The European Proceedings of
Social & Behavioural Sciences
EpSBS

ISSN: 2357-1330

https://doi.org/10.15405/epsbs.2019.08.02.4

III PMMIS 2019
Post mass media in the modern informational society "Journalistic text in a new technological environment: achievements and problems"

INFORMATION CASCADE IN NEWS AGGREGATORS AS MECHANISM FOR FALSE ECONOMIC AGENDA CREATION

Dmitry Konoplev (a)*
*Corresponding author

(a) Chelyabinsk State University, 454001, Br. Kashyrinykh, 129, Chelyabinsk, Russia, dmitrijkonoplev@ya.ru

Abstract

As media companies pursue an unimportant news proliferation strategy, they are increasingly relying on developing news aggregators' capabilities to interact and influence with their audience. Albeit, many lack an understanding of what consequences an unimportant news proliferation in aggregators may have. How out of unimportant news arise information cascades? How these cascades are dive into the news agenda of aggregators? How is the topic considered in information cascades reflected in the media after these cascades become irrelevant? How long is the life cycle for information cascades in news aggregators? To address these questions, the study builds on qualitative data from manual and automated content analysis to conceptualize three underlying subcomponents of information cascades, namely, subject groups, timeframes, and distribution scenarios. The study also addresses the issue of the active formation of so-called news spam and its impact on informational cascades in news aggregators. The study identifies and explains how information cascades enable unimportant news co-creation with aggregators through information occasions' intentional design and automatic processing mechanisms. Research also identifies possible options for the media and aggregators that can help solve the problem with the dissemination of unimportant news. This study contributes to the journalism theory by showcasing how information cascades are enabling false economic agenda creation in automated news aggregators contrary to the declared purposes of the latter.

Keywords: Yandex news, information cascade, news aggregators, economic thinking, journalism.

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

Yandex News service is the largest news aggregator that creates an informational agenda for users of the Russian network segment - Runet. Yandex News offers automated ranking of news from the media in each Russian region, based on the citation of these news, the number of information sources covering such news topics, as well as the distribution of relevant publications over time.

The aggregator mechanism, although not controlled by the media itself (it is created by a software robot), allows media to significantly influence the agenda reflected in Yandex news.

Thus, it is obvious that the media can arrange an important informational occasion from any insignificant news and, fulfilling the requirements of the aggregator, bring it up to the news rating, attracting the attention of the audience to it.

If the construction of an important informational occasion from insignificant news succeeds and the aggregator displays a block of such publications on a visible position on the main page of Yandex News in the region, then this news block begins to live its life on its own.

First, such news block receives the attention of an audience reading the relevant news on the website of any media source. However, if the news is insignificant and not interesting for audience, its distribution stops there. On its own, the reader will not share such news in social networks and contribute to its further promotion.

Promotion of such news is associated with the second scenario, the main role in which is played by other media operating in the respective Yandex news region.

Seeing that the news is gaining popularity (the number of sources who have written on this topic is growing, the news rises in the aggregator ranking and so on), other media begin to rewrite this news and publish it on their own, based on previous publications. At the same time, even if these media were not the initiators and primary source of the publication, then by virtue of the competition they use someone else's news so that the potential reader can come to their site for the same text.

Thus, quite unwillingly, some media popularize the news of other media and create conditions for its maximum proliferation. In some cases, this scenario distorts the real news picture in the aggregator and lowers the rating of the more important news, created in a natural way.

2. Problem Statement

From a series of homogeneous news for a certain period of time (usually within a few hours), a spreading on its own information occasion is formed, which we propose to call the information cascade.

Thus, the information cascade is the information and data that is actively disseminated with an increasing rate, and which form a certain angle of perception of reality in a potential audience.

The main problem with the formation of information cascades in news aggregators is that they can occur not spontaneously in a natural way (for example, the active dissemination of news about a catastrophe, an earthquake, a flood, and so on), but according to a well-thought-out scenario.

In other words, the media can initiate the development of an information cascade. Media create the initial information wave, which is picked up by the news aggregator. Such a mechanism, among other

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

things, can be used for propaganda purposes and to have among its tasks misleading the audience. Also, the purpose can include the promotion of a certain position of the initiator of the news by the media.

In some cases, such information cascades can create problems in terms of the representativeness of news in the aggregator, since Yandex News does not have an effective mechanism for identifying such constructs and their exclusion from the news rating.

As Vysotskaya and Kochetkov (2018) note, the headlines in Yandex News can act as cognitive model of a situation. Pei, Muchnik, Tang, Zheng, and Makse (2015) consider the logic of disseminating such information on the Internet, while Belák, Mashhadi, Sala, and Morrison (2016) focus on the effect of hidden nodes on information diffusion.

Choi and Kim (2016) are concerned with the problem of online news' temporal and spatial exploitation, while Anderson (2013) considers a networked concept of journalistic expertise in the digital age.

The problem of informational cascades is considered by Sattari and Zamanifar (2018) and Guarino, Harmgart, and Huck (2011). According to Bikhchandani, Hirshleifer, and Welch (1998) information cascades can influence the behavior of subjects. At the same time, Mahdi (2017) is considering information cascades in connection with complex networks. Spiwoks, Bizer, and Hein (2008) call into question the influence of the information cascades; albeit, Arkhipov (1997) note that the formation of economic thinking by means of the media provides a certain response in the audience's behavior.

In this article we've combined these approaches for information cascades' emerging scenarios, taking into account the specific characteristics of the Yandex News aggregator.

The study offers basic scenarios for the identification of such information cascades, as well as examines possible options for action to prevent the spread of insignificant information occasions under the guise of important news. The approaches of the media themselves to the interaction with the corresponding information cascades are indicated separately.

3. Research Questions

To achieve this purpose, we have formulated key questions clarifying the essence of the emergence, functioning and disintegration of informational cascades:

1) What conditions must be met in order for the informational cascade to arise not naturally in the framework of the aggregator logic, but by media efforts?
2) What media mechanisms increase the importance of insignificant news, which are present in the aggregator's overall rating, but which are in the lower positions and do not form an information cascade?
3) How does the information cascade disintegrate and what happens next with the information occasion that gave rise to it?
4) Can the same information occasion repeatedly form an information cascade when the initial news goes down in the aggregator rating?

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

5) Is it possible to select certain thematic groups in artificially designed information cascades? In other words - is it possible to capture the main blocks of information about events in information cascades?

4. Purpose of the Study

The purpose of the study is to identify scenarios for the formation of information cascades through news aggregators. We also try to determine the key parameters for the implementation of such information cascades in the media agenda. Research also suggests that the life cycle of the information cascade that emerged as part of a news story in an aggregator should be considered as a related goal.

The study points out the problem of the survival of the subject used in the information cascade, after the information cascade itself ceases to exist in the news aggregator.

5. Research Methods

In the study, content analysis (both manual and automated) was used for the sample database, which amounted to 8052 publications in the Russian online media for one calendar year - from January 1, 2018 to January 1, 2019. Statistical and functional analyses, as well as mathematical modeling were also used to verify the results.

The Vaal, Wordstat and QDA Miner software tools were used for research database content analysis. The results were selectively checked manually. The research was based on publications from electronic media libraries (East View Information Services, East View's Universal Database and Polpred.com database), including Russian online media publications (115 media sources) and state portals with official news (87 media sources). A total of 8052 publications on economic subjects were analyzed.

For processing aggregator data and keyword search, selected publications from the different Russian regions (Moscow, St. Petersburg, Sverdlovsk region, Chelyabinsk region, Kurgan region, Perm region, Republic of Bashkortostan, Kemerovo region, Tyumen region and Kaliningrad region) were analyzed through the Pipes Digital platform.

These data were collected together, using conversion analysis and component expertise methods. The basis for content analysis was formed during the heuristic analysis and mash-up sorting by keywords. XML documents were used as a data source (most often in the form of RSS feeds). For their generation, the extended Pipes Digital fork was used, in which the source base (inputs-feed) was used, the combined controller (control - combine), the filtering system by keywords (search operations varied to check the results - from the most general level item.title to item.content and item.summary). The item.link operation was used only partially to test borrowing news from the official sites of the press services.

As an exception for research database, all base words with a negative connotation (crisis, problems, unemployment, etc.) were chosen to exclude from the news feed the cases of mentioning the main characters of information cascades in a negative context. Data fork in text form loaded into the general database of content analysis.

Branching of information cascades was evaluated in the aggregator Yandex News by statistical methods.

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

6. Findings

The study showed that in all the considered Russian regions there are informational cascades arising on the basis of insignificant and artificially constructed news. This news were initiated by a group of publications (in each region of their own), almost simultaneously starting to cover insignificant topics of economic subjects.

All these topics can be divided into five subject groups: meetings of regional officials (group 1), visits by regional officials to official events (group 2), visits by officials to regional companies and industries (group 3), appointment of regional officials to new posts or transfer of additional functions to them (group 4), the presentation of regional officials with reports (group 5).What do these publications have in common? Their volume is almost always more than 600 characters with spaces, from four to six paragraphs, the title or leader-paragraph contains the official's last name, the text always uses a quote with reference to this official, as well as an illustration taken from the corresponding official site ministries, the site of the governor, the regional government, the city administration, or, in the case of Moscow, the prefectures). In the text of the quotation, officials are accompanied by such phrases as "life improvement", "promising direction of development", "accessible environment", "system improvement" and so on.

Such news is presented as socially significant and the main speakers in the publications are officials. Even if there are other participants in the news story (for example, doctors of the hospital, at the opening of which the mayor of the city came or businessman with whom the governor discussed issues of cooperation), the words are not given to them in most cases - in the text their presence is either indicated by an indirect quote or they are awarded only the captions on the common photo posted with the publication.

Such publications acquire the character of an information cascade within one hour, when several media at once place similar news with sometimes completely identical headlines.

For example, the news story which was in the top of Yandex News aggregator in Moscow on February 4, 2019 under the general title "Sobyanin told about the improvement of the metropolitan medical care system" during the first day formed an information cascade of four publications, after which this cascade left the top of the rating. On the morning of February 5, 2019 this subject story was supplemented with three more news of the same type on the same topic, which enabled for information cascade to hold out in the aggregator's rating until 2:05 pm Moscow time. At the same time, all the news in this story repeated each other both meaningfully and compositionally, and two publications within the 21-minutes interval even gave the news with the same headline: "Sobyanin told about improving the oncological care system" (headlines from "Moscow 24" and "TVC"). The source material itself was taken from the site of the mayor of Moscow.

Regional publications that form rating news from insignificant and imperceptible subjects on economic topics also work with aggregators according to the same scenario. The main difference between regional media and Moscow media is that there are fewer participating sites in the regions that create such information cascades. Albeit, due to the limited number of news stories during the average day, it is easier for them to get into the top lines of the rating and stay there.

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

The capital's news agenda is more diverse and therefore small editions, duplicating the same news on the same topics, are connected to the formation of information cascades besides relatively large web-portals. If the sources of the cascade can be, for example, the websites of TVC or Moscow 24 channels, then, within just a few minutes, the news begins to be actively distributed by such low traffic websites as "Starry Boulevard", "Eastern District", "North of the Capital" and other resources of the prefectures of the administrative districts of Moscow.

In other words, the websites are promoting official news in Yandex news service, building it up according to the rules of the aggregator - at least five sources in the subject, original (made through rewriting) news text that appear within an hour. In most cases, this news is the processing of press releases from official sites.

And if in Moscow and St. Petersburg such news falls in the top five news aggregator on average once every two days (245 and 188 cases in our sample), in other Russian regions examined, similar news are present in the upper part of the aggregators' daily rating almost constantly. The leaders here were Chelyabinsk region (327 cases) and the Kemerovo region (302 cases). However, in the general aggregator rating (the total number of news of this kind in the Yandex news rating) Moscow and the Moscow region were in the first and second positions.

In all these cases, the information cascade took shape within an hour, dialing from 4-5 to 7-8 sources in the Yandex subject news by region and from 7-15 sources in Moscow. After that, it could be expanded due to publications in other media, which quoted already existing sources, or began to leave the rating, since this kind of news did not attract anyone except their initiators.

The study showed that in this case, there were two options for the information cascade development. Either the cascade outlived itself and left the rating during the working day (usually by 8:00 pm local time), or after 12-14 hours a few more publications appeared on the same subject (usually not more than three) and the news topic temporarily maintained in the rating. From now on the information cascade was no longer able to return to its previous positions. This information cascade was replaced by the next information cascade, and so on.

It is noteworthy that the heads of Russian regions, whose surnames were most often present in regional information cascades, were in the top ten of the media rating of the governors in 2018, released by Medialogia Research Company. Sergei Sobyanin (Moscow) took the first place, Andrei Vorobyov (Moscow region) the second, the head of the Sverdlovsk region Yevgeny Kuyvashev was on the seventh line of the rating, the head of the Chelyabinsk region Boris Dubrovsky - on the eighth. At the same time, the head of the Kemerovo Region, Sergey Tsivilev, who took his post less than a year ago, also showed a successful result of media presence, starting right away from the 15th line of the Medialogia rating (Media rating of the governors, 2018).

Comparing the research data with Medialogia rating, we concluded that creating a false economic agenda in the Yandex News aggregator is a conscious process. It can be assumed that informational cascades are formed in order to create a positive image of regional officials and their falling into the overall rating of mentioning.

The study showed that none of the identified information cascades on the basis of the news subjects did not stimulate the growth of traffic to their publishing sites. Moreover, even news position in

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

the top five in the aggregator does not change the traffic structure on the destination site. Thus, the only visible effect of such publications is a quantitative increase in the mention of the relevant officials, albeit, the positive news about officials can partially influence the economic thinking of the audience, creating an artificial picture of economic reality.

Media expert Khomak (2017) names such sites sources as "news spam" and notes that only resources that create information cascades based on press releases from Moscow mayor's office are at least 452.

"In any story, hundreds of sources - but they all repeat the same press release, written in a certain private office <...> This press release is sent to another private office, in which people from the regions for penny salaries are clumsy rewrite it and put it on one of the hundreds of websites that no one will ever read", adds the founder of the Noodleremover project, Korolev (2017).

In fact, this approach violates the very principle of Yandex News to inform network audiences, replacing the important news with the PR materials of officials of various ranks. Thus, Chernetsky (2015) believes that the aggregators are responsible for reducing the quality of news on the Internet.

7. Conclusion

The study showed that the principle of formation of information cascades based on irrelevant news subject always follows the same scenario.

Each news subject that has become a full story in the course of replication for a news aggregator has an official source that is not a mass media and does not possess exclusive data (department, ministry, and press service). In most cases, it is not present as an independent information channel within the framework of the aggregator news range, but only serves as a data source for the media that replicates it.

The final format of the news subject is actually a revised press release, placed with the intended advertising goal and forming a false impression on the current economic agenda of the potential audience (mainly due to quantitative distribution through partner websites). This kind of news has similar qualitative characteristics, as well as the formal logic of complementarity (for example, different sites use different fragments of the same quotation from the initial press release so that the news aggregator does not consider them as repetition). At the same time, the initial press release is replicated (simplified) so that in the information cascade there is an emphasis on a certain part of the original text (for example, the words of the official).

The news aggregator does not affect the formation of such kind of constructed news subjects, does not label them, does not exclude from the ranking. The Yandex News algorithm does not have a special tool for tracking such publications.

At the same time, the replicated news is distributed in Runet according to the principle of an information cascade (repetition of the same type of bigrams, similar parameters of the indexing step by the news aggregator robot, etc.) and do not have features typical for classic news based on direct information subject created without press releases of departments and press services.

In the information cascades, formed on the principle of news spam, there are no original photos, there is no exclusive material (for example, comments that were taken directly by the journalists) as well as there is no problem statement. In such news any criticism of the officials is also completely absent.

https://doi.org/10.15405/epsbs.2019.08.02.4
Corresponding Author: Dmitry Konoplev
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

In some cases, news spam on economic subjects in news aggregators begins to crowd out the classic news in the same subjects. That is, naturally arisen news containing five websites of sources obviously loses to artificially generated news, the number of sources in which is much higher.

For example, in September 2018 in the Chelyabinsk region for the regional rating of Yandex News, an information cascade dedicated to the statements of the Governor Dubrovsky about the upcoming construction of the Chelyabinsk-Yekaterinburg high speed railway, forced out of the rating news about garbage collapse in the city.

Similarly, in February 2018 in the Kurgan region in the aggregator information cascade with a rewritten press release about the celebration of the 75th anniversary of the region forced out of the news agenda the news that none of the heads of neighboring regions came to the celebration, sending a telegrams only and some officials even referred to the disease, not to come on the event.

Research suggests that this is the second function of the information cascades - to jam the negative news agenda in the aggregator, creating the conditions for the widest possible distribution of positive news about regional officials.

At the same time, it is important to note that information cascades built on the principle of news spam exhaust their potential within a maximum of 24-48 hours from the moment of publication and leave the aggregator rating. The study showed that such information cascades did not return to the rating as part of their news subject. Thus, one news subject or press release can form only one information cascade. Therefore, for the constant presence in the ranking of aggregators, the spread of news spam is put on the stream.

This indicates that such information cascades are created artificially. Otherwise, press releases from official sites could not be noticed at all by most resources. In addition, in cases where the media use such an official text, it could obviously be presented in the form of several news subjects, since the authors of the news would highlight different topics within the same press release. If, for example, from a meeting of the regional government or a long speech by the mayor or the governor, the media, as if by agreement, give the same quotation and draw the same conclusion, it means that the goal of focusing on this quotation or conclusion was originally planned.

Thus, it is possible to exclude such materials from the aggregator rating on the basis of simple lexical and compositional analysis, as well as through the analysis of the semantic web of the text (to the extent that mentioning officials in a certain context). That is, technically the problem is solved at the level of the news aggregator itself.

Meanwhile, the role of independent media in this case is limited. They can ignore the news subjects built on the basis of news spam and not participate in the distribution of intentionally designed information cascades only.

In turn, the websites themselves, regularly distributing news spam, at the moment do not have any difficulties with the formation of artificial information cascades. Their only task is to rewrite the initial press release and publish it at a strictly designated time.

References

Anderson, C. (2013). What aggregators do: Towards a networked concept of journalistic expertise in the digital age. *Journalism*, *14*, 1008–1023.

Arkhipov, A. (1997). Utverjdenie novogo economicheskogo myshleniya I zadachi teorii obrazovaniya [New economic thinking's approval and the tasks of the theory and education]. *Russian economic journal*, *8*, 95.

Belák, V., Mashhadi, A., Sala, A., & Morrison, D. (2016). Phantom cascades: The effect of hidden nodes on information diffusion. *Computer Communications*, *73*(1), 12–21.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *Journal of Economic Perspectives*, *3*, 151–170.

Chernetsky, P. (2015). Vliyanie novostnykh agregatorov na kachestvo internet journalistiki [News aggregators influence on the quality of Internet journalism]. *Vestnik PSLU*, *3*, 149–156.

Choi, S., & Kim, J. (2016). Online news flow: Temporal/spatial exploitation and credibility. *Journalism*, *18*, 1184–1205.

Guarino, A., Harmgart, H., & Huck, S. (2011). Aggregate information cascades. *Games and Economic Behavior*, *73*(1), 167–185.

Khomak, D. (2017, September 8). *Sobyachki: kak spam moskovskoi merii ubivayet journalistiku* [Sobyachki: How Moscow City Hall Spam Kills Journalism]. Retrieved from https://noodleremover.news/sobya4ki-31ee651ef358

Korolev, A. (2017, September 6). *Kult lichnosti Sergeya Sobyanina, kto emu sluzhit, kto platit za nego I skolko* [The cult of Sergei Sobyanin's personality, who serves him, who pays for it and how much]. Retrieved from https://noodleremover.news/sobyanin-cult-4aeb0395543c

Mahdi, J. (2017). Information cascades in complex networks. *Journal of Complex Networks*, *5*, 665–693.

Media rating of the governors (2018, January 21). Medialogia Research Company. Retrieved from https://www.mlg.ru/ratings/vlast/governors/6448/

Pei, S., Muchnik, L., Tang, S., Zheng, Z., & Makse, H. (2015). Exploring the Complex Pattern of Information Spreading in Online Blog Communities. *PLoS ONE*, *10*(5), e0126894. https://doi.org/10.1371/journal.pone.0126894

Sattari, M., & Zamanifar, K. (2018). A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks. *Journal of Computational Science*, *25*, 122–133.

Spiwoks, M., Bizer, K., & Hein, O. (2008). Informational cascades: A mirage? *Journal of Economic Behavior & Organization*, *67*(1), 193–199.

Vysotskaya, I., & Kochetkov, P. (2018). Zagolovok internet-novosti kak kognitivnaya model situatsii [A headline of the Internet news as cognitive model of a situation]. *Vestnik NSU. Series: History and Philology*, *17*(7), 89–96.