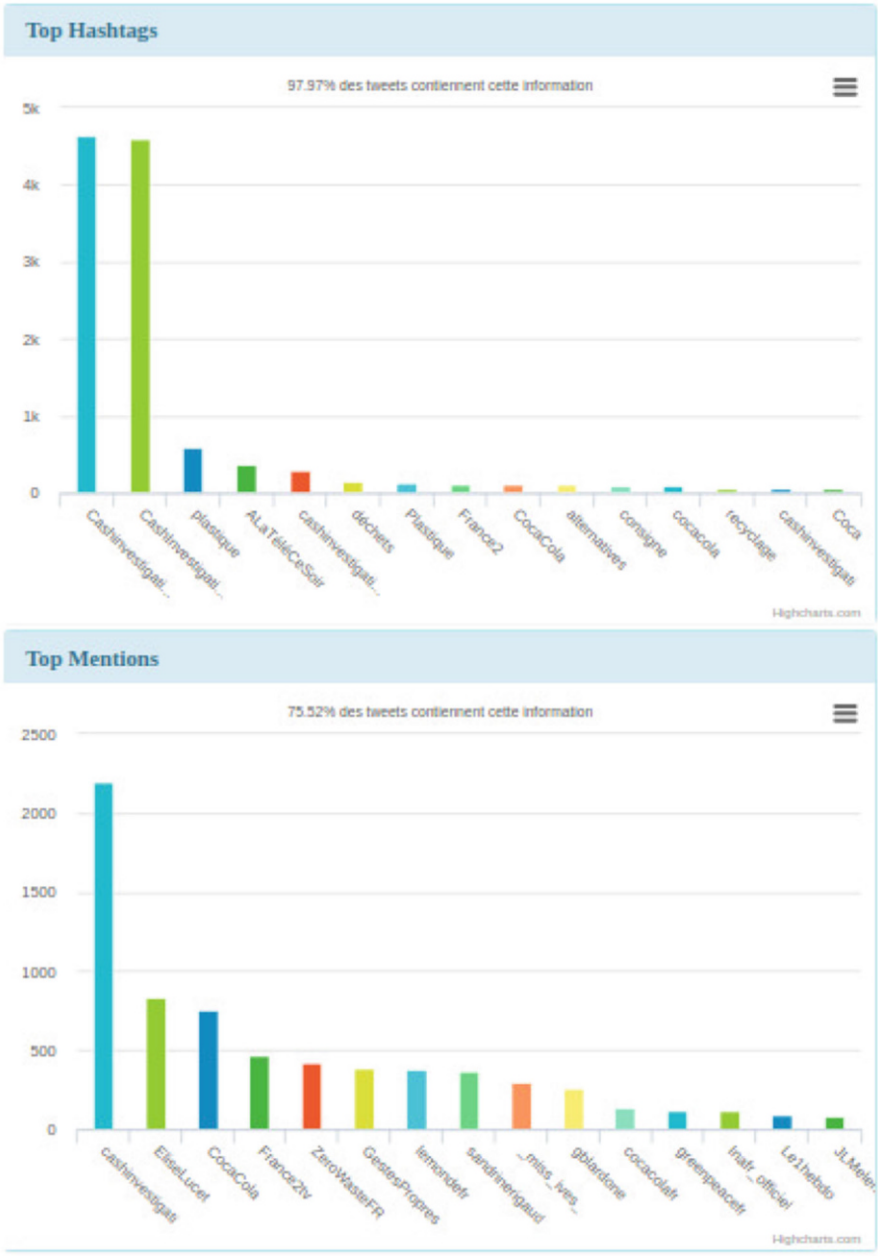

Fig. 4 (continued)

Fig. 4 (continued)

4 Conclusions

This chapter has explored one approach for bringing together televisual and Twitter archives in order to create a new method for analysing archived born-digital data more effectively. The tool developed by INA allows researchers to search multimedia archives simultaneously, to visualize them in a synchronized way and to gain access to contextual statistics like top hashtags, etc. At INA, Twitter archives are created by using Twitter APIs to track accounts and hashtags selected by curators. If a hashtag emerges during a television show, it will only be archived if it is used in association with a tracked account or hashtag. Future work will concentrate on how to visualize this gap, so that researchers know how much data they may be missing; on automatically suggesting hashtags or account names related to particular television shows; and on clustering different keywords, as proposed by Stilo and Velardi (2014), and suggesting new ones to users. For the moment, like many other archives, the focus is on Twitter, but in the future, it is hoped that other social media platforms like Facebook may be added to the dataset. The Web, social media and other forms of media are inextricably linked through the phenomenon of the "second screen", and archives are rising to the challenge of capturing this profoundly networked activity.

Acknowledgements Thanks are due to Thomas Drugeon and Jérôme Thièvre, from the legal deposit team (dlweb) at INA, for their fruitful discussions and inspiration. I am also thankful to Jane Winters for her corrections that greatly improved this chapter.

References

Albarran A, Mierzejewska B, Jung J (2018) Handbook of media management and economics. Routledge, Abingdon

Bellman S, Robinson JA, Wooley B, Varan D (2017) The effects of social TV on television advertising effectiveness. J Mark Commun 23(1):73–91

Bruns A, Burgess JE (2011) #Ausvotes: how Twitter covered the 2010 Australian federal election. Commun Polit Cult 44:37–56

Ceron A, Splendore S (2019) 'Cheap Talk'? Second screening and the irrelevance of TV political debates. Journalism 20(8):1108–1123

De Michele R, Ferretti S, Furini M (2019) On helping broadcasters to promote TV-shows through hashtags. Multimed Tools Appl 78(3):3279–3296

Delgado M, Navarro C, Garcia-Muñoz N, LLuís P, Paz E (2018) TV news and social audience in Europe (EU5): on-screen and twitter strategies. Observatorio (OBS*) 12(4):107–124

D'heer E, Godin F, Evens T, De Neve W, Verdegem P, Van de Walle R (2015) How can Twitter data and viewing rates enrich one another? A conceptual and empirical comparison between viewing rates and Twitter data for TV audience research. In: ICA, proceedings

Filho FC, Santos LA (2015) Second screen and information: history, definition and clues for the future. Comunicação e Sociedade 28:211–227

Gorkovenko K, Taylor N (2016) Politics at home: second screen behaviours and motivations during TV debates. In: Proceedings of the 9th Nordic conference on human-computer interaction, NordiCHI '16. ACM, New York, pp 22:1–22:10

Guerrero-Pico M (2017) #Fringe, audiences and fan labor: twitter activism to save a TV show from cancellation. Int J Commun 11:22

Harrington S, Highfield T, Bruns A (2012) More than a backchannel: twitter and television. In: Noguera JM (ed) Audience interactivity and participation, COST action ISO906 transforming audiences. Transforming Societies, Belgium, pp 13–17

Hu H, Wen Y, Luan H, Chua T, Li X (2014) Toward multiscreen social TV with geolocation-aware social sense. IEEE MultiMedia 21(3):10–19

Katz G, Heap B, Wobcke W, Bain M, Kannangara S (2018) Analysing TV audience engagement via twitter: incremental segment-level opinion mining of second screen tweets. In: PRICAI 2018: trends in artificial intelligence. Lecture notes in computer science. Springer International Publishing, New York, pp 300–308

Lösch S, Willomitzer T, Anderst-Kotsis G (2016) Snapscreen: linking traditional TV and the Internet. In: Proceedings of the 14th international conference on advances in mobile computing and multi media, MoMM '16. ACM, New York, pp 244–249

Montpetit M, Medard M (2012) Social television: enabling technologies and architectures. Proc. IEEE 100(Special Centennial Issue):1395–1399

Nielsen (2018) Juggling act: audiences have more media at their disposal and are using them simultaneously. https://www.nielsen.com/us/en/insights/article/2018/juggling-act-audiences-have-more-media-at-their-disposal-and-are-using-them-simultaneously/

Pfeffel F, Kexel P, Kexel CA, Ratz M (2016) Second screen: user behaviour of spectators while watching football. Athens J Sports 3(2):119–128

Prichett H, Wyckoff E (1953) Winky Dink and You

Seixas SA, Nield GE, Pynta P, Silberstein RB (2018) The neuroscience of social television. Applications of neuroscience: breakthroughs in research and practice, pp 413–426

Shokrpour A, Darnell MJ (2017) How people multitask while watching TV. In: Proceedings of the 2017 ACM international conference on interactive experiences for TV and online video, TVX '17. ACM, New York, pp 11–19

Silva P, Amer Y, Tsikerdanos W, Shedd J, Restrepo I, Murray J (2015) A game of thrones companion: orienting viewers to complex storyworlds via synchronized visualizations. In: Proceedings of the ACM international conference on interactive experiences for TV and online video, TVX '15. ACM, New York, pp 167–172

Stilo G, Velardi P (2014) Temporal semantics: time-varying hashtag sense clustering. In: Knowledge engineering and knowledge management. Lecture Notes in Computer Science. Springer International Publishing, New York, pp 563–578

Strippel C (2017) Praktiken der Second-Screen-Nutzung. Konzeptioneller Rahmen für die Analyse der Parallelnutzung von zwei Bildschirmen. In: Göttlich U, Heinz L, Herbers MR (eds) Ko-Orientierung in der Medienrezeption: Praktiken der Second Screen-Nutzung. Springer Fachmedien Wiesbaden, Wiesbaden, pp 107–136

Wolk A (2015) Over the top: how the internet is (slowly but surely) changing the television industry. CreateSpace Independent Publishing Platform, North Charleston, SC

Zúñiga HGd, Garcia-Perdomo V, McGregor SC (2015) What is second screening? Exploring motivations of second screen use and its effect on online political participation. J Commun 65(5):793–815