

Targeted Romanian Online News in a Mobile Application Using AI

Marius-Cristian Buzea
University Politehnica of
Bucharest, Department of
Computer Science and
Engineering, Bucharest,
Romania
bumarius@gmail.com

Ștefan Trăușan-Matu
University Politehnica of
Bucharest, Department of
Computer Science and
Engineering, Bucharest,
Romania
stefan.trausan@cs.pub.ro

Traian Rebedea
University Politehnica of
Bucharest, Department of
Computer Science and
Engineering, Bucharest,
Romania
traian.rebedea@cs.pub.ro

DOI: 10.37789/rochi.2020.1.1.9

ABSTRACT

Nowadays, being informed is a very important aspect of our life. Few online news broadcasters, such as bbc.com or cnn.com, send real-time mobile notifications with several interactivity facilities with the application to be informed with the latest news. In this paper, we present an application based on Machine Learning (ML), a basic technique used in Artificial Intelligence (AI), available for mobile devices, which detects irony, the articles' polarity and the fake news. The mobile application is based on a machine learning system using manually annotated Romanian corpora, and a crawler network designed to automatically extract data from more than 730 public websites. The aim of the application is to collect information regarding user behavior, to change and adapt the content to user's preferences and requirements in order to process and better understand the user's necessities. Furthermore, the user is able to search through the news, filter the news, create favorites lists and visualize graphical representations for different time periods. Therefore, the main goal of the application is to develop and improve the way that people receive and access reliable information.

Author Keywords

Adaptive Mobile News Platform; Artificial Intelligence; Machine Learning; Romanian Corpora; Graphical Representations;

ACM Classification Keywords

H.5.2: User Interfaces.

INTRODUCTION

Nowadays, the number of mobile devices surpasses 3 billion and is estimated¹ to grow with more than 1 billion in the next few years. Recent statistics show that the global mobile data traffic is increasing every year with 30%², this

¹ <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>, last accessed on 5th July 2020

² <https://mylittlebigweb.com/en/should-you-switch-to-amp-format-for-better-seo-performance/>, last accessed on 5th July 2020

aspect being illustrated by Figure 1, which presents the traffic data from a Google analytics account belonging to a website with more than 2.5 million pages visited per year. The rising number of mobile applications is generating a significant amount of useful data, which can help a company or a public person with online business strategies. At the same time, this phenomenon can have a negative impact for some business if the data collected cannot be selected and filtered by a specific target or topic. Moreover, analyzing and achieving graphical representations of statistical data for different time periods give to users the opportunity to take measures in case of the occurrence of unpleasant situations.

How and when people use their mobile devices is another aspect involved before starting to design and develop the application, because the mobile users often read news on their gadgets in the car or subway. In general, when creating new attractive applications, most programmers adopt the same principles that are normally applied to the web interfaces, but it is known that the issues are more pronounced due to the limited mobile space, a solution being the simplicity. The display is narrow or the network connectivity is slower, but what we lose in presentation layer, we gain in having a carefully designed user interface or better interactions with the technology, having gadgets with touch screen being more natural [1].

There are several Romanian news mobile platforms available in Google store and App store, such as “Pressa - Știri din România” (<https://www.pressa.app/>), DCNews.ro or “Observer” (<https://observatomews.ro/>) but these applications do not have advanced integrated facilities, such as fake news detection, ironic news detection or negative news detection.

Most of the existing applications extract data from different sources using public RSS feeds, browser extensions or poor scrapers. Moreover, what really matters is the number of followers or the number of subscribers on social media platforms. Because of these rules, the online publishers require their employer's exclusivity and the above sources, most of the time, are not immediately updated.

Our developed mobile platform offers a solution to these issues, by using Machine Learning (ML), a basic technique used in Artificial Intelligence (AI) for the extraction process to achieve high performances. Furthermore, the application is adapting the content to the user's preferences and necessities by registering several keywords for every topic or public target.

Another important aspect of our platform is represented by the content, such as images, texts and videos which are stored in our application. The known applications store only the title of the news with a link through the original source and, in most of the cases, when the online publishers delete the information from their website, the user cannot access that source anymore.

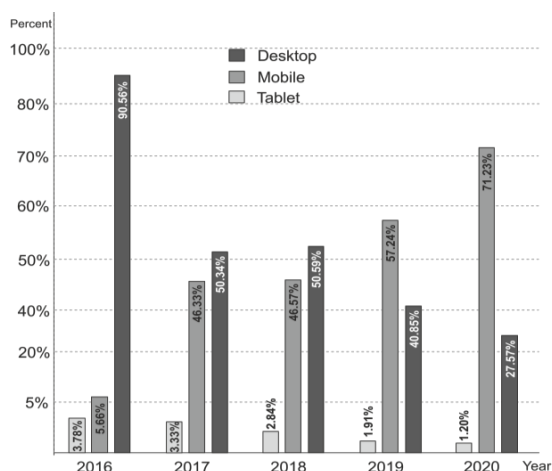


Figure 1: Comparative data traffic

The datasets used in machine learning processes to identify the news polarity, the ironic, fake or negative news in Romanian online broadcasters are another important aspect of this paper. The datasets are based on a Romanian language dictionary, a corpus of non ironic news and ironic news items which were manually annotated by employees in a public institution from Romania.

One of the objectives of this paper is to describe the module for automatic data selection of the mobile application, based on web page properties. This module is used in further extraction mechanisms, representing the main element of the machine learning system. To achieve content for our platform, we used keywords and html selectors, such as "div", "meta tags" or "img".

This paper is also presenting mobile platform techniques, which, alongside machine learning algorithms, might be useful in using these architectures in future mobile application developments.

The paper is organized as follows: next section presents the status and trends related to the developed application; the following section describes the method and main components of the platform; next, the detailed

architecture is presented, followed by the user interface section; finally, last section presents the conclusions of our work.

STATUS AND TRENDS

With so many smartphone users across the world, who could download in 2019 between 2.6 million Android and 2.2 million iOS applications, it is no surprise that the mobile applications industry is expanding [2]. The Apple App Store [3] and the Google Play Store [4] dominate this field, being the most important players in domain. Several applications are free and others have a price, the profit being split between the application's creator and the distribution platform. App usage and mobile devices are still growing [5] at a constant rate and it shows no signs of slowing down. We use our devices at home, in subway and even in our bed, spending most of the time on applications, such as Facebook, Twitter or Instagram. These social platforms are adaptable, innovative [6] and have contributed to our way of communication, creating a way for people to update and share content with their friends and even to learn how these can be used in the business online market campaigns.

Mobile devices offer now more facilities than desktops, due to various functionalities that detect the location or surrounding places, the portability and attractive mobile user interfaces [7]. Compared to desktop user interface, the mobile user interface design requirements [8, 9], are different for every application to ensure usability, consistency, and readability [10]. Due to the smaller screen, the mobile interface [11, 12] focuses on efficiency and discoverability, therefore the user should be able to understand a command through the icons, such as trash or heart pictures, defining the delete action or adding to favorite list.

There are several Romanian applications which follow the above requirements being carefully designed and developed, such as Google News, Biziday (<https://www.biziday.ro/>) and Digi24 (<https://www.digi24.ro/>), designed to receive updates in real time from different public sources, which were the most downloaded Romanian online news applications in 2019.

METHOD

For the past several years, the Google Play Awards have recognized the top applications for Android. The nominees were selected by various teams across Google for nine different categories. These applications should provide the best experience for users, interactive user mobile interfaces, new technologies, but these platforms are not delivering information such as news to the communities. We consider that there is still enough space for the development of these types of applications in online media. Most people would not even consider getting a physical morning newspaper anymore, so they need to use digital sources for daily news. The developed application presented in this paper tries to help users from the Romanian environment to get useful

and reliable information depending on their interest topics, categories, targets or preferences.

Our application has an additional number of new features integrated than the above specified platforms, such as fake and ironic news category, a polarity category or a public targets category.

This application is based on the following components:

- The machine learning component.
- The core and backend web application components.
- The mobile user interface component.

The designed platform serves up news updates from a variety of sources giving to user the opportunity to receive or to be notified with the most relevant news. Moreover, the user is able to create favorite lists or other lists with certain publications saved to be read later.

ARCHITECTURE

Figure 2 illustrates the three main components of our platform, each of them supporting the system in different phases: from the machine learning process to the user interfacing.

In the first component, the machine learning system's components are integrated using PHP, MySQL and Python scripts, constituting the environment that allows the next processes to use and develop the interface between the AI modules and human users, between generated data and user's preferences. This component is based on more than 25,000 items news and a Romanian language dictionary containing more than 42,000 words. To evaluate the news polarity is used a semi-supervised machine learning system based on a three word-level approach [13]. To detect the irony in the online news for the Romanian language is used a corpus based on 14,064 ironic news items from the online environment based on automatic irony detection approach [14].

Another important aspect of this application is the extraction process which is based on more than 730 websites whose properties were manually annotated. After having added new websites to the application, the system evaluates the html properties by searching similar patterns with the best fit from database, to create articles for the latest inserted domain. Otherwise, when the website has not valid properties and the pages cannot be scanned, the platform requires human intervention. The bots network extract and expose only public content from REST API plugin or robots.txt file for the purpose of creating new value from data. Furthermore, it is provided a User Agent string that reveals the application's intention and the requested data based on DOM elements, such as div, are performed at a reasonable rate to consume minimum hardware and bandwidth resources to be never confused with a DDOS attack.

The second component provides the mandatory processes, generating the necessary data by the application core component, designing the connection between UI (User Interface) and AI. In the web application component, several keywords are manually added, to define every target or company/institution. Two alternative ways can be used: an automated process, when the system generates all possible vocabulary combinations of words, or a manual process, when the development team registers only the keywords of interest, gaining more speed for the machine learning processes. When the crawler's network detects some of these words in the news' content, the achieved data are passed to the next step, where several filtered functions and a sentiment analysis approach [13] are applied to remove ads and to set the polarity level.

The third component includes other important functionalities, such as favorite lists or saved for later read lists, which are a part of the mobile user interface component based on users' requirements, making the experience more attractive. In this field it is very difficult to distinguish a standard set because the mobile industry is evolving fast, giving the applications' authors free imagination finding and developing new components.

This component is implemented using the Ionic Framework language, making the developed platform available for both mobile operating systems Android and iOS.

It is not mandatory to use the three modules sequentially. When we have made changes to web application module, the bots network is updating, the machine learning system is changing, but the application core and user interface modules remain unchangeable and accessible.

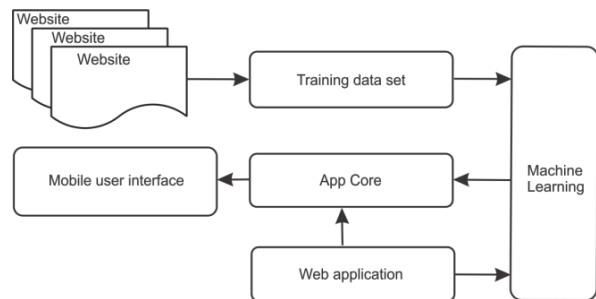


Figure 2: The architecture of the application

USER INTERFACE

We consider that the evolution of human beings in the last few years has been influenced by mobile technologies. Usage rates for those aged from 25 to 65, which are the main target group for this application, increased every year. The target population usually has varying characteristics and necessities, depending on cultural level or the capacity of the people to understand certain concepts. We have to develop elements for the mobile platform and place items or icons in ways that the user can take some desired actions or to interpret some information from graphical

representations. Furthermore, we have to generate information that is easy to access and quickly to share. These are several aspects of the platform that transform the mobile user interface in the most important component of our system.

The developed application has the following features: viewing and selecting news for public targets or companies, selecting the negative, fake or ironic news, searching through news, filtering by company, sharing news, sorting news, adding to the favorite lists, editing users' information or viewing the graphical representation for a selected period. The user must be authenticated to access these features. He/she can require a new account and the credentials are analyzed by the development team.



Figure 3: (a) Start page. (b) Login page.

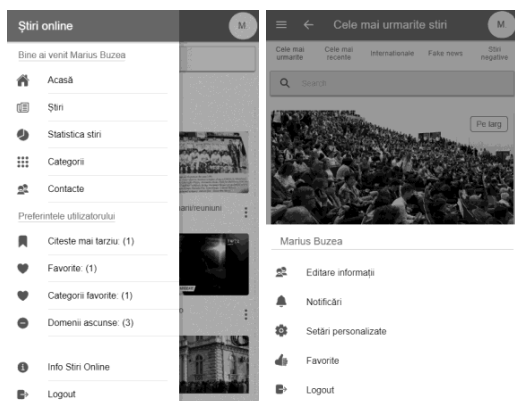


Figure 4: (a) Users' preferences. (b) Users' settings.

Figure 3 (a) presents the start page and Figure 3 (b) the login page, where the user has to choose between username and email address to get access in the developed platform. If the user forgets the password or the account was hacked, he/she can reset the password by pressing the button "Reseteaza parola" (eng., "Password reset"). After this action is completed, the user receives a link through email address to change the password.

After the registered user is logged, the recent stories are displayed with several options in the top of the page, giving

to user the opportunity to select different categories. Also, the settings icon is positioned in the top left corner, as shown in Figure 4 (a) and (b). By pressing this button, the navigation menu is expanded. The user finds shortcuts to the main facilities of the platform, such as news of the day (Figure 5 a), news categories (Figure 5 b), statistics, graphical representation of the selected news, topics, useful contacts, information about this application and the most important aspect, the users' preferences category.



Figure 5: (a) News list. (b) News categories.

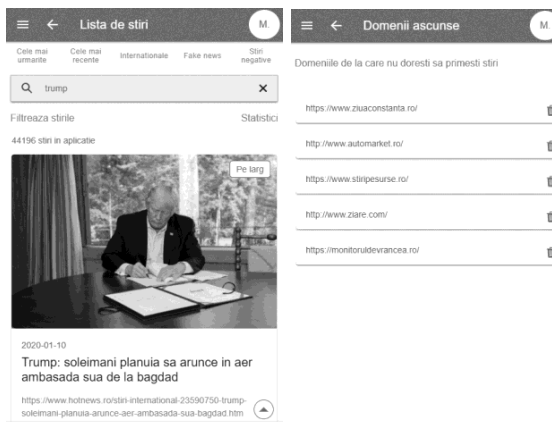


Figure 6: (a) Search option. (b) Hidden domains.

The idea of designing and developing this application came from the necessity of a better connection between users and the online environment, in order to prevent the spread of fake news and misleading stories. The users' requirements category was implemented to ban the online broadcasters of false information or online deepfakes without proper disclosures. This category contains items, such as those saved for read later, favorite lists or hidden domain lists (Figure 6 b), giving the user the opportunity to refute certain domains and receive reliable information. Moreover, by creating the hidden domain lists, the application receives the necessary feedback, thus the programmers are able to

improve or to be more carefully in the selection process for the specified broadcasters.

The search facility (Figure 6 a) of the application allows the user to find a specific story from the lists, providing further details such as image, description and more option (Figure 7 a) with link through the extended story or to the original source.

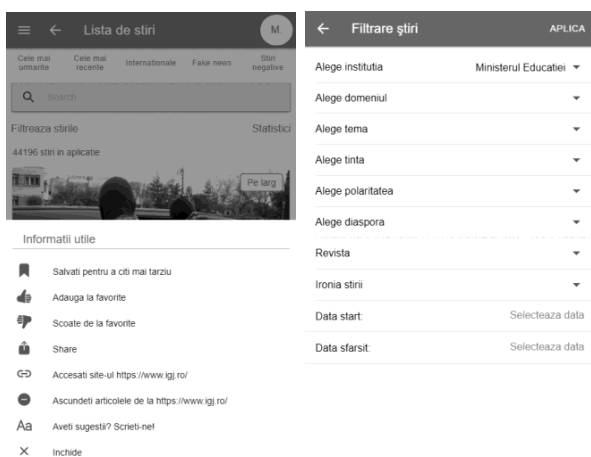


Figure 7: (a) More options. (b) Filter options.

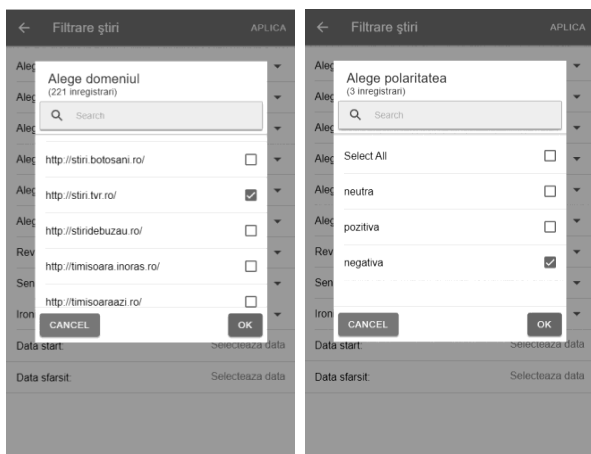


Figure 8: (a) Domain options. (b) Polarities options.

One of the most important facilities of the designed platform is the filter module. The data from this module were manually designed by the management team, choosing relevant keywords for every institution or public target (ex. “ministru educatie” belonging to “Ministerul Educatiei si Cercetarii”). After having added the words in application, the system automatically detects all possible combinations of words, such as “ministrul educatiei”, “ministrilor educatiei”, “ministrului educatiei”, etc. Moreover, these are registered into the database to achieve a better targeting process. As illustrated in Figure 7 (b), the platform has several filtering options, such as topic, target or company/institution which have associated different

domains (Figure 8 a). Furthermore, to gain speed and have a better search experience, we initially loaded only 15 domains, so that the user has to scroll vertically to load more items. Furthermore, the application offers another important feature, the polarity filter, the user having the opportunity to select articles with negative or positive polarity, depending on his/here option (Figure 8 b).

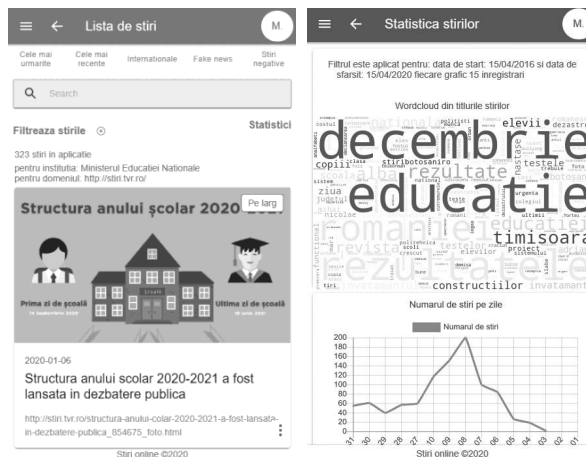


Figure 9: (a) Filter applied. (b) Wordcloud generated by the most used words.

To view the filtered news, the user should pass through options and finish the selection by pressing the button named “Aplica” (eng., Apply) located at the top right corner of the filter page (Figure 8 b). By pressing this button, a page with users’ interest articles is shown, giving more information about the selected items too, as shown in Figure 9 (a).

The user has another option in the right side of the filter page named “Statistic” (eng., Statistics). By clicking this section, a new page is opened with multiple statistical graphical representations for the selected data.

As illustrated in Figure 9 (b), the application generates a wordcloud image from the most used words of the day using python scripts, through application core model. Viewing this image, the user should understand the main topic of the day or the targets involved. Another image representing the number of news distributed on consecutive days is displayed in the same page giving to user the opportunity to browse through different periods.

To select a large time period the user must scroll vertically, loading and displaying more data in the same graphical representation.

Visualizing the total percent of negative, positive or neutral polarity for the daily news is another advantage of this application (Figure 10 a). Figure 10 (b) presents a bar chart containing the numbers of negative and positive news distributed on consecutive days representing a significant detail in discovering and analyzing those days with more

negative news. Furthermore, from the polarity challenge perspective, the user has the opportunity to focus on the impact of negative messages spread in the virtual environment, by selecting and analyzing those online broadcasters with more critical news articles.



Figure 10: (a) Negative and positive polarities percent. (b) Number of news/day

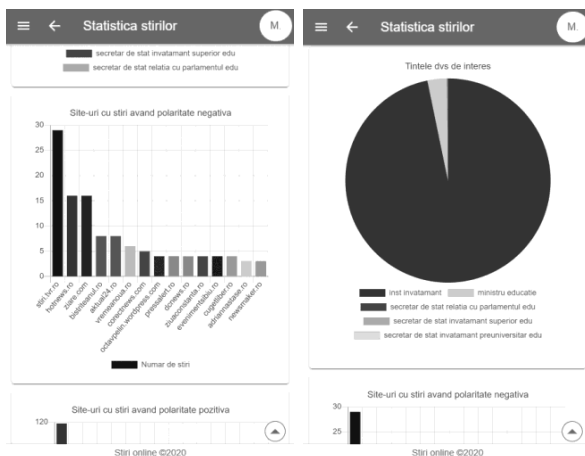


Figure 11: (a) Numbers of negative news/website. (b) Targets percents

Figure 11 (a) shows a chart describing the number of negative news, sorted in descendent order, for those online broadcasters whose most articles are with negative polarity.

As illustrated in Figure 11 (b), the application also generates a pie chart with the most used public targets. To discover these entities the crawler system applies a similar process with the one used by the management team in choosing keywords for companies/institutions.

Another aspect of the platform is represented by the different categories of news, such as international/national category or ironic/non ironic category, as shown in Figure 12 (a).

We tried to maintain similar features for the web and mobile application within the design process, but additional aspects of each application highlight the important differences. Each system has obvious advantages and disadvantages, such as speedup or small screens, but the most noticeable aspect of the mobile platform are the real time notifications. This module was implemented using One Signal notification library of the free JavaScript open source named Node.js and, beside Firebase Console we can push notifications to users with no cost, as shown in Figure 12 b.

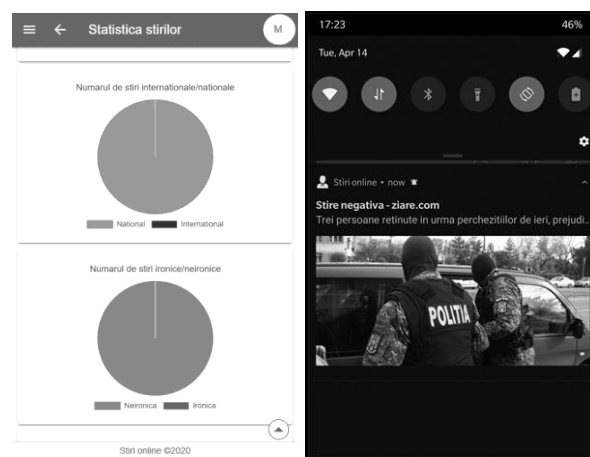


Figure 12: (a) News percent for different categories. (b) Push notification example

CONCLUSIONS

The most important online platforms, such as Facebook³ and Google⁴, announced in 2017 their commitment to use “trust indicators” to help users to select the reliable news of the online broadcasters and prevent spreading of the fake news. Even if these indicators were applied, this year, in the context of coronavirus pandemics, the misinformation circulate through social media applications, including closed stores, unproven methods for disease cure or military helicopters used to spray disinfectants against virus. Thus, the Romanian authorities had taken several measures in order to avoid spreading of false information, launching

³ <https://www.facebook.com/facebookmedia/blog/launching-new-trust-indicators-from-the-trust-project-for-news-on-facebook>, last accessed on 5th July 2020

⁴ <https://www.blog.google/topics/journalism-news/sorting-through-information-help-trust-project/>, last accessed on 5th July 2020

official online platforms⁵ and closing other online domains⁶.

Providing and sharing alternative explanations to fake news with logical meaning and scientific examples have been more effective than the above solutions, the main condition consisting of being immediately informed. Thus, the most important aspect of designing and developing this platform is that the registered user can receive notifications in real time about misinformation from the online environment regarding a public target or company/institution. Also, by having access to this application the user has the opportunity to analyze and visualize various statistical graphical representations.

Our main goal was to offer reliable content for readers and to increase the ability of the used models to detect and support registered users in their activities by receiving information about a specific target, company, institution or topic.

Another aspect of this application was to provide simple pages and well understood elements for displaying information in a structured way, helping to have an intuitive navigation and improve the user interaction and experience.

Nowadays, more online broadcasters are trying to use artificial intelligence approaches in the extraction and selection process of news, designing and developing crawler network.

There are several applications with Romanian content which present to users the latest news from the online environment for different topics, but with no archive or irony and fake detection features.

Evaluating an application is a complex task in the overall development process and we specially focus on creating the best user interface design and machine learning processes, but there is at least one indicator that can be reliable used in this platform, such as Firebase Console, which plays an important role for the future modules.

REFERENCES

1. E. Castledine, M. Wheeler, M. Eftos, “*Build Mobile Websites and Apps for Smart Devices*”, USA, SitePoint, 2011.
2. P. Dominik, “*Mobile UI Design Patterns A Deeper Look At The Hottest Apps Today*”, USA, UXPin, 2014.
3. T. McCann, “*The Art of the App Store - The Business of Apple Development*”, Indianapolis, Indiana, USA, John Wiley & Sons, Inc., 2011.
4. R. Sandberg, M. Rollins, “*The Business of Android Apps Development - Making and Marketing Apps that Succeed on Google Play, Amazon App Store and More*”, USA, Apress, 2013.
5. J. Iversen, M. Eierman, “*Learning Mobile App Development A Hands-on Guide to Building Apps with iOS and Android*”, USA, Addison-Wesley Professional, 2013.
6. L. Sloan, A. Quan-Haase, “*The SAGE Handbook of Social Media Research Methods*”, USA, Sage Publications Ltd, 2017.
7. M. Qiu, W. Dai, K. Gai, “*Mobile Applications development with android Technologies and Algorithms*”, New York, USA, CRC Press, 2017.
8. J. Lumsden, “*Handbook of Research on User Interface Design and Evaluation for Mobile Technology*”, New York, USA, Information Science Reference, 2008.
9. B. Bähr, “*Prototyping of User Interfaces for Mobile Applications*”, Berlin, Germany, Springer, 2017.
10. E. Alepis, M. Virvou, “*Object-Oriented User Interfaces for Personalized Mobile Learning*”, Warsaw, Poland, Springer, 2014.
11. C. Banga, J. Weinhold, “*Essential Mobile Interaction Design Perfecting Interface Design in Mobile Apps*”, USA, Addison-Wesley Professional, 2014.
12. S. Hooper, E. Berkman, “*Designing Mobile Interfaces*”, Canada, O’Reilly Media, 2011.
13. M.C. Buzea, Ş. Trăuşan-Matu, T. Rebedea, “*A Three Word-Level Approach Used in Machine Learning for Romanian Sentiment Analysis*”, 18th RoEduNet, 2019.
14. M.C. Buzea, Ş. Trăuşan-Matu, T. Rebedea, “*Automatic Irony Detection for Romanian Online News*”, 24th International Conference on System Theory, Control and Computing, ICSTCC 2020.

⁵ <https://datelazi.ro/> and <https://stirioficiale.ro/>, last accessed on 5th July 2020

⁶ https://www.ancom.ro/en/decizii-decret-stare-de-urgenta_6253, last accessed on 5th July 2020