A social innovation process for the creation and development of *ad hoc* content for online newspapers and web portals

No Author Given

No Institute Given

Abstract. This paper introduces the description of an under-construction *pipeline*, named *YMemory*, which, starting from recent *Google trends*, allows a decision-maker to monitor the *sentiment* of microblogging platforms regarding these trends, and allows the choice of geographic areas, even narrow ones, and their following auditing. In addition to the positive/negative sentiment about *Google trends*, *YMemory* offers the possibility to display, on the same dashboard, the emotions that *Google trends* arouses in the observed population. As a whole, the pipeline allows to monitor social networks and local sites, in real time, starting from the sensations that emerge from *Google trends*. The pipeline proposed has no expectation of *forecasting* the trends traced. Instead, it aims to provide the user with a guide that, starting from *Google trends*, can support decision-makers (e.g., journalists, analysts, social scientists, doctors and so on) in the creation and development of *ad hoc* content that can satisfy specific geographical reference communities, with respect to different real-world phenomena (e.g. epidemiology, economics, social sciences, and so on).

Keywords: Digital transformation · Social innovation · Inclusive citizenship · Local communities

1   Introduction

Search queries represent what users, from different segments of the population, are interested in at the moment. They capture, for instance, the concerns of populations, showing a fluctuating degree of accuracy on several issues [24, 22, 2]. Although representative of the general population [28], even of a specific location, *search queries*, however, seem to be able to estimate only specific aspects of public opinion, leaving out other relevant cognitive and social aspects that seem to be better captured by social media, such as Facebook or Twitter, which are increasingly used as indicators of *public opinion* and *real-world phenomena* [9, 27]. What seems to be needed is an experimental framework which is capable of giving decision-makers the ability to test *models* and *hypotheses* in a systematic way, integrating both *search query data* [21, 13, 16] and *social data* [18, 7].

   YMemory allows the integration of Google *search query data* and the *social data* of micro blogging platforms, such as Twitter, or local websites. The trends

detected through this *Google service*, demonstrates how the search volume of *keywords* is strongly correlated with as many *regularities*, showing how these types of correlation apply to many phenomena relating to the behavior of citizens of local communities [8, 17].

As new trends emerge in *Google searches*, *YMemory* conducts its queries to trace, even geographically, the feelings and emotions of users of social networks and sites about new trends identified. The core of the pipeline is represented by a sentiment analysis framework that uses an artificial intelligence technique [19, 1], which employs *natural language processing* and *machine learning* modules, to assign emotions and sentiment orientations to a collection of geo-localized posts on micro blogging platforms [11, 26].

2   Literature Review

The availability of the aggregated frequency of *search queries*, such as those provided by *Google Trends*, demonstrated several examples of how the search volume for keywords coincides with as many patterns, showing how these kinds of correlation hold for many local phenomena. For instance [29] shown how the number of searches for the keyword "summer camp" increases when the end of the school year is approaching, because of parents summer recreation plans for their kids. Political decision-makers who, for example, are editors of a news site, may that benefit by looking at the *hot queries* and, consequently, can publish articles on their site that make targeted use of these keywords editing an article on the site focused on such keywords so that Google can easily index the article for a better finding of the same. The practice is well known by SEO practitioners [20, 23].

Jeremy Ginsberg and his colleagues, in their seminal work, highlighted the correlation among *search query data* and one of the major public health concerns, that is the epidemics of seasonal influenza [9]. The methodology proposed selects influenza-like illness search queries such as *influenza complication*, *general influenza symptoms*, *symptoms of an influenza complication*, *antibiotic medication*, *general influenza remedies*, and so forth. Ginsberg's model demonstrated that *Google search query data* can be used to accurately estimate influenza-like illness percentages in health regions for which actual data by traditional surveillance sources were available. The speed with which such data can be modeled (i.e., one or two weeks before), as well as the models' accuracy, make these methodologies exploitable as early warning mechanisms for epidemiological purposes. For instance, such a kind of technique, may be useful for alerting about sharp increase of physician visits, so to induce public health decision-makers to identify the etiology of the outbreak, and providing, at the same time, where available, extra helps or raising local media awareness as necessary.

Another interesting application that clearly shows the (cor)relation among search query data and real-world phenomena was inquired by [26] that highlighted how housing market search data could be used to identify trends home *foreclosures*. The search index increases into the Fall of 2005, during the first

outbreak of actual foreclosures. Search query data and actual foreclosures show the same shape also in Fall of 2007 and with the advent 2008. Probably, when bubble burst, people exploited Google to look for information, legislation and, in general, remedies about foreclosure.

3   Methodology

The beating heart of *YMemory* is the keyword search analysis engine. So we describe some details of the procedure and the main techniques for its implementation. We then refer to the following sections to show how this part of the system practically works when it is in action. The technique employed for the analysis and the exploitation of *search query data* is based on *keyphrase extraction*, a well-known methodology explored by [5]. *Keyphrases*, or *keywords*, are linguistic units, usually, corresponding to one or more words but shorter than a full sentence. *Keywords/keyphrases* identity changes on the basis of the methodology exploited for their identification: statistical motivated *keyphrases* are sequences of words, while *keywords* defined in according to a grammar are known as linguistically motivated ones.

In order to explore, and/or select, the trending stories belonging to a call to Google Trends, in order to explore, and/or select, the trending stories belonging to topo stories of a given a geographical area (i.e. one of the 14 *Y communities* around Italy). Once the trending queries of interests are chosen, the following module inquiries localized tweets and website in order to automatically identify the *mood* of people belonging to the community. This second module applies the same Keyword extraction module in order to identify the emerging sentiment about the given search query. For example, the local website can express the mood of an *influencer* or *opinion leader* who is expressing themselves on a fact of the day regarding a certain search query topic. The problem is that they express a single opinion, albeit supported by some form of authority or reputation. While, the use of the analysis of a significant sample of tweets, that are expressed by a multitude of people, allows us to obtain a *collective mood* on the story. And it is the final product of this double analysis that allows us to build the contents as specified in the following sections. In any case, for the sake of completeness, we try to provide other fundamental elements of the keyword extraction procedure which also contemplates a way to identify the sentiment of the community on each topic identified through the search query data.

The *keyphrase extraction* procedure is based on a supervised learning approach that applies linguistic processing on input text (i.e., tweets and/or web posts). The system uses a Support Vector Machine (SVM) as the learning method and *TFxIDF term weighting* with the *position* of a phrase as features. The methodology employed chooses the candidate phrases using linguistic knowledge. In other words, candidate phrases generated by the system are sequences of *Part of Speech* (PoS) containing *Multiword Expressions* (ME) and *Named Entities* (NE). Extraction is driven by a set of *patterns* which are stored in a

pattern database; once there, the main work is done by the learner device (i.e., the SVM).

Usually, in order to identify the *mood* of a community of tweets and/or, in general, posts, they are exploited a combination of sentiment and emotion analysis modules [4]. The former is used to identify the predominant orientation, i.e. positive or negative of a tweet or a post, while the latter captures the emotions expressed by the current tweet, returning one of the following feelings: anger, disgust, fear, joy, sadness, and surprise. Unlike this canonical procedure, described, for example, in [3], the keyphrase extraction system briefly described above, allows a finer-grained identification of the mood, as that represented, for instance, in Fig. 10 and in Fig. 15. As can be seen from the figures cited, the keywords found allow us to identify not only the classic emotional states but a whole series of categories considered relevant by local communities, with respect to various topics. For example, Figure 10 shows how people are alert to *vaccination*, *fever*, *symptoms*, *treatment*, etc. All these states are not covered by the basic emotions described in the above scientific literature.

The keyphrase extraction procedure reported in this section has been tested in different fields of evaluation, with the aim of establishing the goodness and effectiveness of the extracted keword in different fields of application. For example, one of the versions of the algorithm was evaluated as part of the DUC campaigns[1], organized by DARPA and NIST. In order not to overwhelm the presentation, we refer to [5] for a more complete analysis.

4   *YMemory* - What it is

YMemory engine is composed of a monitoring process carried out by the *keyword search analysis* and *sentiment analysis* module, two components that use algorithms that make extensive use of artificial intelligence. The pipeline is divided into three types of services, further divided into sections, with the aim of creating as many features.

The three services are included within the application through simple and intuitive functions that *YMemory* makes accessible on different devices (e.g., smartphone, tablet or PC). *Inter alia*, the three types of services that make up *YMemory* are:

- *Memoria di Pico* (distance learning service): section dedicated to distance learning;
- *YMood* (citizen sentiment and satisfaction service): section dedicated to the perception of citizens' feelings and the evaluation of their degree of satisfaction with the services and needs of the moment;
- *YTalk* (proposal section): section dedicated to citizens' proposals.

The three services introduced are described in detail in the following sections.

---

[1] DUC evaluation campaign publications can be found here: https://www-nlpir.nist.gov/projects/duc/intro.html - Refer to DUC 2004, DUC 2005, DUC 2006 and DUC 2007 to explore publications about the algorithm reported in this paper.

4.1   *Memoria di Pico*

Memoria di Pico is the distance learning component of *YMemory*. It is based on the indispensable collaboration of physicians, scientists, academics and professionals, at the service of businesses, schools and families, with the dual purpose of *training* and *informing* on all topics perceived as fundamental during and after the health emergency from covid-19.

This functionality of *YMemory* stems from the ongoing experience of *Memoria di Pico*, a distance learning format created by *YPeople*, based on the idea of verticalization of information through a monitoring system. The name recalls Pico della Mirandola's illustrious memory. The *training* and *information* contents are published in video mode, of short duration, in order to be accessible by everyone, without distinction, overcoming any barriers, such as in the case of blind or deaf citizens.

The contents of *Memoria di Pico* module are designed on the basis of the *trend monitoring* and *citizenship sentiment* modules introduced above. In other words, as soon as an analytically verified "peak" of terms emerges, which, for example, speaks "covid-19 symptoms", then it means that a perception of contagion or a feeling of fear is developing in a part of the population in a certain area. Then, one or more videos will be immediately proposed in the Memoria di Pico section, through which the citizen can keep himself informed, in a trusted way, with an expert or a professional in the sector. This applies to all the other sectors present in *Memoria di Pico*, including the cultural ones. If the other monitoring modules still return a feeling of fear, then an open webinar (i.e., *YTalk*) is launched, which will allow to analyze the problem directly in an open forum and with one or more experts in the field.

Within the *Memoria di Pico* section are hosted training and news contents aimed at companies, families, schools and so forth, even as short tutorials. The topics are divided into sectors and each citizen will have access to any category. The categories and their targets are reported below.

News - This category dedicated to all citizenship. It uses information content, insights and discussions with experts and representatives of public institutions on current issues, with reference to the community of reference, during the covid-19 emergency.

Culture - This is the category dedicated mainly to schools and students. It uses training contents and webinars created by university teachers through which students, or anyone interested, can improve their knowledge through distance learning during the covid-19 emergency.

Economy - The category is dedicated mainly to businesses. It uses information, training, webinars and tutorials made by professionals, through which companies, or anyone interested, can obtain advice and insights on their work during the covid-19 emergency.

Health - This category dedicated to all citizens. It uses information, training, webinars and tutorials made by physicians, through which users can get answers and advice for prevention during the covid-19 emergency.

Through the categories introduced above, users will have the opportunity to find, just a click away, all the information they need to deal with the emergency.

4.2   *YMood*

YMood is the component of *YMemory* dedicated to measuring citizens' perception and sentiment, by monitoring local networks, and citizens' satisfaction with the contents offered, through a *voting system*.

Citizens can evaluate the contents of the *Memoria di Pico* section, so as to measure their degree of satisfaction based on the topics covered. The degree of citizens' satisfaction are measured through a graduated scale, simple and intuitive, which makes use of a simple click. Thanks to this module, *YMemory* will have immediate feedback on user satisfaction on a specific service and/or content. The degree of citizens' satisfaction with services represents one of the main purposes of every institution, public and/or private, that has to do with a specific market, according to the principles of *corporate social responsibility* and the criteria of *value co-creation*.

Thanks to *YMood*, *YMemory* represents a way, through which the different *Y communities*, that counts about 150,000 followers around Italy, can find a mean of social and economic comparison, to be able to cooperate with a view to managing the uncertainties and risks present during the health emergency from covid-19, thus allowing the protection of minimum well-being for an *inclusive citizenship* [15].

4.3   *YTalk*

YTalk is the component of *YMemory* designed to play an open and always active webinar role. *YTalk*, that was born during the covid-19 emergency, takes its name from the virtual discussion space of *YPeople.it* and *YSport.eu*, the two newspapers of *Y Media Analytica*.

The *YTalk* format was designed with the aim of allowing the educational and scientific activities to be carried out outside the academic context, as part of a joint initiative of a group of scholars across two Italian universities, the *University of Trento* and the *University of Salerno*. *YTalk* initiative was born, first of all, as an immediate interaction between some teachers and some students, as well as between the readers of *YPeople.it* and *YSport.eu*, to respond to the needs of people to feel involved as active subjects of the health emergency and social life that the country is experiencing during covid-19 lockdown. *YTalk*, from time to time, involves experts and citizens who want to offer insights and questions regarding their daily life during the covid-19 emergency. The readers of *YPeople.it* and *YSport.eu*, thanks to these interactions, can put themselves at the service of their community, offering their contribution by first investigating their respective territories, through the *YLab mode*, and then analyzing solutions,

even only in terms of information and communication, that universities, research centers, and institutions can employ to cope with the difficulties of the moment.

YMemory also works as a *citizen journalism* editorial staff and, as such, it allows to identify/report information and submit it to the competent collaborators (i.e., *YTeam*) of *Y Media Analytica* and its newspapers *YPeople.it* and *YSport.eu*. The *YTeam* then takes care of the elaboration and diffusion of the same, through the application after appropriate *analytical verification.*

The pipeline proposed is part of the territorial platforms of *active citizenship* [10] because it identifies itself as a tool for responsible use of the territory. *YMemory* is also a tool capable of experiencing change as it represents, for the citizens of the *Y communities*, the ability to implement and use innovation during the Covid-19 emergency.

Moreover, the process proposed represents a way through which the different parts of a community can find a mean of social and economic comparison to cooperate in order to manage the uncertainties and risks present during the emergency from covid-19, in order to protect a well-being minimum for an *inclusive citizenship* [15, 12].

Within a special section of *YMemory*, all citizens can propose the improvement of a service through a system of sharing and discussion of their idea. The section also provides a comment platform dedicated to the discussion where users can talk, request insights, cultivate common interests, deal with current information and topics.

5   *YMemory* - How it works

Fig. 1 reports an overall view of the *pipeline* described in the previous section. At the top of the pipeline are the three modules that work in the back office on the server side. The first is the *trend monitoring module*, which mainly deals with *keyword search analysis*. The second module is the one that deals with *social monitoring* [14, 6]. Finally, the last module of this part of the process is the one that deals with the monitoring of local sites close to the *Y Community* (e.g., Verona, Parma, Naples and so on for the other communities).

Once the monitoring of the three modules is completed, *YMemory* proceeds to the creation of the contents and the consequent passage to *Memoria di Pico* and *YTalk*, as described in the previous section.

5.1   *Case study 1 - Economy*

Fig. 2 shows trends monitoring about the search term "Eurobond", peaked from March 15th to April 26th. At this point, *YMemory* conducts its queries to trace the feelings and emotions of users about new trends.

During the above period, the search query "Eurobond" presented a peak in all Italian regions, with particular reference to the communities of Rome and Milan, as shown in Fig. 3 below.
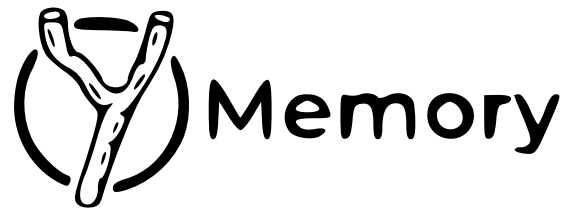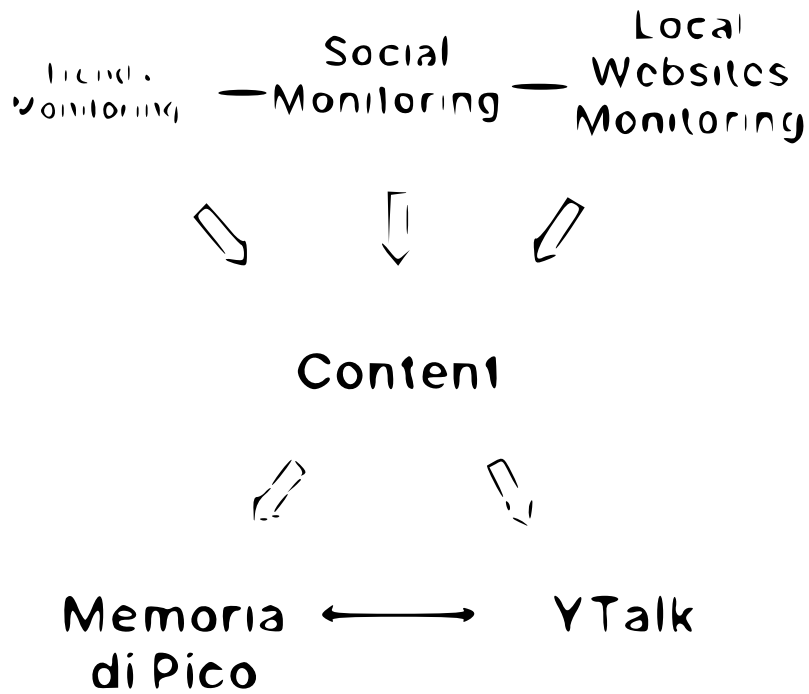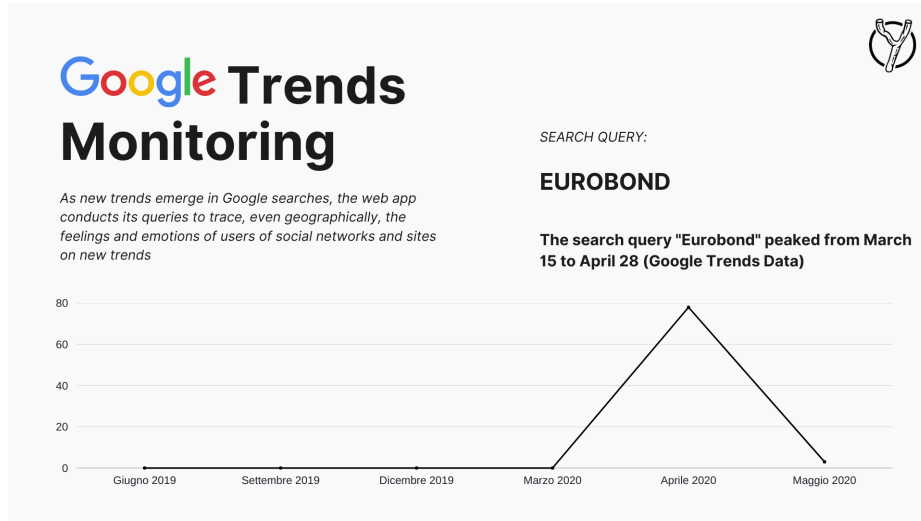
Fig. 1. YMemory architecture
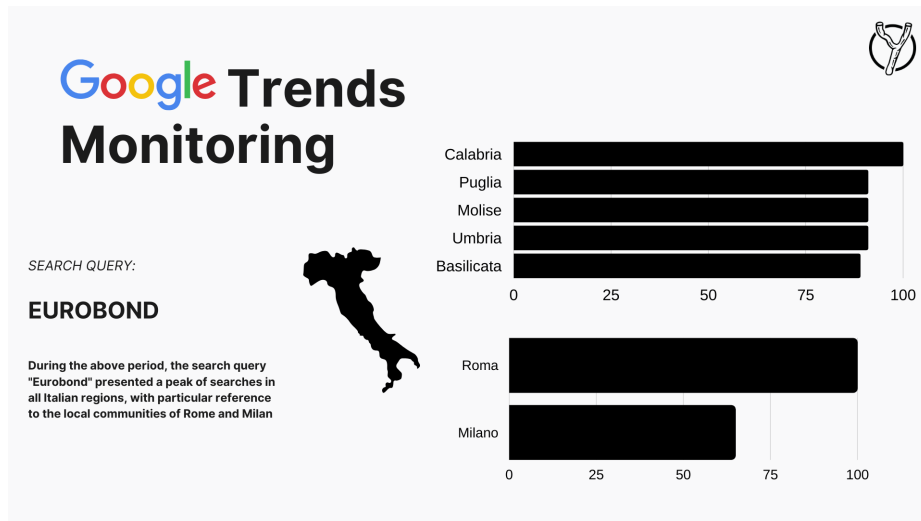
Fig. 2. Trend monitoring for *Eurobond*

Fig. 3. Search query *Eurobond* per regiorn

During the same period, the term "Eurobond" presented a set of associated queries, Fig. 4, such as "mes", "mes Eurobond", "eurobond cosa sono", "Eurobond significato". After that, the social sentiment module builds a wordcloud through a survey done on the social media communities of *YPeople* and *YSport.*
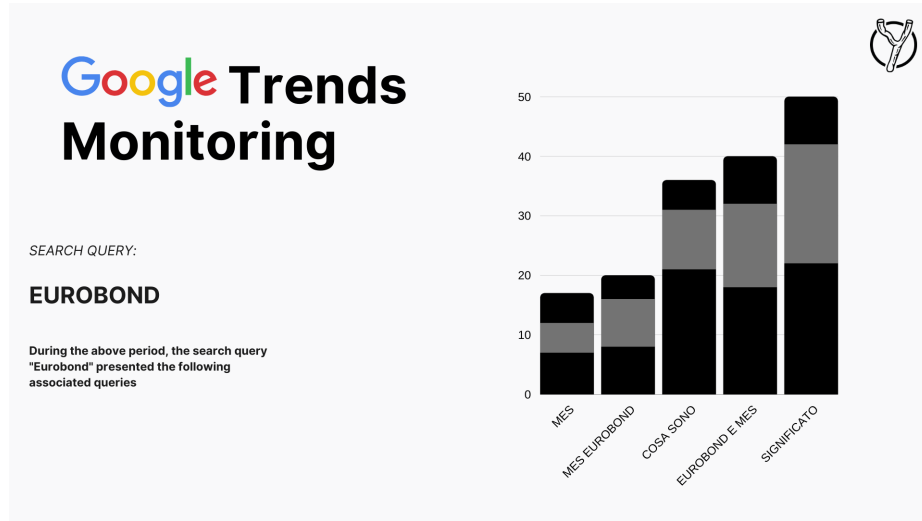

Fig. 4. Associated search query to *Eurobond*

The *sentiment module*, represented in the figure below, shows the main characteristics associated with "Eurobond" search query and that represent the ground of the content that will be delivered through *Memoria di Pico* and *YTalk* modules.

Then comes another turn, where you have to look around, within the *Y community*. This work is carried out by the *local website monitoring module* which constantly monitors websites and all the newspapers that have their own websites (see Fig. 5). This module also performs a strategic function within the process as it allows the creation of content that can give added value compared to what has already been published on the subject by other newspapers or information websites. For instance, given a topic from the *monitoring module* that has not been published yet, then *local website monitoring module* allows to fill the information gap within the *Y community* of reference.

At this point it is the moment of the creation of the actual content that takes place in two ways. The first is represented by a video pill of about 1 minute inside the *Memoria di Pico* module. If this content arouses interest and engagement within communities then a *YTalk* is organized immediately with experts in the sector. For example, with regard to the search query term "Eurobond", and the associated queries, the short video was entrusted to a well-known Italian banker who treated exhaustively, allowed by the short duration, the contents on the
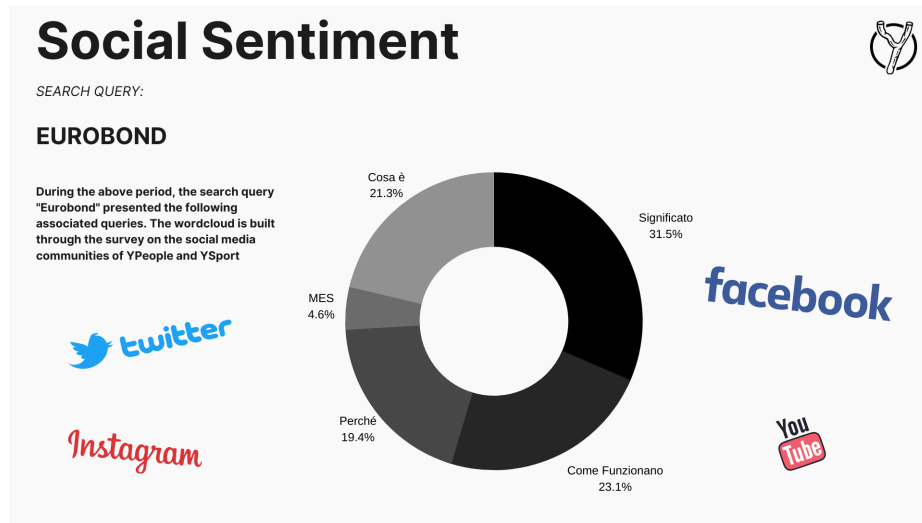
Fig. 5. Sentiment module

about "eurobond". Subsequently, two different *YTalk* were organized, with the same banker and with other experts in the sector, to treat the contents in a more exhaustive way. The graphs below show, in numbers, the level of engagement with the reference community.

As shown in the Fig. 7 the short video pill got more than 7 thousand views and more than 14 thousand impressions only on *Y Tv*, the *Youtube* channel of *Y Media Analytica*. Really, all the unique views and accesses obtained on the sites (the newspapers) and the other social channels must be considered. For the purposes of this work, however, we have limited ourselves in to reporting only the visible data, although it should be considered that the *Y Media Analytica* websites, all together, have about 600 thousand unique views, counted on days per month.

5.2 *Case-study 2 - Epidemiology*

Fig. 8 shows trends monitoring about the search term "Covid", peaked from the end of January to the end of March 2020. Also in this case, *YMemory* conducts its queries to trace the feelings and emotions of users about "covid" trend.

During the same period, the term "Covid" presented a set of associated queries such as "covid 19", "cura Italia", "cura covid 19", "decreto cura italia" and "cura coronavirus"(Fig. 9). Before building a worldcloud on the social media communities of *YPeople* and *YSport*, in this case, the social sentiment module applied a disambiguation algorithm, in order to rephrase the meaning of the two terms "cura Italia" and "decreto cura Italia", that were referred to the law decree containing the word "cura" (i.e., cure) within its name, and the medical term "cure", understood in its meaning as cure for covid-19 disease (Fig. 10).

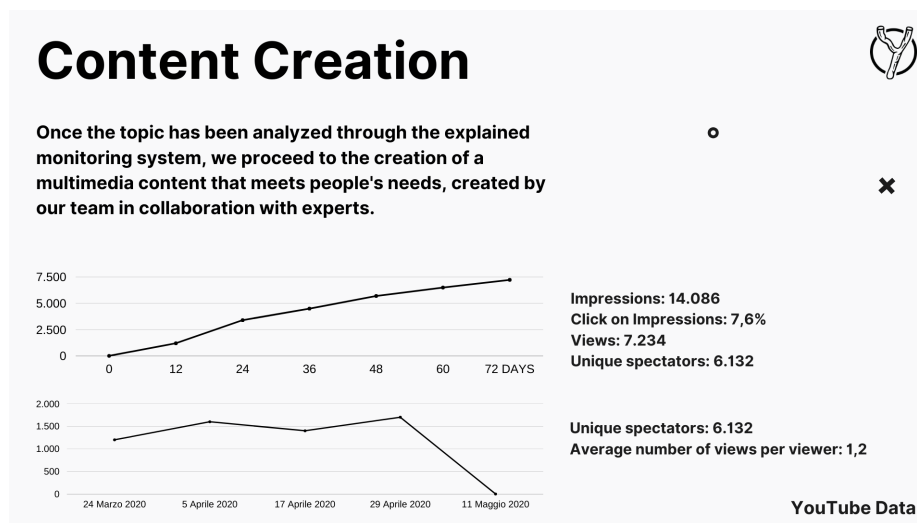**Fig. 6.** Local website monitoring module for websites and all the newspapers

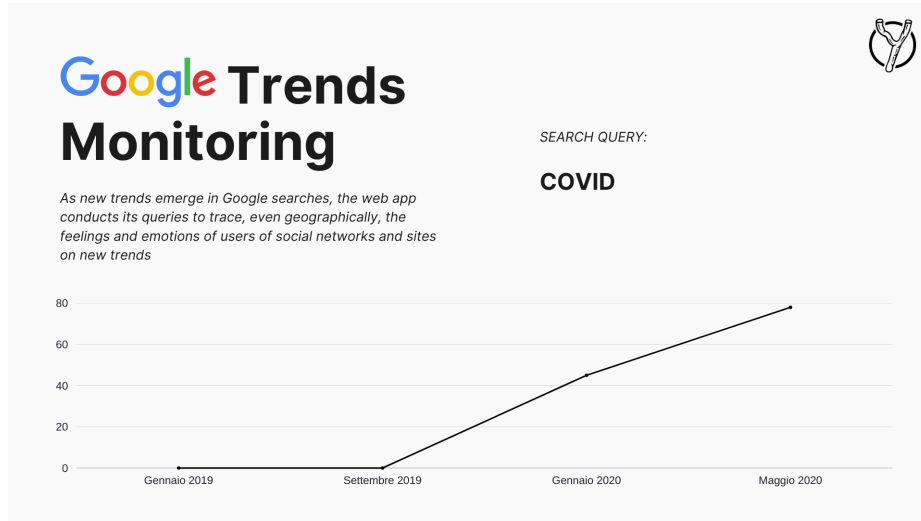**Fig. 7.** Content analytics about Eurobond video
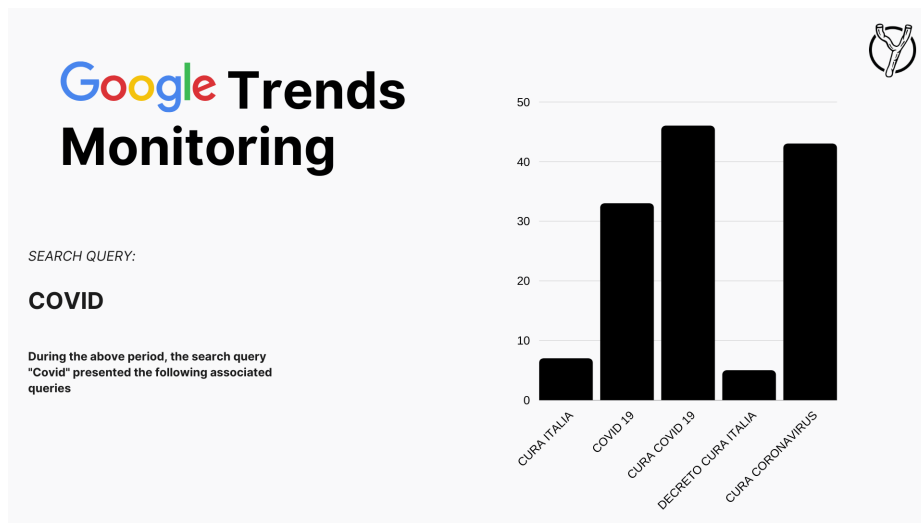
Fig. 8. Trend monitoring for *Covid*

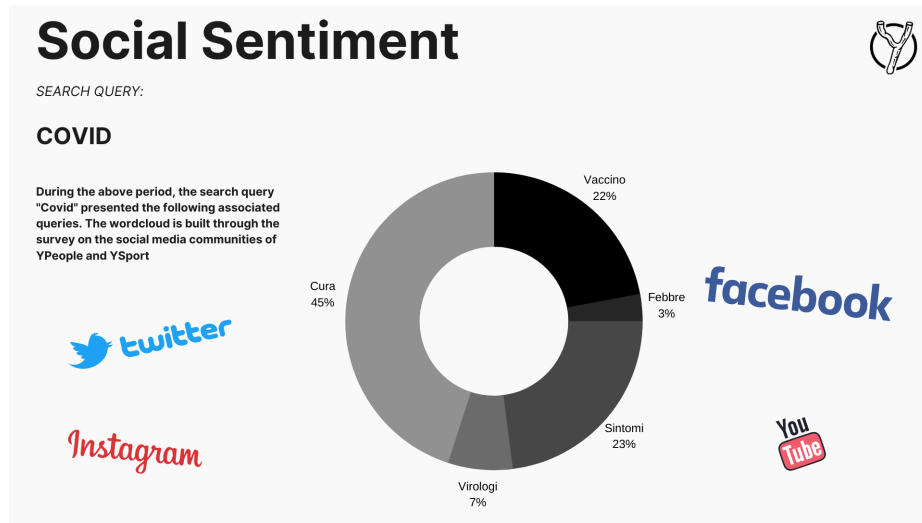**Fig. 9.** Associated search query to *Covid*

Fig. 10. sentiment module for *Covid* analysis

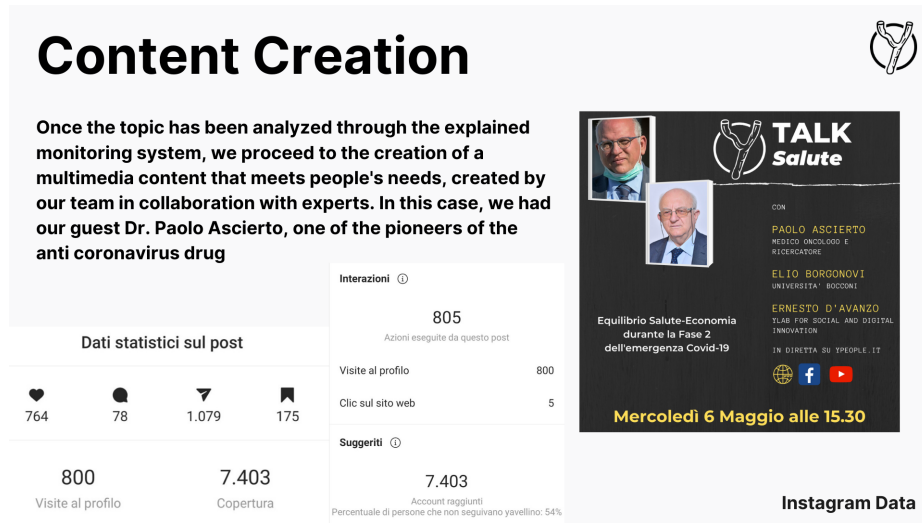**Fig. 11.** Local monitoring module for *Covid* analysis

Fig. 12. Content analytics about *Covid* video

It is interesting to take a look at the results that appear on the dashboard in the *local website monitoring module*. As can be seen from the Fig. 11, many local networks and newspapers, resumed the Ascierto cure, in reference to the Italian scientist Paolo Ascierto, who proposed a cure that uses the drug *tocilizumab* to treat covid-19 patients. Most of the 14 *Y communities* around Italy had a great deal of interest in the Ascierto cure. In addition, some of the communities, as they have many fans abroad, because linked to their homeland for their favorite sports team, also tried to communicate with their friends overseas to reassure their loved ones. Many of the discussions opened on the *YSport* and *YPeople* networks were focused on this topic.

Given this, the *YTeam* immediately contacted prof. Ascierto so that he could give all the information concerning the cure and the drug proposed by him. Professor Borgonovi, one of the leading Italian public health experts, was also invited with the physician. The two scholars confronted all the followers of the communities during a *YTalk*. Below are the data regarding only the announcement post of the guests called to speak. The *YTalk*, transmitted through the various social channels of the *Y communities*, has obtained tens of thousands of views, demonstrating the great expectation regarding the care and interaction with the Italian scientist (Fig. 12).

5.3   *Case-study 3 - Social behavior*

The covid-19 pandemic, and related government restrictive provisions, markedly changed populations' habits during the lockdown. The virus has changed the habits of daily life but has also significantly affected the life plans of people, families and so on. The *YLab* for social and digital innovation has investigated

many of these habits and behaviors, in different areas. For example, the use of *Netflix* during the covid-19 emergency was investigated, as well as *distance learning*, in relation to the student-teacher relationship. It has been also analyzed citizens' trust in institutions during the Covid-19 emergency, and the purchasing behavior of adolescents regarding clothing fashion. An aspect that intrigued particularly, at the level of *social behavior*, was that of couple life and its planning which, in many cases, is resolved on the wedding day.
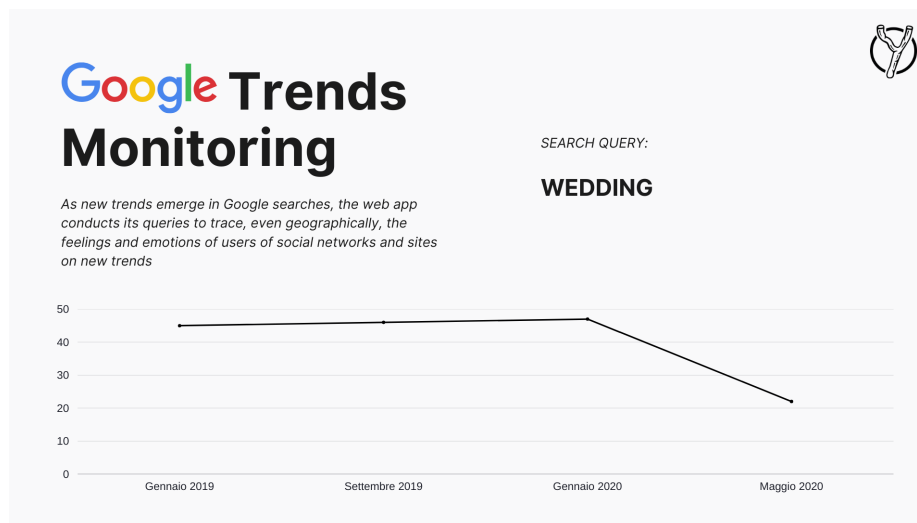

Fig. 13. A drop in Google's search query data regarding "Wedding"

In this latest case-study of the paper, we will briefly describe only the wedding case, since the answer to this trend, emerged from search query and the related behavioral study employing the *YLab mode*, has involved a well-known Italian fashion designer and a wedding planner, whose interaction with during the YTalk .it was highly appreciated both by the *Y communities* and by the young students of the universities involved. It was also an opportunity to create a synergy with a local information portal, *SposIn Campania*, that deals with information and organization of weddings. Last but not least, the wedding case study allows to analyze an aspect that, if analyzed only with the *google search query*, would be of little interest. Thanks to *YMemory*, on the other hand, it is possible to recover aspects that would otherwise be lost and that represent hot and highly requested themes, which emerge from the discussions within the *Y communities* and that *YMemory* manages to grasp.

The dashboard in the Fig. 13 shows, counterintuitively, a drop in Google's search query data regarding "Wedding", starting from January and until May, therefore coinciding with the lockdown. At first glance, this would suggest a disinterest that has developed towards the wedding.
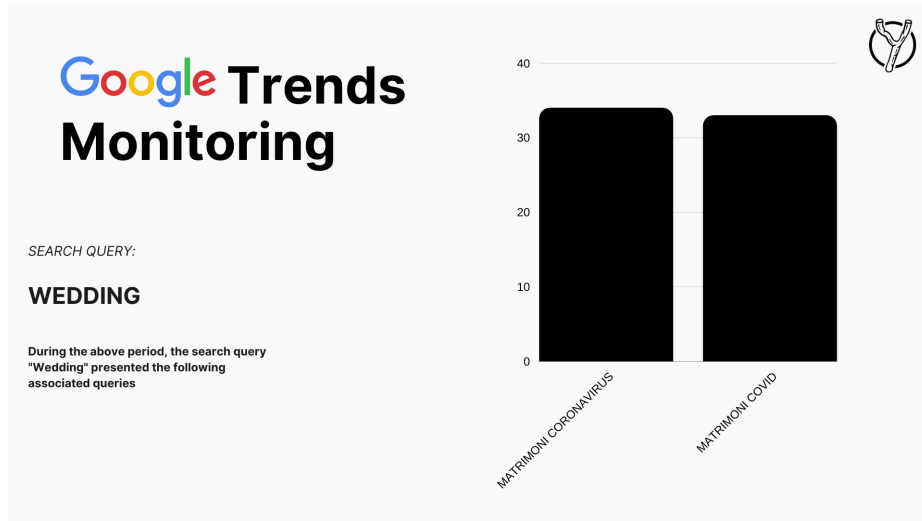
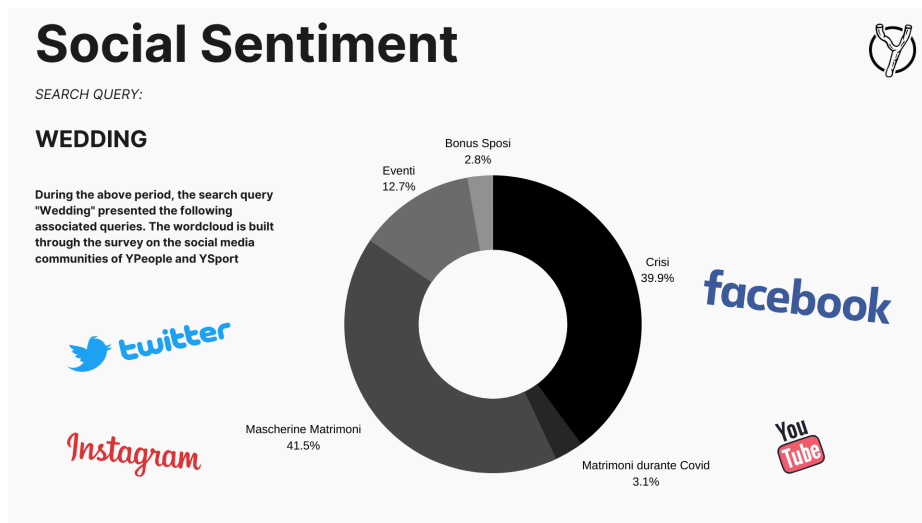**Fig. 14.** Searches related to the terms "matrimoni coronavirus" and "matrimoni covid"

Fig. 15. Social sentiment dashboard about Wedding

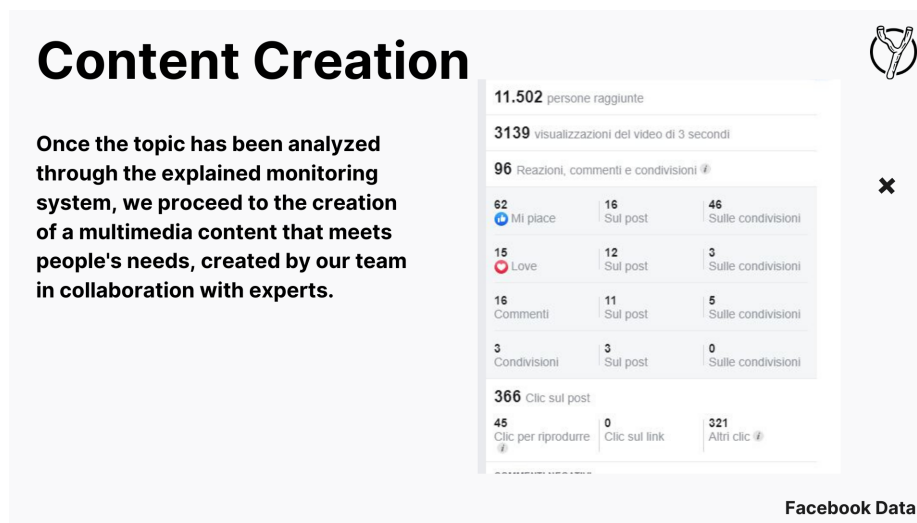**Fig. 16.** Local monitoring about Wedding

Fig. 17. Content analytics about *Wedding* video

If we look at the associated search queries (Fig. 14), however, we can notice searches related to the terms "matrimoni coronavirus" and "matrimoni covid", to underline a form of interest in the subject, but through the "lens" of the term covid and/or coronavirus.

A look at the social sentiment dashboard of *YMemory* (Fig. 15), then, allows you to get a quick idea confirming a "latent" interest that exists in the communities given the sentiment that emerges from the *analytical wordcloud*. For example, the emergence of sentiments such as "mascherine matrimoni, "matrimoni durante covid", "eventi", and "bonus sposi" were the alarm bells to understand, through subsequent analyzes, that in the *Y communities* there was a strong interest in celebrating one of the most important days of one's life.

Subsequently, thanks to the use of the local website monitoring module (Fig. 16), an alternate orientation was seen to deal with the topic. In some cases they were announcements of a recovery, while in others they spoke of alternative employments for wedding planners.

Having said that, a synergy was born with a specialist information portal that also deals with planning wedding ceremonies. The portal operates mainly in the Campania region, where there are 4 Y communities for about 50,000 fans. The synergy saw the involvement of one of the major young Italian wedding planners and a stylist specializing in nationally renowned wedding dresses and beyond. At this point, a YTalk was organized on the channels of the communities involved, with the results that can be seen in the dashboard shown in the figure.

For the sake of brevity, the data shown in the figures refer only to the post of one of the 4 facebook channels of Y Media present in Campania (Avellino, Benevento, Napoli and Salerno). As can be seen, only on this channel, more than 11 thousand people were reached and more than 3 thousand views counted (Fig. 17).

The contents treated during the *YTalk* were very exhaustive, dealing with both the organizational and informational aspects. A great deal of attention has been given to the relational and human resources aspects, given also that wedding planners have to deal a lot with relational aspects with their customers. An interesting testimony in this regard was that of the wedding planner Steve Di Maio. The fashion designer Gianni Molaro, then, attracted a lot of public attention, first of all for his human approach to the problems caused by covid-19, with a splendid testimony. Being Molaro a well-known and appreciated stylist, his testimony was undoubtedly exemplary in terms of leadership for professionals, above all. His opinions regarding the idea of marriage, as a social fact, were noteworthy, giving considerable awareness also in cultural terms and also lending himself to broader cultural debates.

6   *Conclusion*

The paper reported on *YMemory*, a process implemented at *YPeople* and *YSport*, the two online newspapers of *Y Media Analytica*, an innovative start up based in Trento. Starting from Google trends, YMemory, can support decision-

makers (e.g., journalists, analysts, social scientists, doctors and so on) in the creation and development of ad hoc content that can satisfy specific geographical reference communities, with respect to different real-world phenomena such as, for example, epidemiology, economics, social sciences, and so on.

Three case studies have been described in detail to show *YMemory* in action. The case studies show how the use of the YMemory pipeline allows the identification of ad hoc trends on the basis of which to build dynamic content, which provides real-time interactions with the public, in order to create new content, in a sort of virtuous circle. Although the use of google trends is a fairly widespread practice in web and *seo oriented journalism*, the other modules of YMemory make up for the shortcomings of searches that are not registered on google trends but that are part of the social activity of citizens. In this sense, the *sentiment module* and the *local monitoring module* allow to have the pulse of the situation in the reference communities where the YSport and *YPeople* channels are spread. An example is the case-study 3, related to the wedding. The google search queries did not suggest anything interesting for the citizens, but a look at the local communities, thanks to YMemory, revealed the interest of thousands of people for whom the topic was of extreme interest. Without mentioning the cases in which physicians were involved, for health, or economists, to understand the state of the art on the various social and economic solutions that were proposed at government level.

In this sense, the proposed pipeline represents a contribution of *social innovation process* because it aims to satisfy social needs better with respect to the existing solutions, coming from the community. Moreover, thanks to the process described, it could be created a platform able to involve civil society, for instance, through *citizen journalism processes*.

References

1. Alaei, A. R., Becken, S., Stantic, B.: Sentiment Analysis in Tourism: Capitalizing on Big Data. Journal of Travel Research, **58**(2), 175–191 (2019)
2. Chen S., Mao J., Li G., Ma C., Cao Y.: Uncovering sentiment and retweet patterns of disaster-related tweets from a spatiotemporal perspective ? A case study of Hurricane Harvey. Telematics and Informatics, **47** (2020)
3. D'Avanzo, E., Pilato, G. and Lytras, M.: Using Twitter sentiment and emotions analysis of Google Trends for decisions making. Program: electronic library and information systems, **51**(3), 322–350 (2017)
4. D'Avanzo, E., Pilato, G.: Mining social network users opinions to aid buyers' shopping decisions. Computers in Human Behavior,**51**, Part B, 1284–1294 (2015)
5. D'Avanzo E., Kuflik T., Elia A.: Online Advertising Using Linguistic Knowledge. In: D'Atri A., Ferrara M., George J., Spagnoletti P. (eds) Information Technology and Innovation Trends in Organizations. Physica-Verlag, (2011)
6. Holsapple, C. W., Hsiao, S. H., Pakath, R.: Business social media analytics: Characterization and conceptual framework. Decision Support Systems, **110**, 32–45 (2018)
7. Contreras, D., Salam, M.: Data-driven decision making in critique-based recommenders: from a critique to social media data. J Intell Inf Syst, **54** , 23?44 (2020).

8. Goyanes, M.: Why Do Citizens Pay for Online Political News and Public Affairs? Socio-psychological Antecedents of Local News Paying Behaviour, Journalism Studies, **21**(4), 547-563 (2020)

9. Ginsberg, J., Mohebbi1, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature, **457**(7232), 1012–1014 (2009)

10. Hoskins, B.L., Mascherini, M.: Measuring Active Citizenship through the Development of a Composite Indicator. Soc Indic Res **90**, 459–488 (2009)

11. Jagdale R.S., Shirsat V.S., Deshmukh S.N.: Sentiment Analysis on Product Reviews Using Machine Learning Techniques. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, **768**. Springer, Singapore (2019)

12. Jones, L. M., Mitchell, K. J.: Defining and measuring youth digital citizenship. New Media & Society, **18**(9), 2063?2079 (2016)

13. Kaplan, A., Haenlein, M.: Rulers of the world, unite! The challenges and opportunities of artificial intelligence.Business Horizons, **63**(1), 37–50 (2020)

14. Lee, I.: Social media analytics for enterprises: Typology, methods, and processes. Business Horizons, **61**(2),199-210, (2018)

15. Lister, R.: Inclusive Citizenship: Realizing the Potential. Citizenship Studies, **11**(1), 49–61 (2007)

16. Makarius, E. E., Mukherjee, D. Fox, J. D., Fox, A. K.: Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. Journal of Business Research, **120**, 262-273, (2020)

17. Nblega Carriquiry, A.; Sauri, D.; March, H. Community Involvement in the Implementation of Sustainable Urban Drainage Systems (SUDSs): The Case of Bon Pastor, Barcelona. Sustainability 2020, 12, 510.

18. Olteanu, A., Castillo, C., Diaz, F., Kiciman, E.: Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front. Big Data, **2**(13), (2019).

19. Rezaeinia, S. M., Rahmani, R., Ghodsi, A., Veisi, H.: Sentiment analysis based on improved pre-trained word embeddings. Expert Systems with Applications, **117**, 139–147, (2019)

20. L. Ruoxin, W. Xuan and S. Kang, "Research on the Application of Big Data in Newspaper Industry," 2019 6th International Conference on Dependable Systems and Their Applications (DSA), Harbin, China, 2020, pp. 505-506

21. Shrestha, Y. R., Ben-Menahem, S. M., von Krogh, G. . Organizational Decision-Making Structures in the Age of Artificial Intelligence. California Management Review, **61**(4), 66?-83, (2019)

22. Sims K.M., Weber E.M., Bhaduri B.L., Thakur G.S., Resseguie D.R.: Application of Social Media Data to High-Resolution Mapping of a Special Event Population. In: Griffith D., Chun Y., Dean D. (eds) Advances in Geocomputation. Advances in Geographic Information Science

23. Tsuei, H., Tsai, W., Pan, F. et al: Improving search engine optimization (SEO) by using hybrid modified MCDM models. Artif Intell Rev **53**, 1–16 (2020)

24. Udanor, C., Aneke, S. and Ogbuokiri, B.O.: Determining social media impact on the politics of developing countries using social network analytics. Program: electronic library and information systems, **50**(4), 481–507

25. Valencia, F., Gmez-Espinosa, A., Valds-Aguirre, B.: Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy, **21**(589), (2019)

26. Webb, G. K. (2009). Internet Search Statistics as a Source of Business Intelligence: Searches on Foreclosure as an Estimate of Actual Home Foreclosures. Issues in Information Systems, 10 (2), 82-87
27. Wei Wang, Quan-Hui Liu, Junhao Liang, Yanqing Hu, Tao Zhou: Coevolution spreading in complex networks. Physics Reports, **820**, 1–51, 2019
28. Zhu, J., Wang, X., Qin, J. and Wu, L.: Assessing public opinion trends based on user search queries: validity, reliability, and practicality. In: Annual Conference of the World Associate for Public Opinion Research, Hong Kong, 2012
29. Zhu, C.: Big Data as a Governance Mechanism. The Review of Financial Studies, **32**(5), 2021–2061, (2019)