

Development of Information System for Aggregation and Ranking of News Taking into Account the User Needs

Vasyl Andrunyk^{[0000-0003-0697-7384]1}, Andrii Vasevych^{[0000-0003-4338-107X]2}, Liliya Chyrun^{[0000-0003-4040-7588]3}, Nadija Chernovol^{[0000-0001-9921-9077]4}, Nataliya Antonyuk^{[0000-0002-6297-0737]5}, Aleksandr Gozhyj^{[0000-0002-3517-580X]6}, Victor Gozhyj⁷, Irina Kalinina^{[0000-0001-8359-2045]8}, Maksym Korobchynskyi^{[0000-0001-8049-4730]9}

¹⁻⁴Lviv Polytechnic National University, Lviv, Ukraine

⁵Ivan Franko National University of Lviv, Lviv, Ukraine

⁵University of Opole, Opole, Poland

⁶⁻⁸Petro Mohyla Black Sea National University, Nikolaev, Ukraine

⁹Military-Diplomatic Academy named after Eugene Bereznyak, Kyiv, Ukraine

vasyl.a.andrunyk@lpnu.ua¹, andriy.vasevych@gmail.com²,
lchirun21@gmail.com³, nadija.m.chernovol@lpnu.ua⁴,
nantonyk@yahoo.com⁵, alex.gozhyj@gmail.com⁶, gozhyi.v@gmail.com⁷,
irina.kalinina@gmail.com⁸, maks_kor@ukr.net⁹

Abstract. The purpose of the work is to develop an intelligent information system that is designed for aggregation and ranking of news taking into account the needs of the user. During the work, the following tasks are set:

- 1) Analyze the online market for mass media and the needs of readers; the purpose of their searches and moments is not enough to find the news.
- 2) To construct a conceptual model of the information aggression system and ranking of news that would enable presentation of the work of the future intellectual information system, to show its structure.
- 3) To select the methods and means for implementation of the intellectual information system.
- 4) Design an online resource for aggregation and ranking of news, news feeds and flexible settings, a list of available sources of information, compliance with specified media and personal aggregation results.

Object of research is processes of aggregation of news and intelligent ranking of necessary news according to the needs of the user.

Subject of research is methods and means of aggregation and ranking of news and building an information system that implements them.

Practical value of work is to develop an intellectual information system for aggregation and ranking news according to the needs of the user.

There are many similar open and paid systems. Realized specifically for the public or for one or another media. Many of those information systems do not have a wide-ranging functionality for flexible news feed set-ups, which is mostly often demanded by regular users. This system will have a functional enough to solve this problem and offer a multifunctional service to users. The system will be implemented as a free software product under an open distribution license, and this approach will ensure its further development.

Copyright © 2020 for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Keywords. News, News Aggregation, Content Ranking, User Needs, Intelligent System, Content Analysis, Data Mining, Context Filtering, Bayesian Clustering, Bayesian Networks

1 Introduction

We live in the age of information - the time of unrestricted access to information resources, the time when the amount of information published by various sites, news feeds and other sources increases exponentially [1-5]. Every day, thousands of electronic newspapers publish tens of thousands of articles on various topics. Each of them may be potentially interesting to a particular reader. An overabundance of news sources creates a situation in which a person can spend more time searching for news of interest to himself than when reading this news. A significant number of news aggregation algorithms are designed to overcome the excess information, allowing a person to read immediately what is interesting. Like information retrieval systems, news aggregation systems allow the user to find the information he needs. Since the needs of an individual user can vary significantly in different people, the system of aggregation of news should be tailored to a particular person. The main distinguishing feature of media aggregators from other sources of information is the unique ability to provide the most up-to-date information, regardless of the time of day and its volume.

Now, it's easy to find out what happened at the other end of the globe in a matter of minutes or even seconds [6-9]. An entirely new space has emerged that destroys all possible boundaries [10-14]. This unique opportunity gives us modern means of communication, means of information transmission, including radio, television, telephony, e-mail and the global Internet network with its practically unlimited possibilities [15-21]. Meanwhile, along with an unprecedentedly large potential for informing the society, there were the same opportunities in scope for its misinformation. The news aggregator solves the problem of fake publications in the media, in particular, makes it possible to filter out unreliable news sources [22-26]. The Internet combines visual, audio, print and video representation of data and provides any necessary information at any level of users interested in it. It involves dialogue, feedback, and not a monologue that is typical of print media, radio and television [26-34].

2 Analytical review of literary and other sources

2.1 The basic principles of the study

Electronic newspapers have long been an integral part of network content. Thousands of users do not go down to a kiosk or to a mailbox for a fresh press, and choose from their bookmarks Ukr.net, Tsn.ua, Pravda.com.ua, Radiosvoboda.org or other sites that work as Internet media. Internet as a media has no equal, the specifics and advantages of electronic publications are, of course, determined by the wide possibilities of the network [35-41].

Technological advantages. First, it is the efficiency of providing information. In our time, information is becoming obsolete very quickly, and the Internet allows you to develop the maximum speed of access to it. There are many different opportunities for widespread use of on-line streams (i.e., real-time information transfer). These are various broadcasts, news feeds and more [42-45].

At the same time, the website has no restrictions on the amount of information and the time of the show. Unlike television or radio broadcasting, the website may provide fresh information in its entirety at any time. Any network editions can update articles as many times a day as they like and give out a news feed in real time. On the other hand, the Internet allows you to accumulate and store information. That is to create archives of any depth. Instead of describing a problem in detail and its development, the author can simply put a hyperlink to the source material, thereby saving space.

There are masses of possible technical solutions in the e-newspaper: the search function, the provision of information in any form (text, picture, graphic, audio file, video file), push-channels and others.

As the number of people who use the Internet services is growing, the importance and significance of network media also increases. People are increasingly attracted by the fact that the Internet media has more independence, as opposed to the classical types of mass media: television and print media. In addition, it bribes the convenience and speed of providing information: any user, spending a minimum of time and energy has the opportunity to find out everything that interests him.

New media, and in particular news sites, are definitely different from traditional ones and have their own features, the main ones being multimedia, interactivity and hypertextualization. In the book "Media Convergence and Multimedia Journalism", M. Lukin gives the following definition of multimedia: this is "the characteristic of providing information through various media platforms - verbal text, photographs, audio, video, animations and other forms derived from them."

In traditional media, there was always a delimitation: newspapers and magazines supplied it in plain text (after the appearance of colour printing, more visual information was added to it as design and photos), radio - in audio, and television - in audiovisual. Now, if you look at the articles on the Internet, you can see more photos or entire galleries, videos, and infofiles in addition to the text. This combination of formats is possible thanks to the Internet, and is multimedia. In addition, due to hyperlinks and news collections, it is really possible to trace the development of events from and to. Hypertextuality is another no less important attribute of news sites. By definition, hypertext is "a text consisting of a potentially infinite number of texts combined by a system of built-in hyperlinks, which allows you to read it not only horizontally, but also through the internal and external references," in-depth ". The use of hypertext in journalism gives the materials a third dimension - depth.

It is noteworthy that hyperlinks can be found in the text itself, so close to it, for example, on the side. So on the website of the RBC news agency; if you open one or another article, you can see a small window on the left with the inscription "Link on the topic", while on the site of another agency, RIA Novosti, hyperlinks are located inside the material. In addition, as already mentioned above, the hyperlinks are external and internal. External leads to other sites, while internal ones allow the user to

navigate the same resource. Another important component of new media is interactivity, which allows users of the news site to leave comments under the materials placed on it, send e-mails to the editorial staff, participate in surveys and engage in other activities. This way, users can communicate with each other and contact the author of the Internet publication or the editorial staff of the Internet publication, and, of course, express their opinions. Thus, media engagement with its audience has become more accessible than before. If before the emergence of interactive Internet resources people had to write letters to the editor or call to express their opinions or attract the attention of the media to a particular problem, then it's just enough to just go online and leave a comment on the site[46-49].

The advantages of the Internet compared to television. On the Internet, unlike television, the time for the transfer of information is unlimited. ABC broadcasts news for several hours a day, and you can meet them on the ABCNEWS.com website for 24 hours a day. Information on webpages may differ by another valuable quality, which is not able to afford either television or print newspapers: the depth of disclosure of the topic. For example, the article about global warming is limited to two minutes on television or 500-1000 characters in a newspaper. In an electronic version, the article on the same subject can be as large as possible, accompanied by photographs and charts, sound interviews with environmentalists and specialists, video recording of the consequences of cataclysms. The site will also be given a table of references to institutions dealing with the causes of global warming [50-54].

Thanks to the Internet, television and broadcast companies can now offer information to a specific narrow circle of people without tired of it by a mass audience. For example, a natural disaster destroyed the city. On TV in CNN there is no time to list all the dead and wounded. However, on the web page, such information may appear which, in fact, was made in 2017, when the tornado swept over the state of Arkansas. Some companies went even further. Yes, MSNBC.com is offering its users an opportunity to rate the articles they read on the site to find out which of them is most popular with readers. The result is a completely interactive tool of information. In 2017, ABCnews.com hosted an average of 100-150 thousand a day. Answers to such surveys.

Professionals recognize the Internet as well. Alexander Tkachenko, General Director of the 1 + 1 telekalan believes that "the Internet and related new media technologies have fundamentally provided new opportunities for the publication of materials prepared by journalists on the Internet. As a rule, not only newspapers, magazines, news agencies, but also TV companies and radio stations have their own Internet representation. Visitors to sites of certain mass media are available as current editions, as well as archives of newspaper and magazine articles, information and analytical television and radio programs. This allows you to search and play the necessary information at a user-friendly time, regardless of the broadcast grid, from programs geared towards different time zones.

"Active Media Group" experts state that Internet media has become an effective media platform for advertising, as the audience's loyalty to them is increasing more

and more. Experts refer to TNS data, according to which the audience's loyalty to the Internet in 2017 amounted to 81%, besides television (75%) and radio (70%).

"The Internet has expanded opportunities for advertising. On television, the viewer immediately recognizes that before him the advertisement and may interrupt the contact with the media carrier. While on the Internet there are many non-standard ways of placing target-based contextual advertising that does not cause user annoyance. However, many advertisers are trying to save money and use standard Internet advertising such as regular banners," says Daria Leiderman, Media Division Manager, Active Media Group. Activists of Media Group also argue that the high level of loyalty of the audience to the Internet is determined by the hypertext of this source. "Users are involved in creating content online with professionals, so information becomes more reliable. In addition, due to the hypertext of the Internet, there is the possibility of revisiting information on several resources, which increases the level of trust in this media. In this regard, Internet advertising posted on similar sites may be effective. At the same time, advertisers are in no hurry to abandon television advertising, as the brand of the TV channel is of great significance", says Darya Leiderman.

Thus, the loyalty of the audience to the Internet media increases due to the possibility of re-examination of information on several resources. This factor attracts the attention of advertisers who regard the Internet as an effective channel for the distribution of advertising.

2.2 Place online media on the World Wide Web

Convergent newsroom in Germany. According to Internet World Stats statistics, as of October 2017, the number of Internet users in Germany was about 67.4 million people, which is equal to 82.7% of all residents of the country. According to the study, in March 2017, 76% of Germans regularly entered the network, with 47% using the mobile Internet. According to comScore for October 2017, the number of Internet users in Germany was the largest in Europe. On average, they spent 24.5 hours online per month and viewed approximately 2,710 pages during that time. According to a study by Langzeitstudie Massenkommunikation [10], as of 2016, Germans aged 14 years on average spend 220 minutes a day on television, 187 on the radio and 86 on the Internet. At the same time, these indicators vary greatly depending on the age of the audience. So the younger generation between the ages of 14 and 29 prefers the Internet: 136 minutes a day, 130 against television and 127 on the radio. Despite the fact that in general the development of the Internet and news Internet resources in Germany shows a positive trend, the existence and development of a converged news audience in Germany is quite difficult due to the peculiarities of the media sector. It is known that for listening to the radio and watching television in the country charged a subscription fee and for listening to the radio on the mobile phone charged money. Printed products, of course, is also not free.

In the 2000s, Internet information resources began to be actively developed in Germany, and editorial offices began publishing their own websites and online versions of newspapers. Despite the fact that television as it was and still remains the

most common news source in Germany, the online media audience has been steadily increasing. In contrast, print publications gradually lost their subscribers and readers.

As a result, many of them were forced to close or fully switch to the network. One of the most prominent events was the closing of the print version of the Financial Times, which took place in mid-December 2012. Since then in Germany, you can only read the Internet version of the publication. The reason for the closure was that the circulation of the newspaper in the period from 2001 to 2012 fell from 23.7 million copies to 18.4. The subscription to the newspaper fell from 62 thousand in 2006 to 42 thousand in 2012. Revenues from advertising also declined at times.

At a time when the media editors began to explore the Internet, their online versions were free. In this connection, the outflow of visitors from the traditional media, especially the printed press, is quite understandable: instead of paying for news content, it could be freely available on the network. However, since the profit from advertising on the Internet, despite its constant growth (10 years ago in 2008, Internet advertising revenue exceeded the advertising revenue on the radio), the lack of ability to cover the cost of online editorial attracted entering paid subscriptions. The policy of many online media has changed: users have to subscribe to a resource if they want to get full access to the content. Usually small news notes are posted on the site in open access, and for the opportunity to view all the materials of the site you need to pay. There are three types of subscriptions: electronic and printed versions, and the electronic version for mobile phones. The development of mobile applications is one of the main tasks for many online editions. So WirtschaftsWoche, puts a priority on the development of iOS - their editions. The subscription fee for the WirtschaftsWoche electronic edition will be lower than its iOS version. In addition, the tablet application is more technologically advanced. Thus, the situation with the Internet media has developed as follows: full online versions of print media and online TV viewing are usually paid, and access to content is made by subscription. Mobile versions and gadget apps are also paid. In this regard, the search for news on Internet sources is extremely difficult: finding detailed information about an event or analyst in open access is not so easy. Nevertheless, it is leaked even from closed sources, and can be found through search engines (Google's most popular search engine in Germany). For the audience of the Internet media in the country is characterized by a review of news, selected by news aggregators, and the subscription to electronic editions and their versions for mobile devices goes to the background. Since subscriptions are usually not expensive, one person reads no more than 2-3 resources. Of course, the technical support of the online media, their design and the ability to interact with the audience through interactivity are very high, because it is their account holds the audience. But they can not compete with news aggregators. If you take the ranking of 50 most popular sites in Germany as of July 2017, compiled by Google, then in the first place it is itself with an indicator of 38.6 million unique visitors, which make up 74.7% of the total number of Internet users in the country. The news sites of the news agency bild.de (10 million unique visitors, accounting for 19.3% of the total number of Internet users) and the magazine Spiegel.de (6.2 million unique users, which is 12% of the total number of Internet users) rank the thirteenth and twenty-first line of the ranking. In total in the rating of one hundred, most popular sites of the country included five

news resources. Because news aggregators allow users to watch the news, without having to go to the site of their sources, online resources lose a large amount of potential profit from advertising. This problem has recently been discussed at the state level. In early March 2013, a law was passed in Germany on the protection of the work of journalists and photographers, according to which all search engines and news aggregators will be forced to pay "license fees" for posting full texts of articles on their sites. Publishers who publish online versions of their newspapers and magazines will determine the cost of one line of material. At the same time, journalists will be paid their share. Search engines and news aggregators, according to the new law, will be allowed to publish on their portals only individual words from articles and small text fragments. Thus, the new law completely changed the situation on the network: the search for news on the Internet with the help of aggregators and search engines for German users will be practically impossible, because Google strongly disagrees with the introduction of the new law. It is likely that the online newsletter model that exists now will change. News sites will receive new unique users who previously read news on aggregator sites. It is also possible that links to social networks will increase, as the Germans have data on the most active users of social networks in Europe, and social networks to some extent help to select the content. It is possible that the number of subscribers to online publications will not increase through mobile devices as it will be the only way to get high quality and detailed news information online. Accordingly, it can be assumed that the emerging newsgroup in Germany will be transformed and, by its behaviour, approach those who learn information from traditional media. Thanks to the new law, much of the Internet's ability to search for information and news will be suppressed, and the development of media platform as well as Internet technologies in general will be slowed down. In my opinion, such a decision rejects Germany back, as the mass media of other countries at this time will, on the contrary, actively develop on the network. The vacuum in which electronic media will appear in Germany, and the limited choices available to users of the Internet, may even serve as their outflow to traditional media. It is unlikely that this will be a printed press, but television, which is the most common source of news, can get new viewers.

Convergent newsroom in the United States. In the United States, 94% of the population use the Internet, according to the research The Generations Online, conducted by Pew Research Centre Internet & American Life Project. Among users of the network, people aged 18 to 44 are dominant, while the number of users per year has increased in all age groups. Max's minimum growth was recorded in older age groups: so the number of people aged 70 to 75 who use the Internet has increased from 26 to 45% per year, and the number of people aged 76 years and over who use the Internet and the network, has grown from about 17% to 27%.

At the same time in the United States, as well as in Germany, television, especially local TV channels, is the main source of news, and the Internet is in second place. According to a study by Audience Segments in a Changing News Environment conducted by Pew Research Centre with 3612 people in the US, the audience of print media and radio is steadily declining, while the online media audience is showing

significant growth. It is also possible to note the general tendency to reduce the audience of television, despite the fact that in some years it did not significantly increase.

Together with the redistribution of sources of information from which people learn news, the time spent on their review has changed. Thus, in the period from 2010 to 2017, according to the North American Technographic, it fell the most: from newspapers by 26%, and by magazines by 18%. Radio also surrendered its positions, and now it is listening to 15% less than before. Television on the contrary showed a good result: in 2017, it began to look at 5% more than in 2010. The time that people spend on the Internet has increased by a colossal figure of 181% and, like television, is about 13 hours a week. Accordingly, it is the television and the Internet that have become leaders in time, which are ready to spend on their viewing. The second place is radio, which Americans are ready to spend twice less time: 6 hours a week. Reading newspapers and magazines, they spend less than three hours a week. The Pew Research Centre for The People & The Press, based on its study, identified four groups of news consumers in the United States. The first and the main of them - Integrators, which is 23% of the total news audience of the country. It includes those who receive information from traditional and online media. Unlike those who prefer only traditional media, they are better educated, successful and involved in social processes. Also, People - Press analysts led by Andrew Kohut identified three other groups in the typology of news users. So in the Net - Newsers group, there are 13% of the adult population, representing a younger generation of people than Integrators. As a source of news information, they rely on the Internet more than on traditional media such as television, newspapers and radio. They are more secure and even better educated than Integrators. According to statistics, about 8 out of 10 representatives of the Net-Newsers group at least graduated from college. They not only basically rely on the Internet as a source of news, but actively use its new features and technologies.

Traditionalists make up 46% of the population. These are more adult, less educated, and less well-equipped news content consumers who choose television as the main source. They have one characteristic feature that distinguishes this group from Integrators and Net - Newsers: looking at photos and videos telling them about events in the country, they learn more about the information than when they read it.

The last group is Disengaged - people who are little interested in news and do not always read them. It is 14% of the total news audience. Integrators and Net - Newsers are those people who move away from the traditional media. Instead of watching the newscast on certain hours on the TV or reading the newspaper they find in their porch in the morning, these groups watch the news online on the Internet when they want more, they spend so much time on the net that "stumble" on news even when they check email or engage in any other internet activity.

Net - Newsers is a particularly interesting group, because for them, the use of news on the Internet is an integral part of life. They are happy to comment on news, watch news on recommended sites such as Digg, Reddit or NewsTrust, read blogs, visit YouTube, and sometimes even post their own news articles online. Integrators are lagging behind Net - Newsers in this regard, but the most "advanced" of them gradually begin to use interactive and all the capabilities of the network, due to which their

behavior is approaching Net - Newsers. At the same time, representatives of Traditionalists who also retreat from television replenish the Integrators group.

Of the total number of Internet users who read news from the Internet at least three times a week, equal to 37%, most of them (55%) do it every day, with men interested in news more than women (59% vs. 51 %). It is also characteristic that the age group from 30 to 49 years is interested in getting the most up-to-date information from the Internet. The level of education also affects the interest in news: the higher it is, the higher the interest. In order to find news, Americans are using a variety of channels and technologies: from social networks to news generated through mobile gadgets.

In addition, 22% of Americans have customized home-based webpages, such as iGoogle, with an inalienable attribute of news. Among those who read news on the Internet every day, namely news embedded in a home page, their number is 44% of people. One third of all Americans and half of those entering the network say they regularly or occasionally watch news programs or video clips on the Internet. In addition, every tenth user of social networks receives news information from there. In addition, 35% of Americans are daily looking for news on the Internet. The most popular online resources from which users find information is Yahoo (28%), MSN (18%) and CNN (17%). The New York Times, The Wall Street Journal, USA Today and The Washington Post are popular with online editions. It is noteworthy that access to most of the content in them, as well as in Germany, is carried out by subscription. True, on the sites of American editions, there are full articles, not notes, and subscriptions are only two types: electronic (distributed to the site and all mobile applications) and combined: printed version and electronic. If a subscription is signed in Germany for a year, then in the United States it is usually a month (like The Wall Street Journal) or three (as in The New York Times). Searching for news 83% of Americans use search engines, with 31% of their number resorting to search engines at least three times a week. Characteristically, young people with higher education are more likely to be looking for news online than older people or graduates from schools and colleges. When looking for news, 51% of users do not go to unknown sites. In addition, news sources can be subscribed by e-mail and RSS feeds. Significantly, 64% of Internet users over the age of 25 trace news, moving on links to them, without going to the main page of the resource. The category of people aged 25 to 29 years old does the opposite: 48% come first on the main pages of news sites. Most Americans track local and national news and international only when there are important events. 57% of the audience of news sites always keep track of local news regardless of whether something is important or not. In addition, 55% of Americans are tracking national news. 34% of the US residents are interested in international news, and in the period of important events, their number is up to 56%. Comments on news published on news resources are read by every fifth, but only 7% of users are ready to comment on the entries. A total of 75% of Americans have never left comments on sites. If we consider segmentation of the audience in groups, then you can see that Integrators and Net - Newsers are the most similar in their characteristics. These are people with a high level of education, earnings and the ability to get to the network at work. It is on these factors that the most important thing is the gap with other audiences. Integrators and Net-Newsers are also younger than Traditionalists are. They actively use the op-

opportunities of news online resources: 16% Integrators and 30% Net - Newsters watch the news online, 10% and 19% of these groups listen to the radio online and 19% and 26% of them read blogs. In other groups, these values range from 1% to 3%. By the degree of technical accountability, these groups are almost identical, and are undoubtedly leaders. Thus, the features inherent in these two groups and their behaviour describe the converged audience in the United States, which, by virtue of its demographic characteristics, is even less numerous but more promising than the conservative part of the news consumers.

Convergent newsroom in the UK. According to Internet World Stats, about 52.7 million people use the Internet in the UK, accounting for 84.1% of the country's population. If we consider the structure of the audience by age group, then more than 98% of people in age groups from 16 to 24 and 25 to 34 go to the network. People between the ages of 35 and 44 lag behind them by 2%: their number is 96%. In addition, the generation from the age of 45 to 54 is not too lagging behind: 90% of people in this group use the Internet. A group of people between the ages of 55 and 64 is also nearby, with 81% of the people using it on the Internet. In other age groups, indicators are significantly lower. Overall, you can see that older Britons use the Internet to a lesser degree than those who are younger. However, the number of users in all age groups increases year by year. It should be noted that in all age groups up to 64 years, the percentage of men and women using the Internet is practically the same (difference ranges from 0.2 to 4%) while in groups of people over 65, men prevail. According to the Office for National Statistics survey conducted in 2016, 39 million Britons are getting into the network every day. According to ONS data, 51% of Britons use mobile gadgets to access the network. 87% of them are people aged 16 to 24, and 81% are Britons aged 25 to 34. The older the people, the less they use the mobile Internet. It is noteworthy that 83% of people over 45 years of age use the mobile Internet, going online over the Wi-Fi during the day. 21% of them use a tablet for these purposes.

According to a 2016 survey by Direct Marketing Association UK, 64% of Britons use the mobile Internet to read news. If we consider the British's attitude to the Internet as a news source, then it is second only to television as a whole in the country, and to the first in people under the age of 34, according to a study by McKinsey and Company. The main sources of online news are electronic versions of newspapers and TV channels, while viewing videos and releases of television broadcasts online is a priority source of information. Significantly, the most popular site in the country is Google UK, the seventh line of BBC Online, and the thirteenth dailymail.co.uk, sites of news agencies and radio are not very popular. It's also worth noting that Google is the most sought-after search engine for news searches, with 93% of its users coming to its service, according to a study from UK Digital Future in Focus, conducted by comScor. Blogs are the least authoritative source of news, and the UK trusts them less than other European countries (6% vs. the average 10%). News activity in social networks at the same time is developed quite high, and a certain part of the transitions to news sites is precisely from them. It is also worth noting that more than 70% of the news in the UK is open access, and therefore finding material about an event is not difficult. The culture of subscriptions to online publications, including electronic ver-

sions of newspapers, is poorly developed due to the large amount of information that is widely available. The subscription usually takes a week, with both electronic and mobile gadgets (each with its own apps). In addition, with the electronic version you can subscribe to the printed one.

When comparing the news audience in Germany, the United States and the United Kingdom, you can see that in all countries, the main source of information is television, and the Internet is in second place. At the same time, in the younger generation, the worldwide network is the main provider of news content. At the same time, older age groups of audiences prefer television and other traditional media. In addition, in all countries, the consumption of news from the Internet, in particular through mobile Internet, is increasing, while traditional media outlets are losing their audience. You can see that the audience of news online resources in all three countries is very similar: these people have a higher level of education and income compared to other audiences. They also go to the network several times a day: from home, office, and sometimes from mobile gadgets, whose penetration among online news users is quite high.

The main content provider for them is news aggregators and search engines, with Google being the undisputed leader in all countries. While reading and retrieving news, social networks play an important role, among which Facebook is the most popular. There are mixed attitudes to blogs in different countries: in the US, more than 20% of the audience reads blogs, in particular political ones, while in Germany this percentage varies around 10%, while in the UK the blogosphere is on the last place in popularity among news sources. It is also worth noting that in the United States, as in the UK, the important role is played by news videos from which network users receive information. With regard to online versions of print media, you can say the following: the freer users of the network can search for information and receive it from other sources, the less you need an online subscription. In addition, its duration depends on it: in Germany, where there is a tendency to increase the volume of paid Internet content and limiting the work of news aggregators, users are offered to subscribe for a year. At the same time in the USA there is a similar tendency, but on a smaller scale, and the offer is offered for a period of one month to three. However, because advertising revenue, despite their stable growth, does not bring the expected profit, many publications are still free, also have to enter a subscription. In the UK, users are more loyal to advertising, and due to this, the profit from it is greater than in other countries, so the question of paid subscription is not so acute. There are news sites with partially free access in these countries and in their content: so in Germany, you can see mainly news notes, and in Great Britain, full materials. Among the main topics of online news, which are interesting to users, are local news, political and public. In Germany and the USA, more men than women watch news from the Internet, and in the UK their numbers are virtually identical. At the same time, the gap between men and women who receive information from the Internet is significantly lower than the gender gap in the traditional media audience, due to the higher level of education in the first category as a whole. Local news, political and public are highlighted. In Germany and the USA, more men than women watch news from the Internet, and in the UK their numbers are virtually identical. At the same time, the gap between men and women who receive information from the Internet is significantly

lower than the gender gap in the traditional media audience, due to the higher level of education in the first category as a whole. Local news, political and public are highlighted. In Germany and the USA, more men than women watch news from the Internet, and in the UK, their numbers are virtually identical. At the same time, the gap between men and women who receive information from the Internet is significantly lower than the gender gap in the traditional media audience, due to the higher level of education in the first category as a whole. Thus, the portrait of a converged audience in foreign countries and its behavioural features are very similar, despite some differences due to the peculiarities of the development of news resources on the Internet.

2.3 An analysis of known problem solving tools

During the work the analysis of modern intelligent information systems responsible for news aggregation was carried out, their functionality, advantages and disadvantages were determined. The most common news aggregation systems are the following: www.ukr.net, www.smi2.ua, ww.smi.today, <https://www.rbc.ua>.

Information site www.ukr.net. UKR.NET is the most popular Ukrainian Internet news service. Every day in the news stream on the portal UKR.NET comes about thirty thousand news from almost one hundred thousand sources from all over Ukraine. Among them are popular and authoritative news sites, international and Ukrainian news agencies, online media versions, regional online resources. News are collected every minute and are placed in the relevant thematic sections "Politics", "Economics", "Events", "Technology", "Science", "Show business", "Auto", "Society", "Sport", "Healthy I", « Abroad », « Interesting », « Photoreport », « Video », news of every region of Ukraine. In this case, in the news feed on the main page of UKR.NET algorithm, the most recent and up-to-date news is generated. (Fig. 1-2).

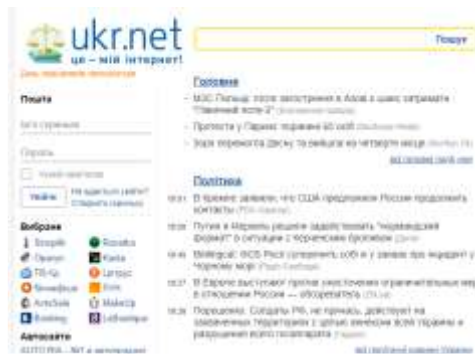


Fig. 1. The main page of Ukr.net

Advantages:

- availability of regional news feeds;

- UKR.NET portal does not publish news, but only controls automatic heading and clustering;
- News are grouped in groups;
- availability of own e-mail service;
- availability of mobile version.

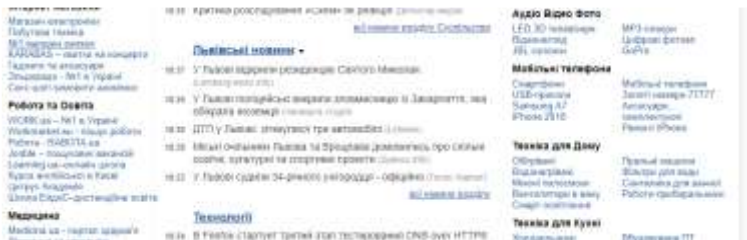


Fig. 2. Block of regional news of the site Ukr.net

Disadvantages:

- the lack of tweets of the news feed;
- there is no news search;
- meager functional.

Information site www.smi2.ua. In its present form, SMI2.ua is a content aggregator of the media, in which news in real time automatically rank in the ranking, depending on its popularity in the audience, using the data exchange news network SMI2.net, which is part of the same company. In addition to the two general ratings (per hour and per day), there are similar thematic ones, where the news on economics and finances, cars, sports and real estate is collected according to the same principle. There are regional ratings - cities of Ukraine and the USA. By clicking on the header of this or that news or the "read more" button, you can get the full text of this news on the site of the corresponding edition. There is also an opportunity to share news in the main social networks.

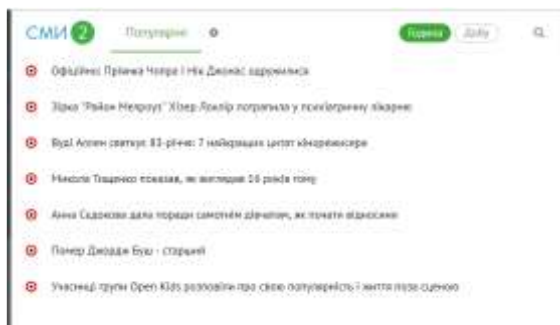


Fig. 3. The news line www.Smi2.ua

Advantages:

- nice appearance of the interface;
- possibility to search in news;
- the news is divided into categories;
- news aggregates by automated algorithms;
- searchable by the text of the publication;
- availability of a currency exchange widget;
- the ability to change the font size;
- comfortable adaptive design;

Disadvantages:

- impossibility of flexible adjustment of the news feed;
- not enough information sources;
- absence of user registration
- lack of news headlines.

Information portal <https://www.smi.today>. SMI.TODAY is a free automatic news aggregator of Ukraine and other countries. The project was created to reduce the time to search and analyze news, as well as to accelerate the opportunity to receive only fresh news from verified sources on one page. Every day, more than 50,000 people from around the world visit this project, and more than 100 popular media outlets have posted positive reviews on this site. The aggregator newsletter smi.pp.ua is updated every minute, with the developed intelligent algorithm distributing news in sections, which allows you to keep abreast of the most important events 24 hours a day and 7 days a week. The news aggregator accepts only proven resources, as the editorial department verifies all editions.

Only new materials appear on the project's main page. For the convenience of using the resource, each user can register on the site and view it news in a convenient form, as well as add your favourite news and sources to bookmarks. In addition, each registered user can add any bookmarks news to read news later.

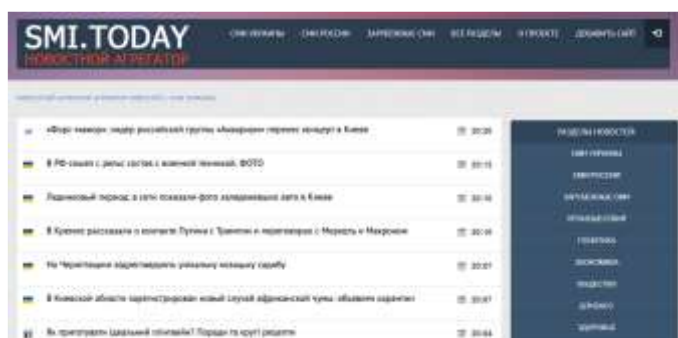


Fig. 4. <https://www.smi.today>

Advantages:

- possibility of registration of the user;
- only checked resources are accepted into the news aggregator;
- an opportunity to offer your own news site;
- a good description and a large number of sections;

Disadvantages:

- outdated graphical interface;
- lack of Ukrainian localization;
- there is no adaptive design;
- meager functional.

Information portal www.rbc.ua. RBC-Ukraine is a Ukrainian information portal that covers important events in Ukraine and in the world. Specializes in financial, economic and political news. Included in the top three most visited news sites.

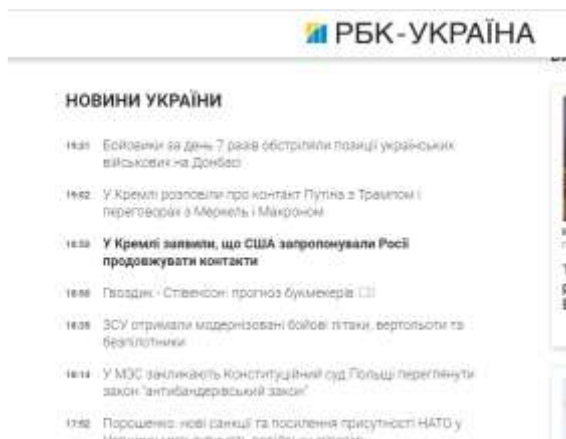


Fig. 5. Information Portal <https://www.rbc.ua>

Advantages:

- the interface is clear to the user;
- there is a possibility to search the site;
- convenient user interface.

Disadvantages:

- meager functional;
- impossibility of flexible adjustment;
- news only from resources belonging to the holding of RBC;

Having analyzed the online news aggregators who publish their own news feeds, one can conclude that the disadvantages are present in many resources, and in some of their more than advantages, these services provide only the basic functionality for a specific task.

3 System analysis and problem solving

3.1 System analysis of object of research and subject area

First, you should define the concept of the system, and then system analysis and system design. The system is a set of detailed methods, procedures and procedures, established or formulated for the implementation of specific activities, duties or problem solving. System analysis is a process that collects and interprets the facts, identifies problems, and splits the system into its components. The purpose is to study the system or its parts to determine its goals for improving the system and ensuring that all components of the system work effectively in accordance with their goals. In short, System Analysis determines what the system should do. System Design is a process of planning a new system or replacing an existing system by defining its components or modules to meet specific requirements. Before planning, one needs to deeply understand the old system in order to determine how modern computers will be best used for efficient functioning. In short, Systems Design focuses on how to achieve system goals.

Tree goals. To achieve the desired result within a certain time interval, it is necessary to decompose the subsystem or system. In this case, you can decompose the system by designing a tree for it, which describes the problem as a whole and will serve as the basis for the structural data used to solve the problem. "Tree of Goals" - provides a graphical representation of the subordination and interrelation of goals, representing the distribution of the general (general) goal or sub-mission, individual actions and tasks. Using the "target tree" defines the "target framework" of the information system, organization or phenomenon. The general rule for constructing a "tree of goals" is the completeness of the reduction. Completeness of reduction - a process of bringing complex phenomena, systems or process to the simplest components.

To implement this rule (completeness of reduction) the following elements of the system approach are used:

- the goal of the higher level is represented by a landmark - this is the basis for the development (decomposition) for the purposes of the lower level;
- lower-level goals are ways to achieve a goal that is level higher and should be presented so that their aggregate contributes to the achievement of the original goal.

Requirements for building a "target tree":

- completeness of the presentation of a phenomenon or process;

- at each level, the set of sub-targets should be sufficient to describe the goal of a higher level;
- consistency of goals that are at the upper and lower levels;
- the decomposition of the target at each level on a sub-target should be carried out using one methodological approach.

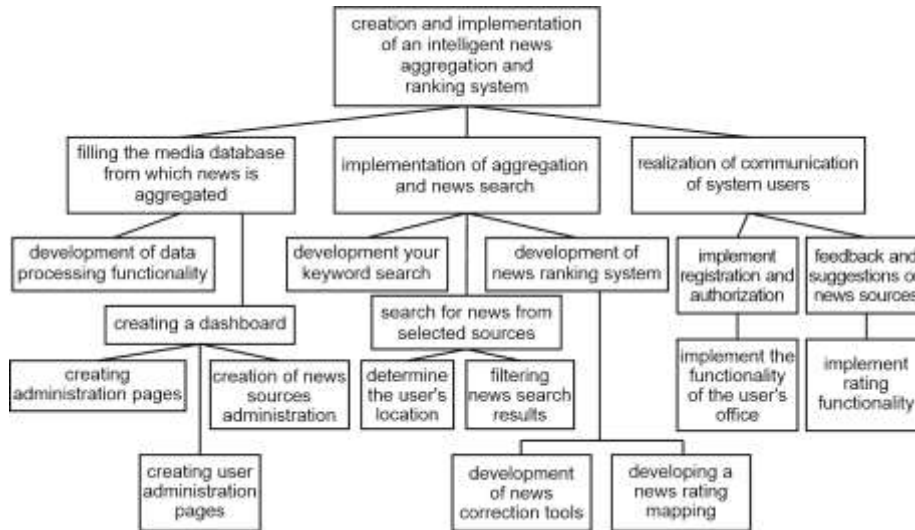


Fig. 6. Tree of Goals

In Fig. 6 shows the main goal - it creates and implements intelligent information system of aggregation and ranking of news taking into account the needs of the user. The main goal consists of three subsidiary branches with sub-targets. The first sub-target is the filling of the media base from which the news is aggregated by the intellectual information system. It is subdivided into a subtask to create a content fill panel, namely, "creating a administration panel" and "developing a data processing functionalities" that is used for input. The second goal itself consolidated the implementation of aggregation and news search, which includes such tasks as keyword search, news search from selected sources, and the development of a news system. To display regional news for a particular user, there are subtasks for defining its location and filtering the search results within a given radius. The system for ranking news can be attributed to the development of system correction tools and a segment of the ranking of news. The third sub-target for implementing the system as a full-fledged implementation of communication capabilities between users and the system is to this end, for the most part, the development of a graphical interface and a way to identify the user. Its underlying objectives include the implementation of registration and authorization, which includes the implementation of the functional cabinet user. In addition, this task includes reviews of news sources, which in turn include a subtask for the implementation of the function of the ranking of news.

Chart of options for use. The Usage Diagram is a dynamic or behavioural diagram in UML. The Usage Diagram simulates the functionality of the system using actors and cases of use. The use of cases is a set of actions, services and functions that the system must perform. In this context, the "system" is something that is being developed or exploited, such as a website. "Actors" are people or organizations that work under certain roles within the system. The general external entities of the system are:

- Administrator is a moderator who has the ability to add, edit and delete content as news sources as well as user moderation functions.
- User - this person is considered to be the end user of a resource that has direct access to the program, has the ability to manipulate the system, and such a person is able to receive the necessary useful information.

The main uses are: user registration, site search, login to the system, news feeds, moderation of the content of reviews, work with news sources.

In fig. 7 shows a diagram of usage options. Detailed research of Fig. 7, you can see, logging in should be user-registered and pass authorization procedure, both for the regular user and the administrator. Unlike a user, the administrator has the ability to add, edit and delete available news sources, and moderate feedback as well as perform all other manipulations that are available to the regular user. The user after entering the registration will be able to log in to the system and carry out the following manipulations with her: search for news, including the ability to search for a keyword, search by name, and filter the results; generate a news feed based on your own preferences, or if the user is not logged in, view the general news feed and view the ratings.

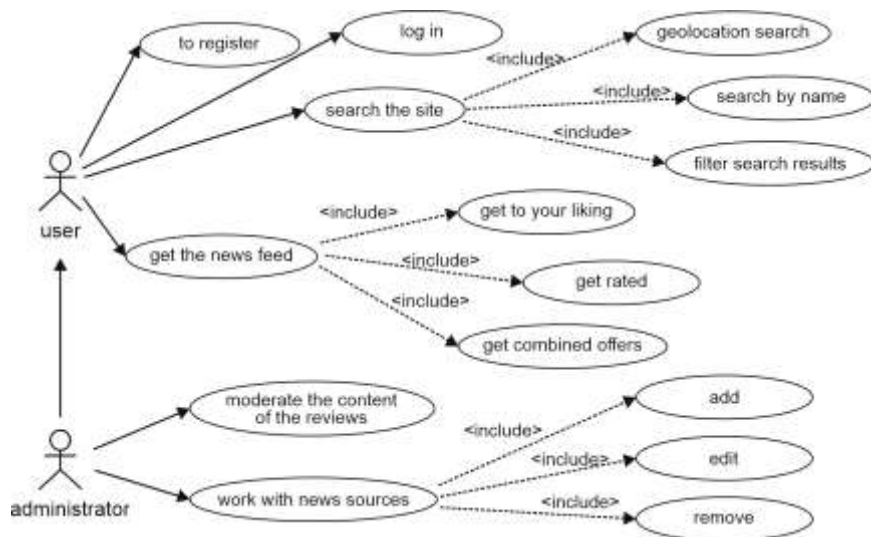


Fig. 7. Diagram of use options

Class Chart. In Fig. 8 shows a class diagram for the aggregation information system and ranking of news; it shows the following classes: t_map_user, Zmi, Marker, Comment, which are responsible for data storage;and UserViewSet, UserSerializerService, SearchViewSet, RecomendationService, MarkerSerializerService, PoinSerializer as auxiliary classes with data processing logic.

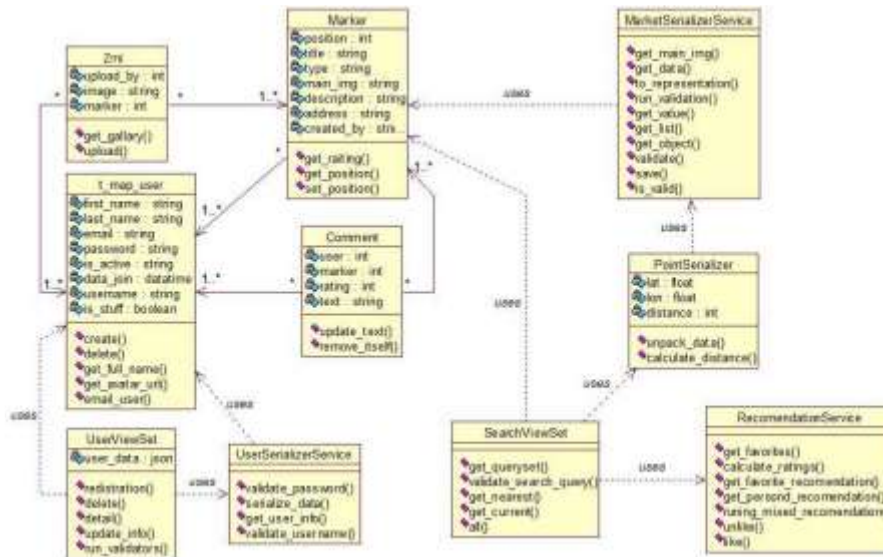


Fig. 8. Chart of classes

The class t_map_user - displays the user object with attributes: is_stuff type string, username type string, date_joined type date, is_active type bollean, email type string, last_name type string, first_name type string, last_login type string, password type string, int type int. The Zmi class represents the class responsible for representing the object in the database; it contains the following fields: upload_by - a link to the user's record, image, marker. And also the methods: upload (), get_gallery ().

Class Marker - represents the class that is responsible for representing an object in a database and has the following fields: position, title, type, main_img, descriptron, address, created_by - which is a reference to the user who created it. And also methods for obtaining position and rating and redefining the same position.

The Comment class represents the class that is responsible for representing the object in the database and stores the following fields: user, marker - user reference and marker, rating - the rating that the user has put.

Class UserViewSet - Represents a class controller that is responsible for the logic of working with a user object, processing requests and manipulating user data.

Class UserSerializerService - Represents the class responsible for serializing user data and contains some logic that should be displayed along with user data.

The SearchViewSet class represents the base class that is responsible for site search, processing of search queries, and issuing results; it uses the next service to give the user along with the search result recommendations.

The RecommendationService class represents a class service that includes the logic of getting a personal news feed that uses user classes and a marker to get the desired results. Class MarkerSerializerService - Represents a set of logic services that is used to represent the presentation of a marker class object, geolocation, and the closest markers on it. PoinSerializer class - represents the serialization class for the marker, getting all the necessary data to display it.

To enter the system, the system expects the entry of authorization data, after their input is checked for their correctness, if these data are entered incorrectly - the user remains a choice - to re-enter the data, or continue without authorization. Depending on whether the user authorized him to open additional features, consider, at first, the capabilities of the user who did not authorize:

- When choosing the "Search by word" activity, the user enters the search words in the appropriate input field, after which he receives the corresponding search results which, by means of the "Add additional filters" activity, can filter the result according to the specified criteria, after which the user receives the necessary news sample.
- If the user did not use the word search, then the "Search by location" option remains. In this activity, the user is offered the news that took place near him.

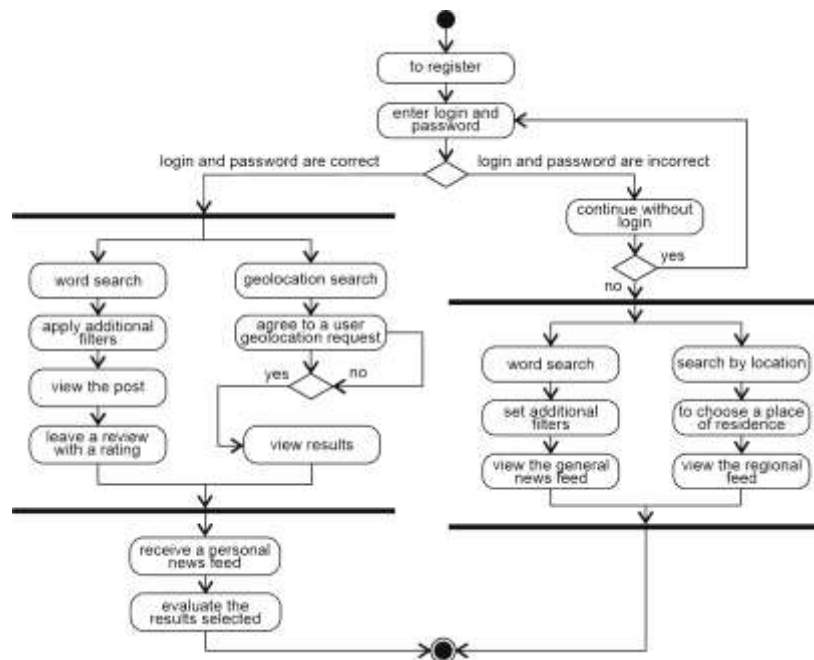


Fig. 9. Chart of activity

For an authorized user, additional features are opened in the above types of activities, then he will be able to:

- to use the "Word Search" activity after "Add additional filters" and "View Post", after which it will have access to additional "Leave Feedback with Evaluation" activity in which the user can leave feedback on the news or the source of news that will have a positive impact on quality of the generated news feed.
- after searching for news, the user is given the opportunity to use the "Get a personal news feed" activity in which the user will be selected a subset of the news collection, which will be based on his preferences and ratings.
- he will also be able to evaluate the results of the selection and exclude the current results from it, so that in the future similar are not included in the selection. This is provided in the "Evaluate Selected Results" activities.

After that the user has the choice to re-search and view the results, or just leave the page.

Sequence diagram in the context of UML is a collaborative object and is used to determine the sequence of events between objects for a particular result. A sequence diagram is an important component used in the analysis, design and documentation processes. A sequence diagram is also known as a synchronization chart, event diagram and event scenario. Object interactions usually begin at the top of the chart and end at the bottom. In the sequence diagram, the interaction of objects occurs through messages in the vertical and horizontal dimensions and are indicated by horizontal arrows and message names. The initial sequence diagram starts at the top and is located on the left side of the chart. The following messages are added directly to the previous messages. The sequence of messages can be divided by type based on functionality. The sequence diagram for the preceding "Obtaining aggregation results" is shown in Fig. 10. The sequence diagram depicts three objects: "User" is a user of a resource that has access to the program, after having made certain manipulations with the system, finds the necessary information; "System" is responsible for the processes, "Database" stores data reflecting the structural data of the system.

To get recommendations, the user enters the login information into the system; the system sends a request to the database to obtain user data, if any, after performing a password check, by hanging the password entered by the user and checking the hash with that received from the database. After successful identification, the system authorizes the user. After the user's permission, the user sends him the geolocation to his system. The system generates a search query for news that takes place within a radius from the user, and sends them to the system. Following this, the system sends additional requests for user's preferences, to combine the results and further generate offers. After that, the system processes the data stored and generates a new request for news for the user. The system sends this request to the database that returns the filtered results. After that, the System combines the ratings and results and sends them to the user. After receiving the news, the user will not be able to choose one of them, then he sends the corresponding request to the system, it processes it and adjusts further suggestions while keeping it not like the database.

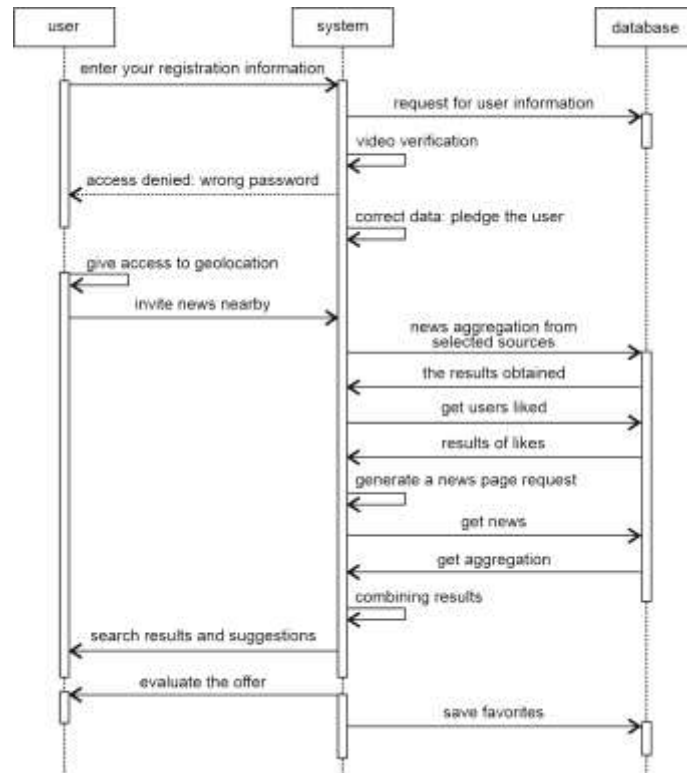


Fig. 10. Diagram of sequence for the precedent "Obtaining results of aggregation"

State Diagram is used to indicate the state of a system or its part in the event of a time event. This is a behavioural diagram, and it reflects behaviour using end referals. State diagrams are also called state machines and state diagram charts. These terms are often used interchangeably. So, simply, the state diagram is used to model the dynamic behaviour of the class in response to the time and change the external stimuli. We can say that each class has a state, but we do not model each class using state diagrams. It is preferable to simulate a state with three or more states. The state diagram is depicted in Fig. 11. The diagram shows that for the start of operation, the system is waiting for user input, that is, the introduction of authorization or registration data. Once the user has entered the data for registration, the system checks them, and if the data is entered incorrectly, then the system will wait for the input, but if the data is entered correctly - the user gets the opportunity to enter them again or continue without authorization. After that the user enters the search data on which the system was waiting, if the data were not entered the system goes to the end state. If the search data has been entered, the system processes and selects a news feed.

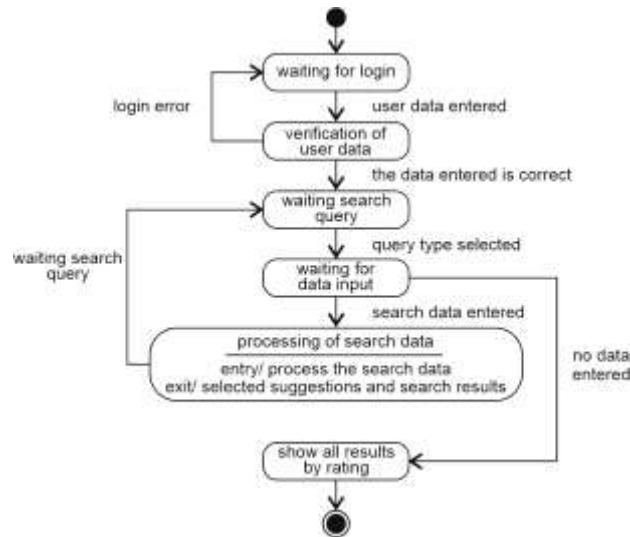


Fig. 11. State Diagram

Component Diagram. In fig. Figure 12 shows a diagram of the components in the system for organizing and arranging news, taking into account the needs of the user.

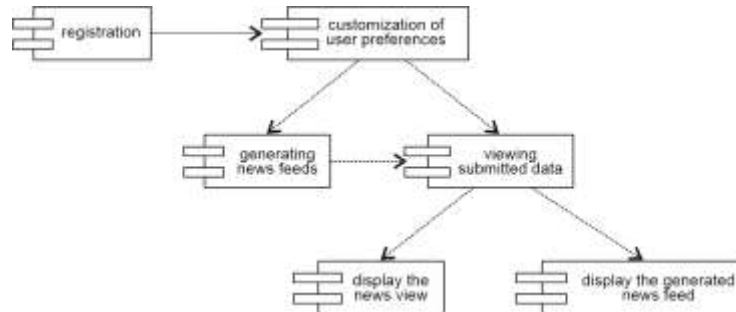


Fig. 12. Components Diagram

To design this system, the Design Model MVC (Model-View-Controller) for the backend of the Service and the Model View Model (MVVM) for the front part of the site was taken as the basis. Since the system itself is designed as distributed with a separate client part and server, it is divided into the following components:

- "Registration / atomization" - the backend part is responsible for processing the authorizing data of the user, and the front, for the possibility of entering these data by the user, and sending to the backend of the part;
- The "Generation of Ribbon New" component is responsible for preparing the news request and formatting it for the front part of the unit;

- "View» is the system component responsible for displaying data for the user.
- "Display of the ranking of news" - the component responsible for displaying the ranking of the individual news and the rating of the news source.
- "News Feed Display" is a component that displays a custom news feed based on user preferences.

3.2 Statement and substantiation of the problem

A conceptual model is a system that uses concepts and ideas to formulate a given presentation. Conceptual modelling is used in many industries, ranging from sciences to socio-economic theory to software development. Using the conceptual model to represent abstract ideas, it is important to distinguish conceptual model from conceptual model. That is, the model is actually a matter for itself, but this model also contains the notion that such a model represents - which model is, unlike what is a model. Without deep immersion in philosophy, recognizing these differences between the model itself and what it represents is crucial to understanding the proper use of conceptual models in the first place. Then, one should not be surprised that conceptual models are often used as an abstract representation of real-world objects. To develop a conceptual model for aggregation and ranking of news, it is necessary to set out the output and input data typical for the development of this system. Input - data coming from the outside of the system:

- user registration data - you must enter your name, last name, email address and password. The password must be confirmed, and validation is made on it;
- data for authorization - the user must enter the password and email address of the mail. To authorize in the admin panel, you must enter the same data, but the administrator;
- data for search - entered in a text format in the specified search field;
- news source - the administrator must add information about the source of news, its rating and reliability of this online media.
- review - the text information entered by the user to the selected news.

Output is data that is received by the user after processing by the system or other external entities. The source data includes: various types of user search results, in which there are keywords, a general news feed, and a rating for each publication. For authorized users, the opportunity to receive a personal news feed is another source information provided by the user system based on intelligent selection based on user preferences.

The purpose of the development is to develop an intelligent system that will be used to aggregate and rank news based on user preferences. This system is a web-resource that provides the following basic functions:

- searching for news on keywords;
- search for news about the location of the user;
- viewing and selecting the right news sources;

- reviewing the rating of publications;
- leaving feedback on a particular news or source;
- means of administering the aggregator of news;

To develop this web resource, client-server architecture is considered, it is considered one of the architectural software templates and carries an important concept for the development of network applications, and also provides for the exchange of data between them and interaction. It includes the following essential components: a set of servers, a set of clients and a network.

Appointment of the system. This resource is designed to aggregate and rank news based on the needs of the user, and in order to facilitate the choice of the user, the system will generate a personal news feed, but only for registered users. The user will be able to search for the receipt and search for news by keywords, put filters on the search results, receive news that is next to him, that is within a certain radius. To improve the results of the news feeds, the user will be able to leave feedback; this will be one of the factors influencing the subsequent selections of news publications.

Place of application of the system. The system will be useful for readers of the online media who do not want to spend time searching for the news they need and to hang out the pre-selected set. This system will help the user to find news on the keywords, filter the results by the specified type. A registered user is also able to post reviews about individual news. And based on his assessments to get a more interesting for him a selection of news.

Justification, development and implementation of the system. To date, Ukrainian Internet media have already had a significant and very interesting way of development. Publishers, through trial and error, accumulated experience; the market was strong and developed along with the improvement of new technologies. As a result, today we see high-tech multimedia media with great services and wide opportunities. In addition, one can safely assert that this is just the beginning. Currently, many media experts are seeing a great future in the online media market. Instead of the expected decay of print media, convergence occurs: many publications that want to develop and receive new interactive capabilities go online. There is also a reverse trend: convergent not only the Internet with print editions - has it also connected to television.

Expected effects from the implementation of the system. Like information retrieval systems, aggregation systems will allow the user to quickly find the information he needs. Since the needs of an individual user can vary significantly in different people, the system of aggregation of news should be tailored to a particular person.

The main distinguishing feature of the aggregator of the Internet media from other sources of information will be a unique opportunity to provide the most up-to-date information, regardless of the time of day and its volume.

4 Methods and means of solving the problem

4.1 Choice and justification of problem solving methods

Group and context filtering. Most approaches to solving the problem of information filtering can be divided into two main categories: Contextual (Content Based Filtering, CBF) and Groupware (Collaborative Filtering, CF).

CBF's approach is based on the assumption that news that is interesting to the user is similar to those that were interesting to him before. The CF approach, in turn, tries to find users similar to the one and then recommends to the user that information that seemed interesting to similar users. A large number of studies have recently been conducted to combine these two approaches. Such approaches are called hybrid.

Context Filtering. The CBF system deals with similarity calculations, between fresh news and user profiles. The most common and simple method in this category is keyword matching. Based on this simple method, systems like vector space model were developed that allow better filtering and searching of information.

Group filtering. The collaborative filtering task is to predict the benefits of elements for a particular user, based on the user preferences database of other users. Consider two types of collaborative filtering: memory-based and model-based. Memory-based algorithms operate over the entire database to create recommendations. Model-based, on the contrary, uses a database to study or customize a model, which is then used in the formulation of recommendations. Collaborative filtering systems often vary by feature: they operate by implicit or explicit means of expressing user interests. Explicit methods mean that the user deliberately describes his needs, usually based on a discrete integer scale. Implicit methods mean the interpretation of user behaviour or choice to determine preferences. Implicit ways of expressing interests can be based on viewing information (for Internet applications, for example), shopping history (for stores), or other templates for access to information. Despite the type of available preference data, collaborative filtering - algorithms need to cope with data deficiencies. Normally, we do not have a complete set of preferences for all item names. It can not be assumed that the absence of some element is a coincidence, since users are inclined to express preferences for those elements that were viewed by them, and therefore they are interested.

Memory-based algorithms. In general, the task of collaborative filtering is to predict the user's preferences based on the custom base of preferences. The database consists

of users, so a set of preferences (votes) v_i , which corresponds with the user i of element j . If I_i votes the set of elements on which user i has defined its estimates, then you can make an average rating User i as:

$$\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{ij}. \quad (1)$$

In memory-based collaborative filtering, the user's evaluation algorithms, which we denote as a , are predicted based on incomplete information about it and the set of weights calculated based on the user database. Assume that the predicted p_{aj} estimate by a user of element j is the weighted sum of ratings of other users:

$$p_{aj} = \bar{v}_a + \kappa \sum_{i=1}^n w(a,i)(v_{ij} - \bar{v}_i), \quad (2)$$

where n is the number of users in the *collaborative filtering* database with non-zero weights. Weights $w(a,i)$ can reflect the distance, correlation or similarity between each user i and the current (active) user a . Next, we will consider the details of various collaborative filtering algorithms that relate weighting. There are other possible characteristics of memory-based collaborative filtering, but in this work, we restrict ourselves to the wording described above.

Correlation. The general formulation of collaborative filtering statistical methods (as opposed to verbal or high-quality annotations) first appeared in the context of the GroupLens project, where the Pearson correlation was the basis for weighting. Correlation between users a and i is expressed as:

$$w(a,i) = \frac{\sum_j (v_{aj} - \bar{v}_a)(v_{ij} - \bar{v}_i)}{\sqrt{\sum_j (v_{aj} - \bar{v}_a)^2 \sum_j (v_{ij} - \bar{v}_i)^2}}, \quad (3)$$

where the sum of j is spent on the elements for which both users (a and i) have determined their estimates.

Similarity of vectors. In the area of information retrieval, the similarity between two documents is usually measured through a comparison with the word-frequency vector document and the calculation of the cosine of the angle between two vectors of frequencies. Can we use this formalism in the collaborative filtering task, where evaluations will play the role of the frequency of words? Note that by following this algorithm, measured ratings indicate positive feedback, and negative reviews are not counted, and invaluable items get a zero estimate. Accordingly, weights are expressed as

$$w(a, i) = \sum_j \frac{v_{aj}}{\sqrt{\sum_{k \in I_a} v_{ak}^2}} \frac{v_{ij}}{\sqrt{\sum_{k \in I_i} v_{ik}^2}}, \quad (4)$$

where the factors in the denominator serve to normalize ratings so that users who rate more actively than others will not be more like the others. Other schemes of normalization are also possible. The method can be supplemented by the "default estimation" scheme, which allows you to expand the set of user-evaluated elements. Another important addition may be the use of so-called inverse frequency estimates. When searching for text documents, the comparison of documents is based on the frequency vectors of individual words, with each word having a weight reflecting its specificity, so that the commonly used vocabulary has a lower priority. A similar pass can also be used in collaborative filtering by introducing a new user evaluation of a element j :

$$\bar{v}_{aj} = f_j v_{aj}, \text{ where } f_j = \log \frac{n}{n_j}.$$

Model-based methods. From the probabilistic point of view, the collaborative filtering task can be considered as calculating the mathematical expectation of the value of the estimate based on the available information about the user. For an active user, we want to anticipate ratings for items that have not yet been viewed. If we assume that the estimates are integers in the range from 0 to m , then we get:

$$p_{aj} = E(v_{aj}) = \sum_{i=0}^m \text{Prob}(v_{aj} = i | v_{ak}, k \in I_a) i, \quad (5)$$

where $\text{Prob}(v_{aj} = i | v_{ak}, k \in I_a)$ is the likelihood that the active user will evaluate the element j , precisely, for such a value, provided that there is an observation of the estimates made.

Bayesian clustering. Let C takes a small discrete set of values denoting clusters of users. We divide users into clusters, and we will consider their advantages through conditional probabilities:

$$\text{Prob}(C = c, v_1, \dots, v_n) = \text{Prob}(C = c) \prod_{i=1}^n \text{Prob}(v_i | C = c). \quad (6)$$

Bayesian networks. The method of learning the Bayesian network is to form such a network that each node of it is an element to be evaluated. Each node has a finite set of states - estimates of the corresponding element. In this model, the training algorithm of the bay network defines the best predictors for each element, such that it becomes possible to construct a decision tree, which, depending on the state of the root element, determines the high probability of the value of the sheet element.

In general, the Bayesian method, as well as the correlation method, work faster than others, respectively, sharing the primacy among different sets of data.

Algorithm for choosing news sources to show to the user. The number of sources complicates the task of finding news sources interesting to the user. In order for the algorithm given in the previous section to collect statistics for all agencies, it would have been necessary for decades. In order to overcome this problem, group-filtering algorithms were used in this paper. Accordingly, the best way with the task of group filtering news is to handle the Memory Based algorithm with weights equal to the correlation of the user vectors. In general, Memory Based methods are characterized by the possibility of effective implementation on the database, which is also the advantage of this algorithm for systems with a large number of users.

Consider the algorithm described by the following two formulas:

$$p_{aj} = \bar{v}_a + \kappa \sum_{i=1}^n w(a,i)(v_{ij} - \bar{v}_i), \quad (7)$$

$$w(a,i) = \frac{\sum_j (v_{aj} - \bar{v}_a)(v_{ij} - \bar{v}_i)}{\sqrt{\sum_j (v_{aj} - \bar{v}_a)^2 \sum_j (v_{ij} - \bar{v}_i)^2}}. \quad (8)$$

So, at the beginning of each custom session, we have v_a , the vector of explicit user ratings and the vector p_a , obtained using the group filtering algorithm. You must select k sources that will be shown to the user. All news sources from the point of view of this user are in one of the states shown in the diagram below:

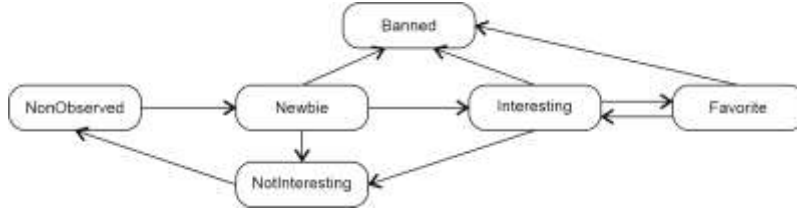


Fig. 13. All news sources from the point of view of this user are in one of the states shown

The source, about which the user does not know anything, is in the NonObserved state. The Newbie group has a fixed size and comes with NonObserved sources that have received the highest ranking by the group filtering algorithm. What got into the group Newbie, i equal v_a . In the state of Newbie channels are within 4 sessions after getting into it, in which they were displayed. This time should be enough for the user to evaluate the content. After the end of the 4 sessions, the source passes either to the interesting or to the NotInteresting, depending on the amount of interest to it from the user. In the state of Interesting are those channels that are of greatest interest to the user. From it the stream passes to the state of NotInteresting in the event of a fall in the value of interest in it. I will be banned and favourite will be issued because of explicit user actions (by pressing the buttons add to favourite and ban). In these states, the channel is located regardless of what the algorithms give it.

The display algorithm works as follows: preferred and interest channels are displayed whenever they have news. If there is not enough of these channels, channels are added in the state of Newbie.

4.2 3.2. Choice and justification of solutions to the problem

To develop the information system used:

- HTML hypertext markup language. It is used to format web pages when displayed in a user's browser;
- formal language for describing the appearance - Cascading Style Sheets CSS. Applies to stylistically designed web pages when displayed in a user's browser;
- language management script for viewing hypertext WEB pages on the user side of JavaScript;
- general scripting language, actively used to develop web applications PHP. This language generates dynamic pages on the website, which is important, and reduces the hard work.

As a means of development, the following methods were used:

- program for creating, editing, debugging, analyzing and publishing PHP web pages and applications phpDesigner;
- Adobe Photoshop CS Raster Graphics Editor 2014;
- free text editor with open source for Windows with syntax highlighting of a large number of programming and markup languages Notepad ++;
- distributions and software shells designed to create and set up Open Server sites.

HTML. Using HTML allows you to format documents for displaying those using fonts, lines and other graphic elements on any system that views them.

Most documents have standard items, such as headings, paragraphs, or lists. Using HTML tags you can mark these elements, providing WEB-browsers with minimal information for displaying data elements, while preserving the overall structure and information completeness of documents. Everything you need to read an HTML document is a WEB browser that interprets the HTML tags and reproduces a document on the screen in the form provided by the author.

In most cases, the author of the document strictly defines the appearance of the document. In the case of an HTML reader (based on the capabilities of the WEB browser, to some extent, can control the appearance of the document (but not its contents). HTML allows you to note where the document should have a header or paragraph using the HTML tag, and then provide the WEB browser For example, one WEB browser can recognize a paragraph start tag and submit a document in the right form, and the other does not have this capability and provides the document in one line. Users of some WEB-browsers also have the ability to customize the size and type of font, circle and other settings, changing the mapping document.

HTML tags can be conventionally divided into two categories about tags that:

- determine how the body of the document as a whole will be displayed in WEB-browsers
- describe the general properties of the document, such as the title or author of the document.

HTML documents can be created with any text editor or specialized HTML editor. The choice of the editor to be used to create HTML documents depends exclusively on the concept of the convenience and personal passions of each author. In the thesis work using HTML, we describe the structure of the document. HTML allows you to highlight separate logical parts in the text (headings, paragraphs, lists, etc.), put a photo or picture on a web page, and organize links on the page to communicate with other documents. This is the basis for the site, we create the structure of the HTML.

PHP is a widely used open source general-purpose script. Simply put, PHP is a programming language specially designed for writing web-based applications (scripts) that run on a web-server. The PHP abbreviation means "Hypertext Preprocessor (Hypertext Preprocessor)". PHP is quite easy to learn. The advantage of PHP is to allow the web developer to quickly create dynamic web pages. An important advantage of PHP language in languages such as Perl and C is the ability to create HTML documents using PHP commands. A significant difference between PHP and any code executed on the client side, such as JavaScript, is that PHP scripts are executed on the server side. The practical nature of the NRC is due to five important characteristics: traditionality; simplicity; efficiency; security; flexibility. PHP provides developers and administrators with flexible and efficient security features that are conventionally divided into two categories: system-level tools and application-level tools.

Due to the fact that PHP is an embedded language, it has an exceptional flexibility in relation to the needs of the developer. Although PHP is usually recommended to be used in conjunction with HTML, it integrates with JavaScript, WML, XML and other languages as successfully. In addition, well-structured PHP applications are easily expanded as needed (however, this applies to all major programming languages).

There are no problems with browsers, because before sending a client, PHP scripts are fully compiled on the server side. Essentially, PHP scripts can be transmitted from any device with a set browser, including cell phones, electronic notebooks, tablets and laptops, not to mention the traditional PC. With PHP there is a connection between the main blocks of the site with the database, its connection to the site.

JavaScript is the language of managing web pageviews on the user's side. To be more precise, then JavaScript is not only the programming language on the client side. Liveware, the parser of JavaScript, is a substitute tool on the side of the Netscape server. However, the most popular JavaScript has been programmed on the user's side. The main idea behind JavaScript is the ability to change the values of HTML container attributes and properties of the display environment in the process of viewing HTML pages by the user. This does not require reloading the page. In practice, this is manifested in the fact that you can, for example, change the background colour

of a page or an image integrated into a document, open a new window or issue a warning. The name "JavaScript" is a registered trademark of Sun Microsystems. Now, JavaScript completely takes up the niche of browser languages. The Java language, where the JavaScript name came from; like Java, JavaScript language is object-oriented, influenced the JavaScript syntax. However, on this their connection ends: Java and JavaScript are different languages; none of them is a subset of another. Language standardization was initiated by Netscape and implemented by the European Computer Manufacturers Association (ECMA) Association. The standard version is called ECMAScript and is described by the ECMA-262 standard (available online: English, Russian). As with any programming language, the main task of JavaScript is to create a sequence of actions that will lead to a certain result. Using JavaScript, we create a drop-down menu. Similarly, using JavaScript and cookies, we save user preferences by re-visiting and reloading the page.

CSS. Until the appearance of CSS, the execution of web pages was carried out solely by means of HTML, directly within the contents of the document. However, with the advent of CSS it became possible to fundamentally separate the content and presentation of the document. Due to this innovation, it became possible to easily apply a single style of design for a mass of similar documents, as well as a quick change in this design. Advantages:

- Multiple page designs for different devices. For example, the design will be wider on the screen, the menu will not be displayed while printing, and the design will be adaptive on the smartphone.
- Reduces the loading time of the pages of the site by transferring the rules of view data into a separate CSS file. In this case, the browser loads only the document structure and data stored on the page, and the view of these data is downloaded by the browser only once.
- Simplicity of further design change. You do not need to change each page, just change the CSS file.
- Additional design options. For example, using CSS-layout, you can make a block of text that the rest of the text will flow around (for example, for the menu) or make the menu always be visible when scrolling the page.

Disadvantages:

- Different display layouts in different browsers (especially outdated), which interpret differently the same CSS data.
- There is often a need to change not only one CSS file, but also HTML tags associated with CSS selectors, which sometimes nullifies the simplicity of using single file styles and significantly extends the time of editing and testing.
- Often, when page layout, you need to use the same value many times: one and the same color, the same font. And if this value needs to be changed, then you will have to change in many places ..

In dissertation work with CSS we create the style of a site: fonts, indentations, background, frame, links, buttons.

phpDesigner - A powerful program for creating, editing, debugging, analyzing and publishing web pages and applications in PHP language. The program is noteworthy that in addition to PHP, it supports HTML, MySQL, XML, CSS, JavaScript, VBScript, JAVA, C, Python and Ruby. Thanks to the tools for auto-highlighting the code, phpDesigner is also suitable for WAMP / LAMP- and AJAX-developers. In addition, the program contains capacious libraries that contain more than 3,000 functions that can easily be accessed during the programming process.

Basic features of phpDesigner:

- Syntax highlighting for many languages;
- Syntax check for PHP, HTML and CSS;
- Automatically complete code for PHP, HTML, CSS and JavaScript;
- Support for such JavaScript libraries as JQuery, Ext JS, YUI, Dojo, MooTools and Prototype;
- Built-in prompt system for PHP and JavaScript;
- Code inspector for HTML;
- Code browser for PHP, CSS and JavaScript;
- Transition to any PHP, CSS, and JavaScript announcement;
- Debugging and profiling PHP scripts using Xdebug;
- Support for FTP, SFTP and TortoiseSVN;
- Intuitive navigation on the objects of your code;
- Availability of built-in assistants;
- Ability to change skins.

phpDesigner is the main program for writing code.

Adobe Photoshop is a multifunctional graphic editor developed and distributed by Adobe Systems. Usually works with raster images. Photoshop supports the following colour schemes or ways to describe the colours of the image: RGB; LAB; CMYK; in grayscale; black and white; Duotone; with a 256-color palette (Indexed); Multichannel. Image processing is supported, with a colour depth of 8 bits (256 gradations per channel), 16 bits (15 bits plus one level, 32769 levels) and 32 bits (using floating-point single-number precision numbers). It is possible to save additional elements in the file, such as Guide, Channels (for example, Transparency Channel - Alpha channel), Clipping path, layers containing vector and text objects. The file may include colour profiles (ICCs), transfer functions (colour transfer functions). Non-Square Pixels (Pixel Aspect Ratio) are allowed. Features of Adobe Photoshop CS5:

- cross-platform support for 64-bit and accelerated GPUs;
- Trueedge technology to simplify the selection of complex objects;
- Content-awarefill function;
- creation of HDR images;

- Unmatched algorithm for processing Raw images;
- natural painting effects;
- Puppetwarp function;
- automatic adjustment of lenses;
- Adobe Photoshop CS5;
- quick correction of the output;
- automatic correction of lens distortion;
- Improved lens correction filter;
- auto-editing and image enhancement;
- Advanced layout.

Adobe Photoshop is used to create a logo and site buttons.

Notepad ++ is a free open source text editor for Windows with highlighting the syntax of a large number of programming languages and mark-up languages. It is based on the Scintilla component written on C ++ with the use of the STL, as well as the Win32 API and is distributed under the GPL license. The basic functionality of the program can be expanded both at the expense of plugins, and third-party modules, such as compilers and pre-processors of the syntax substrates. Basic features:

- Syntax highlighting;
- Autocomplete and auto-close brackets and tags (if activated);
- Bookmarks;
- Regular expressions for searching and replacing;
- Record and play macros;
- File comparison;
- Project manager;
- Document card;
- Redefining any hotkeys;
- Backup of stored files (included in the settings);
- Text transformation using the connected TextFX plugin;
- Support and conversion of ANSI, UTF-8 and UCS-2 encodings;
- Block selection of text, simultaneous selection of several different places (with Ctrl);
- Multi-line editing (using Alt). Notepad ++ is used to edit the code.

Denver - a set of distributions (Apache, PHP, MySQL, Perl, etc.) and a shell for the development of sites on a local computer running Windows without access to the Internet. Denver Basic Package:

- Apache web server with support for SSI, SSL, mod_rewrite, mod_php.
- PHP interpreter with support for GD, MySQL, SQLite.
- MySQL with transaction support (mysqld-max).
- A virtual host management system based on templates.
- Start and end control system.

- The phpMyAdmin panel for DBMS administration.
- Perl interpreter kernel without standard libraries (supplied separately).
- sendmail emulator and SMTP server with support for working with PHP, Perl, Parser, etc.
- The installer.

A web server is a server that accepts HTTP requests from a client's web browser and issues HTTP responses in the form of HTML pages, images, files, and other media streams or other data. Apache is a popular web HTTP server. The Apache configuration system is based on text configuration files. Has three conventionally configured levels:

- Server configuration (httpd.conf).
- Configure the virtual host (httpd.conf c version 2.2 extra / httpd-vhosts.conf).
- Configuring the directory level (.htaccess). SSI (Server Side Includes) is a language for dynamically "collecting" web pages on a server from individual components and issuing a received HTML document to a client, for use of instructions the file must end with .shtml, .stm or .shtm extensions.

PHP (Hypertext Preprocessor - "PHP: Hypertext Preprocessor", English Personal Home Page Tools - "Personal Web page creation tools") - a general-purpose scripting language used for developing web applications and creating dynamic web sites.

MySQL is a relational DBMS, which is a structured set of data, data is stored in separate tables, has links with each other, thus providing the ability to merge data from multiple tables when the request is executed.

PhpMyAdmin is a web-based interface for administering MySQL DBMS.

Denver (Denwer) is used to view the site and work with the database.

Client-server interaction architecture. Web applications are a type of program built on the client-server architecture. The client-server model is a program structure that distributes tasks and loads between resource providers and services, servers, and those who send a request i.e. a client. Essentially, clients and servers are software. As a rule, they are located on different computers and exchange data on a computer network using network protocols, but sometimes the client and the server can be on the same computer. The server host runs one or more server programs that distribute their resources between clients. The client asks for the content of the server, but does not transmit anything. Servers are waiting for requests, and customers initiate communication sessions with them. Customer requests are handled on a server - where the Database is located and the Database Management System (DBMS). This gives you the advantage of not having to send large volumes of data, and the query is optimized in such a way that it consumes a minimum amount of time. All this increases the system performance and reduces the waiting time for the result of the request. When performing queries, the server significantly increases the security of data, since data integrity rules are determined in the database on the server and are unique to all applications that use this database. Customer functions:

- Initialization of the server request
- Processing of the results of requests received from the server
- Representation of the results of the request to the user in a user interface form.

Server functions:

- receiving requests from the client
- processing requests
- execution of requests to the Databases and their optimization
- sending results of client requests
- providing security
- -providing stability to multi-user mode of operation

Productivity. Software development has evolved over the years, from manual testing, but in those times the requirements were much lower, the sites were text and downloaded in a few minutes. Therefore, the Web developer had much less incentive to pre-test. But rates have grown as ecommerce gains momentum in the world of web development. Therefore, testing began to be conducted in the development environment. But with the growth of applications began to automate and test. Developers began to write automated tests. In the end, testing has matured to such an extent that it has spread beyond the simple set of test modules and integration tests in the playback style. Organizations began to build increasingly sophisticated, thin test cases.

To date, the rates for applications have become higher than ever. The tests of the program have long been de facto standard. Since the applications became too complicated for manual testing, test frameworks were created to automate testing. And any good code starts with writing tests. But this does not let you know how the application behaves in the real environment. Testing the performance of web applications fixes it. Performance testing is a form of software testing that focuses on how the system works under a certain load. Performance testing should give organizations the diagnostic information they need to identify and troubleshoot. The slow work of apps affects paid subscribers, and new subscribers become less influential on revenue.

Most often, the problem of performance is very difficult due to the fact that it is difficult for developers to reproduce such "bugs". Performance issues do not directly affect the behaviour of the software. Rather, they are related to how the software reacts to the chaotic world of environments in which the application is launched. Therefore, it is necessary to conduct performance testing. Usual QA testing is to observe how the app handles one person. To produce the test, you need to simulate the harsh conditions, so you can detect how behaving the application under heavy load, so-called load testing. In the test environment, you can choose the load for the application, for example, the simultaneous use of the application by a thousand users in the normal operations and measure the behaviour of the program. Does it keep track of speed, slows down, or even drops? Of course, a thousand real people will not conduct such testing. To do this, software is created to help simulate the load.

In addition to stress testing, a stress test and endurance testing are performed.

5 Practical implementation

5.1 Description of the task realization

General Information. The developed software product is called "Intelligent Information System and News Ranking". The purpose of this product is to help users, based on their preferences, to compile a mix of news from different media, usually based on the criteria given automatically.

Functional purpose. The aggregate information system and news ranking allows the user to automatically scan news sites and aggregate algorithms to form a news feed. Since the needs of an individual user can vary significantly in different people, the system of aggregation of news adapts to a specific person.

Functional restrictions are imposed on users with outdated versions of browsers, since they do not support the latest standards used to develop this product.

Description of the logical structure. The database of the web service consists of 8 tables. This system was developed as distributed, where one part is responsible for displaying the data, the other for its processing, where the existence of the first without the second does not make sense. The bulk of the system being developed includes sub-modules for processing, shaping, validating data, and the logic of working with them. Thanks to well-chosen software implementations, software solutions have a high degree of declarative, which provides ease of understanding of executable code and simplifies the development of the system.

Call and download. Before running the program, you must run the executable web server locally or deploy it on a dedicated server with a static IP address. If we start the site locally, then we need to open a localhost with the port specified in the parameters in the browser. If the site is deployed on a dedicated server, we should open the server address or its domain name in the browser. The entry point in the program can be both the main page and the admin panel page.

The input data can include: search data; News sources; data for user registration; reviews. **Output data** for the system of automation of contextual advertising can be: generated news feed; News rating; user-wanted news.

5.2 4Instruction for the user

Introduction. The Aggregation and Ranking system works as a web page, and in order to take advantage of its capabilities, the user must have a pre-installed browser, which version is not younger than the version released in 2016, in order to go through

all the functionality. The web page requires permission to execute JavaScript scripts, since it is completely built on them.

General information about the program. The service consists of two parts:

- The component responsible for displaying and working with users, that is, the interface of the system, implemented under the web page;
- A component that acts as the heart of a system that implements the entire functionality, and can connect to other interfaces such as mobile applications.

Classes of tasks to be solved. The service solves the problem of aggregation and ranking of news based on user preferences. Provides flexible news feed setting.

Description of the main characteristics and features of the program. Using the aggregation service and ranking of news will save time in finding the right news, as the system analyzes the user's preferences and automatically generates a suitable news feed for him. The system also allows the user to flexibly configure aggregation, select news sources that he likes to read and type of publication.

Information on functional restrictions on use. The service will function correctly with browsers released in 2016 and later. Without permission to run JavaScript scripts, the service will not be able to function. Requirements for the technical characteristics of the device on which this service is used is the same as the requirements of the browser in this device. Necessary Internet connection.

5.3 Analysis of the control example

The web service has two types of roles: administrator and user.

The image shows a registration form with the following fields: 'Имя, фамилия' (Name, surname), 'E-mail' (Email), 'Пароль' (Password), 'Повторите пароль' (Repeat password), 'Номер телефона' (Phone number), 'Генерация пароля' (Generate password), 'Имя пользователя' (Username), 'Почта' (Email), 'Пароль' (Password), 'Повторите пароль' (Repeat password), and 'Мобильный телефон' (Mobile phone). There are also checkboxes for 'Мужчина' (Male) and 'Женщина' (Female) under the 'Имя пользователя' field.

Fig. 14. Registration form

Login to the web service is carried out by the login and password issued by the administrator of the service. The administrator's office provides the following functions: viewing, adding and editing existing news sources, moderating existing users. Configuring the aggregator news algorithm.



Fig. 15. Control panel



Fig. 16. Configure news sources in the admin panel



Fig. 17. Registered User News Feed

Керування джерелами

3 джерела з 1 неактивний

Ім'я	Діяльність	Ім'я джерела
Без обмежень	Будь-який	Фільтр новин
<input type="checkbox"/>	Активний	НАШІ ГРОШІ - Львів
<input type="checkbox"/>	Активний	ДК.ua - Львів
<input type="checkbox"/>	Активний	Львів формат

Fig. 18. Managing news sources

Створити нову стрічку

Привітне колекція надійних джерел, які ви хочете прочитати

Назва стрічки

Fig. 19. The form of the news feed

Початок сторінки

Моя стрічка новин

- НАШІ ГРОШІ - Львів
- ДК.ua - Львів
- Ця стрічка

Наші гроші - Новини за ключовими словами

Отримувати новини за ключовими словами

РЕЗУЛЬТАТИ

"Львів" - Google Новини

Ваші новини

Додаткової інформації про новини у відповідних новинах

- Новини про Львівський "Львів" за миття "Додаток"
- Львівський Львівський Львівський Львівський Львівський Львівський
- Новини Львівського Львівського Львівського Львівського Львівського Львівського

Fig. 20. The form of adding news by keyword

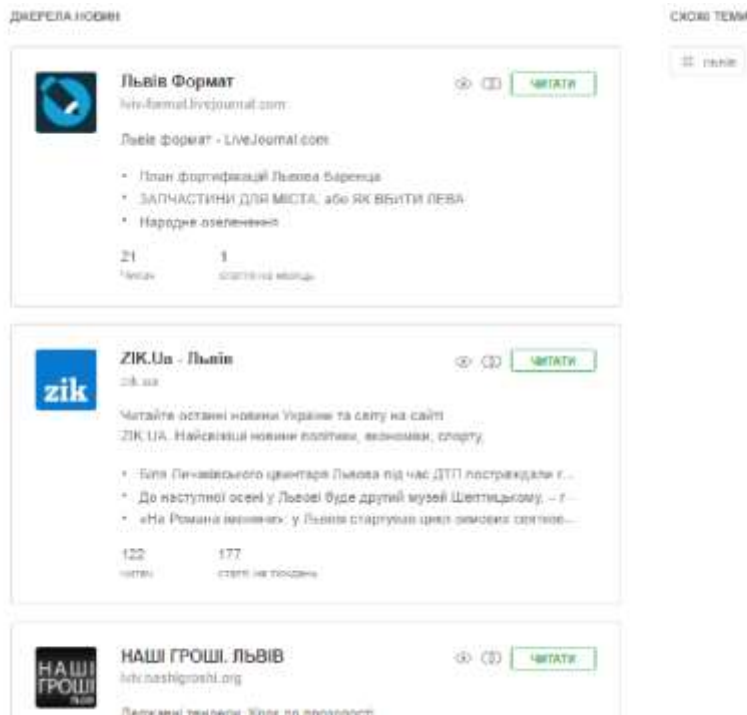


Fig. 21. List of available news sources

6 Conclusions

In this article a review of literary and information sources from the Internet was conducted. In general, the relevance of this topic was proved and the analysis of directions to which the information system evolves, depending on the market of mass media. A variety of resources has been explored, which provides valuable information that will be used in developing a system, such as analyzing the needs of a particular news reader and preferences according to user audiences. An analysis of such information systems was also carried out, and their weaknesses and advantages were investigated. The set of results obtained for the study of the issue of this system will be used in the detailed design and development of the whole.

In the work of the system analysis of the information aggregation system and ranking of news taking into account the needs of the user, an object approach was defined, which allowed to construct diagrams: components, states, sequences, activities, classes, usage options, and design a target tree. By establishing and justifying the expediency of creating this system, connections and necessary external entities are defined in order to achieve the desired results, as well as determined: the purpose of development, the purpose and place of application of the system, the development of a conceptual model (input and output data). In the article, the choice and justification of the methods for solving the problem was made

- for realization of this intellectual information system of aggregation and ranking
- news Selected and grounded list of various solutions to this problem
- problems among which: software (libraries, database extensions, frameworks, package managers), systems that significantly accelerate and facilitate development
- of this system, and in some cases, it is possible to solve all set for
- development of the task.

This article describes the key features of the system, described the creation of a software product according to GOST 19.402-78 "Description of the program", which describes such properties as general information, functional purpose, description of the logical structure, input data, call and download, output data. The user manual describes the features of the system and the possibilities of use, the functional limitations that may be imposed on the user due to the non-compliance of the environment. The description of the control example, which demonstrates the realized possibilities of the system of aggregation of news, shows the drawings, which confirm this and describes the main way of using the system.

References

1. Antonyuk N., Chyrun L., Andrunyk V., Vasevych A., Chyrun S., Gozhyj A., Kalinina I., Borzov Y.: Medical News Aggregation and Ranking of Taking into Account the User Needs. In: CEUR Workshop Proceedings, of the 2nd International Workshop on Informatics & Data-Driven Medicine (IDDM 2019), Vol-2362, 369-382. (2019)
2. Lerman, K.: Social information processing in news aggregation. In: IEEE Internet Computing, 11(6), 16-28. (2007)
3. Paliouras, G., Mouzakidis, A., Moustakas, V., Skourlas, C. PNS: A personalized news aggregator on the web. In: Intelligent interactive systems in knowledge-based environments, Springer, Berlin, Heidelberg, 175-197. (2008)
4. Masthoff, J.: Group recommender systems: aggregation, satisfaction and group attributes. In: recommender systems handbook, Springer, Boston, MA, 743-776. (2015)
5. DeVito, M. A.: From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. In: Digital Journalism, 5(6), 753-773. (2017)
6. Bharat, K., Schmitt, M., Curtiss, M., Mayer, M., Pelczarski, K., Rakowski, B., Acharya, A.: U.S. Patent No. 8,676,837. Washington, DC: U.S. Patent and Trademark Office. (2014)
7. Setty, V., Anand, A., Mishra, A., Anand, A.: Modeling event importance for ranking daily news events. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, 231-240. (2017)
8. Kavanaugh, A., Krishnan, S., Pérez-Quñones, M., Tedesco, J., Madondo, K., Ahuja, A.: Encouraging civic participation through local news aggregation. In: Information Polity, 19(1, 2), 35-56. (2014)
9. Lee, A. M., Chyi, H. I.: The rise of online news aggregators: Consumption and competition. In: International Journal on Media Management, 17(1), 3-24. (2015)
10. Basyuk, T., Vasyliuk, A., Lytvyn, V.: Mathematical Model of Semantic Search and Search Optimization. In: CEUR Workshop Proceedings, Vol-2362, 96-105. (2019)

11. Antonyuk N., Medykovskyy, M., Chyrun, L., Dverii, M., Oborska, O., Krylyshyn, M., Vysotsky, A., Tsiura, N., Naum, O.: Online Tourism System Development for Searching and Planning Trips with User's Requirements. In: *Advances in Intelligent Systems and Computing IV*, Springer Nature Switzerland AG 2020, 1080, 831-863. (2020)
12. Lozynska, O., Savchuk, V., Pasichnyk, V.: Individual Sign Translator Component of Tourist Information System. In: *Advances in Intelligent Systems and Computing IV*, Springer Nature Switzerland AG 2020, Springer, Cham, 1080, 593-601. (2020)
13. Rzhеuskyi, A., Kutyuk, O., Voloshyn, O., Kowalska-Styczen, A., Voloshyn, V., Chyrun, L., Chyrun, S., Peleshko, D., Rak, T.: The Intellectual System Development of Distant Competencies Analyzing for IT Recruitment. In: *Advances in Intelligent Systems and Computing IV*, Springer, Cham, 1080, 696-720. (2020)
14. Rusyn, B., Pohreliuk, L., Rzhеuskyi, A., Kubik, R., Ryshkovets Y., Chyrun, L., Chyrun, S., Vysotskyi, A., Fernandes, V. B.: The Mobile Application Development Based on Online Music Library for Socializing in the World of Bard Songs and Scouts' Bonfires. In: *Advances in Intelligent Systems and Computing IV*, Springer, 1080, 734-756. (2020)
15. Vasyl Andrunyk, Volodymyr Pasichnyk, Natalya Antonyuk, Tetiana Shestakevych: A Complex System for Teaching Students with Autism: The Concept of Analysis. Formation of IT Teaching Complex. In: *Advances in Intelligent Systems and Computing IV*, Springer Nature Switzerland AG 2020, Springer, Cham, volume 1080, PP. 721-733. (2020)
16. Babichev, S.: An Evaluation of the Information Technology of Gene Expression Profiles Processing Stability for Different Levels of Noise Components. In: *Data*, 3 (4), art. no. 48 doi: 10.3390/data3040048 (2018)
17. Babichev, S., Durnyak, B., Pikh, I., Senkivskyy, V.: An Evaluation of the Objective Clustering Inductive Technology Effectiveness Implemented Using Density-Based and Agglomerative Hierarchical Clustering Algorithms. In: *Advances in Intelligent Systems and Computing*, 1020, 532-553 doi: 10.1007/978-3-030-26474-1_37 (2020)
18. Bidiuk, P., Gozhyj, A., Kalinina, I.: Probabilistic Inference Based on LS-Method Modifications in Decision Making Problems. In: *Lecture Notes in Computational Intelligence and Decision Making*, 1020, 422-433. (2020)
19. Chyrun, L., Chyrun, L., Kis, Y., Rybak, L.: Automated Information System for Connection to the Access Point with Encryption WPA2 Enterprise. In: *Lecture Notes in Computational Intelligence and Decision Making*, 1020, 389-404. (2020)
20. Kis, Y., Chyrun, L., Tsymbaliak, T., Chyrun, L.: Development of System for Managers Relationship Management with Customers. In: *Lecture Notes in Computational Intelligence and Decision Making*, 1020, 405-421. (2020)
21. Chyrun, L., Kowalska-Styczen, A., Burov, Y., Berko, A., Vasevych, A., Pelekh, I., Ryshkovets, Y.: Heterogeneous Data with Agreed Content Aggregation System Development. In: *CEUR Workshop Proceedings, Vol-2386*, 35-54. (2019)
22. Chyrun, L., Burov, Y., Rusyn, B., Pohreliuk, L., Oleshek, O., Gozhyj, ., Bobyk, I.: Web Resource Changes Monitoring System Development. In: *CEUR Workshop Proceedings, Vol-2386*, 255-273. (2019)
23. Gozhyj, A., Chyrun, L., Kowalska-Styczen, A., Lozynska, O.: Uniform Method of Operative Content Management in Web Systems. In: *CEUR Workshop Proceedings, Vol-2136*, 62-77. (2018)
24. Chyrun, L., Gozhyj, A., Yevseyeva, I., Dosyn, D., Tyhonov, V., Zakharchuk, M.: Web Content Monitoring System Development. In: *CEUR Workshop Proceedings, Vol-2362*, 126-142. (2019)

25. Veres, O., Rusyn, B., Sachenko, A., Rishnyak, I.: Choosing the Method of Finding Similar Images in the Reverse Search System. In: CEUR Workshop Proceedings, Vol-2136, 99-107. (2018)
26. Vysotska, V., Hasko, R., Kuchkovskiy, V.: Process analysis in electronic content commerce system. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT 2015, 120-123 (2015)
27. Lytvyn, V., Vysotska, V.: Designing architecture of electronic content commerce system. In: Computer Science and Information Technologies, Proc. of the X-th Int. Conf. CSIT'2015, 115-119 (2015)
28. Lytvyn, V., Vysotska, V., Veres, O., Rishnyak, I., Rishnyak, H.: The Risk Management Modelling in Multi Project Environment.. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 32-35 (2017)
29. Korobchinsky, M., Vysotska, V., Chyrun, L., Chyrun, L.: Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News, In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 52-57 (2017)
30. Lytvyn, V., Vysotska, V., Burov, Y., Veres, O., Rishnyak, I.: The Contextual Search Method Based on Domain Thesaurus. In: Advances in Intelligent Systems and Computing, 689, 310-319 (2018)
31. Kanishcheva, O., Vysotska, V., Chyrun, L., Gozhyj, A.: Method of Integration and Content Management of the Information Resources Network. In: Advances in Intelligent Systems and Computing, 689, Springer, 204-216 (2018)
32. Naum, O., Chyrun, L., Kanishcheva, O., Vysotska, V.: Intellectual System Design for Content Formation. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 131-138 (2017)
33. Su, J., Sachenko, A., Lytvyn, V., Vysotska, V., Dosyn, D.: Model of Touristic Information Resources Integration According to User Needs, 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 – Proceedings 2, 113-116 (2018)
34. Vysotska, V., Fernandes, V.B., Emmerich, M.: Web content support method in electronic business systems. In: CEUR Workshop Proceedings, Vol-2136, 20-41 (2018)
35. Vysotska, V., Lytvyn, V., Burov, Y., Gozhyj, A., Makara, S.: The consolidated information web-resource about pharmacy networks in city, CEUR Workshop Proceedings, 239-255 (2018)
36. Gozhyj, A., Kalinina, I., Vysotska, V., Gozhyj, V.: The method of web-resources management under conditions of uncertainty based on fuzzy logic, 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 – Proceedings 1, 343-346 (2018)
37. Gozhyj, A., Vysotska, V., Yevseyeva, I., Kalinina, I., Gozhyj, V.: Web Resources Management Method Based on Intelligent Technologies, Advances in Intelligent Systems and Computing, 871, 206-221 (2019)
38. Su, J., Vysotska, V., Sachenko, A., Lytvyn, V., Burov, Y.: Information resources processing using linguistic analysis of textual content. In: Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications, Romania, 573-578, (2017)
39. Vysotska, V.: Linguistic Analysis of Textual Commercial Content for Information Resources Processing. In: Modern Problems of Radio Engineering, Telecommunications and Computer Science, TCSET'2016, 709–713 (2016)
40. Lytvyn, V., Sharonova, N., Hamon, T., Vysotska, V., Grabar, N., Kowalska-Styczen, A.: Computational linguistics and intelligent systems. In: CEUR Workshop Proceedings, Vol-2136 (2018)

41. Vysotska, V., Chyrun, L.: Analysis features of information resources processing. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 124-128 (2015)
42. Lytvyn, V., Kuchkovskiy, V., Vysotska, V., Markiv, O., Pabyrivskyy, V.: Architecture of system for content integration and formation based on cryptographic consumer needs, 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018 – Proceedings 1, 391-395 (2018)
43. Vysotska, V., Chyrun, L., Chyrun, L.: Information Technology of Processing Information Resources in Electronic Content Commerce Systems. In: Computer Science and Information Technologies, CSIT'2016, 212-222 (2016)
44. Vysotska, V., Rishnyak, I., Chyrun, L.: Analysis and evaluation of risks in electronic commerce, CAD Systems in Microelectronics, 9th International Conference, 332-333 (2007).
45. Lytvyn, V., Vysotska, V., Demchuk, A., Demkiv, I., Ukhanska, O., Hladun, V., Kovalchuk, R., Petruchenko, O., Dzyubyk, L., Sokulska, N.: Design of the architecture of an intelligent system for distributing commercial content in the internet space based on SEO-technologies, neural networks, and Machine Learning. In: Eastern-European Journal of Enterprise Technologies, 2(2-98), 15-34. (2019)
46. Vysotska, V., Lytvyn, V., Burov, Y., Berezin, P., Emmerich, M., Fernandes, V. B.: Development of Information System for Textual Content Categorizing Based on Ontology. In: CEUR Workshop Proceedings, Vol-2362, 53-70 (2019)
47. Lytvyn, V., Vysotska, V., Kuchkovskiy, V., Bobyk, I., Malanchuk, O., Ryshkovets, Y., Pelekh, I., Brodyak, O., Bobrivetc, V., Panasyuk, V.: Development of the system to integrate and generate content considering the cryptocurrent needs of users, Eastern-European Journal of Enterprise Technologies 1(2-97), pp. 18-39 (2019)
48. Demchuk, A., Lytvyn, V., Vysotska, V., Dilai, M.: Methods and Means of Web Content Personalization for Commercial Information Products Distribution. In: Advances in Intelligent Systems and Computing, 1020, 332–347. (2020)
49. Vysotska, V., Burov, Y., Lytvyn, V., Oleshek, O.: Automated Monitoring of Changes in Web Resources. In: Advances in Intelligent Systems and Computing, 1020, 348–363. (2020)
50. Lytvyn, V., Vysotska, V., Shatskykh, V., Kohut, I., Petruchenko, O., Dzyubyk, L., Bobrivetc, V., Panasyuk, V., Sachenko, S., Komar, M.: Design of a recommendation system based on Collaborative Filtering and machine learning considering personal needs of the user. In: Eastern-European Journal of Enterprise Technologies, 4(2-100), 6-28. (2019)
51. Vysotska, V., Chyrun, L.: Methods of information resources processing in electronic content commerce systems. In: Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015-February. (2015)
52. Andrunyk, V., Chyrun, L., Vysotska, V.: Electronic content commerce system development. In: Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015-February. (2015)
53. Aliksieieva, K., Berko, A., Vysotska, V.: Technology of commercial web-resource processing. In: Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015-February. (2015)
54. Emmerich, M., Lytvyn, V., Yevseyeva, I., Fernandes, V. B., Dosyn, D., Vysotska, V.: Preface: Modern Machine Learning Technologies and Data Science (MoML&T&DS-2019). In: CEUR Workshop Proceedings, Vol-2386. (2019)