Using Big Data in E-tourism Mobile Recommender Systems: a Project Approach

Olga Artemenko[1][0000-0002-4057-1217], Volodymyr Pasichnyk [2][0000-0001-9434-563X],

Halyna Korz[2][0000-0002-0296-7069], Pavlo Fedorka[3][0000-0002-9242-5588],

and Yaroslav Kis[2][0000-0003-3421-2725]

[1]PHEI "Bukovinian University", Chernivtsi, Ukraine
[2]Lviv Polytechnic National University, Lviv, Ukraine
[3]Uzhgorod National University, Uzhgorod, Ukraine

olga.hapon@gmail.com, vpasichnyk@gmail.com,
halyna.i.korzh@lpnu.ua, fedorkapavlo@gmail.com,
Yaroslav.P.Kis@lpnu.ua

Abstract. This paper describes main modern tendencies for the design and development of e-tourism recommender systems with big data analytics. This study is an attempt to systematize and summarize knowledge about the possibilities of using e-tourism big data in mobile e-tourism recommender systems. In particular, to analyze the sources and types of tourist data generated by the tourist gadget, that can be related to e-tourism big data. This research focuses on the first stage of the project lifecycle for creating a mobile recommender system using e-tourism big data to filter those that best meet the interests of a particular user. Some solutions have been designed and methodological tools analyzed for more efficient use of various types of e-tourism big data from a user's gadget to be operated by a recommender system. In this study, big data for the e-tourism industry will be considered not only as a set of approaches, tools and methods for processing structured and unstructured touristic data of huge volumes.

Keywords: e-tourism, mobile recommender systems, trip support, big data, context analysis.

1   Introduction

The project activity is being implemented in all spheres of society. Project management is particularly active on a multidisciplinary basis. The development and implementation of recommender systems is based on the PMBOK methodology, which promotes the efficient organization, planning, management of labor, financial

and logistical resources throughout all stages of the project cycle. Tourism is one of the most important divertissements for modern society. Nowadays one of the most common ways to plan your trip is to look for information in various digital information resources. Moreover, the modern user of digital information space is interested not only in the background information. With the development of mobile technology and social networks, an increasing number of users are sharing their travel experiences through social media, which in turn is accumulating large amounts of thematic (travel and related) data. Such information arrays form a special category of data that can be called e-tourism big data.

Large amounts of information come from users of social networks, wireless networks and mobile devices. Therefore, information technologies of the "smart tourism" class have developed. They function in the paradigms "tourist as a sensor (data source)" and "every tourist is an expert" [1]. Consequently, information and technology processes for analyzing large amounts of data in the field of tourism, including personalized ones, are the funds for developing new "smart" destinations in tourism businesses, services, and ways of managing tourist and financial flows.

A wide range of digital communication tools and context-oriented recommender systems, along with Big Data analysis processes, form the technological basis of smart tourism. Eventually, smart e-tourism recommender systems must simultaneously communicate with the user (group of users, user with physical disabilities), the peripheral background, as well as with the community and society [2].

With the resources of the Internet, the amount of linearly accumulated data is growing at an incredible rate. According to statistics, users create social traffic that accumulates big data via the Internet in 60 seconds. This means, within a few minutes, new data fill up and change the information array to be analyzed, transformed and visualized to build a recommendation for a particular tourist object or direction. Much of this growth is due to social networks, which, in addition to providing researchers with the ability to use large amounts of accessible and constantly updated data, also allow the e-tourism recommender systems to provide more personalized, effective recommendations [3].

In this study, big data for the tourism industry will be considered not only as a set of approaches, tools and methods for processing structured and unstructured touristic data of huge volumes [4]. These methods can be applied to both large and small data sets. After all, gadget users are also a source for large arrays of useful e-tourism data, which has all the characteristics of big data [5].

There are many different techniques for analyzing large and complicated data sets, but almost every one of them is based on tools borrowed from statistics and data mining (eg machine learning, visualization, etc.) [6-20].

This study is an analysis of the technologies and methods for extraction, accumulation and processing of emerging in the tourism industry big data, in particular those exported from a mobile device of a tourist, and the potential opportunities for their use in e-tourism mobile recommender systems.

2   Project Planning

After formulating the need for project implementation, a detailed plan was developed that takes into account the project scope, risks, timing, the need to create an effective team.

The project of creating an e-tourism recommender system is implemented in four stages:

Step 1. Analysis of the project environment, which has a direct impact on the project.

Step 2. Formulation of the project concept - goals, objectives of the project implementation strategy.

Step 3. Identification of the ways and methods to achieve the project goals.

Step 4. Project implementation and achieving the goals.

Adequate identification of the environment in which the e-tourism recommender system will operate as a result of project implementation is of supreme importance, since the project is a product of that environment and is designed to meet its needs. Therefore, the viability of the project depends largely on how deeply the project environment is analyzed from the point of view of its relationship with the external environment.

Since the completeness and consistency of planning and execution of all work determine the success of the project, at the planning stage all the main tasks are identified as well as all the necessary time and resources are carefully calculated. The implementation of the project consists of many activities, such as assessment of the team's ability to implement the project, terms of reference, resource planning and project workflows. At the stage of strategy formation, they determine the ultimate goals of the project and identify ways to achieve them. An important requirement for defining project goals is the ability to quantify them by volume, timing, etc.

3   Using Big Data in Mobile E-tourism Recommender Systems

This research focuses on the first stage of the project lifecycle for creating an e-tourism recommender system as a class of intelligent systems that provide recommendations for users of various travel services, based on the processing of multiple information resources in order to filter those that best meet the interests of a particular user according to the information in his profile.

The basis for successful operation of mobile e-tourism recommender system is the fast and correct processing of real-time data. There are certain technological features in working with input and output data while designing e-tourism recommender systems, in particular mobile ones.

Making recommendations by a mobile e-tourism recommender system is based on three main sources of input:

1) the user as an information source: generates queries, leaves feedback, distributes messages about himself and services received on social networks

2) the user's gadget - information about the tourist's external background, contextual

data of all kinds, etc.

3) transactions for searching content and the Internet of things - data from guidance resources both tourist and external, work schedules, lists of tourist places and establishments, public transport timetables, etc., including web search data, user net surfing history and online booking data.
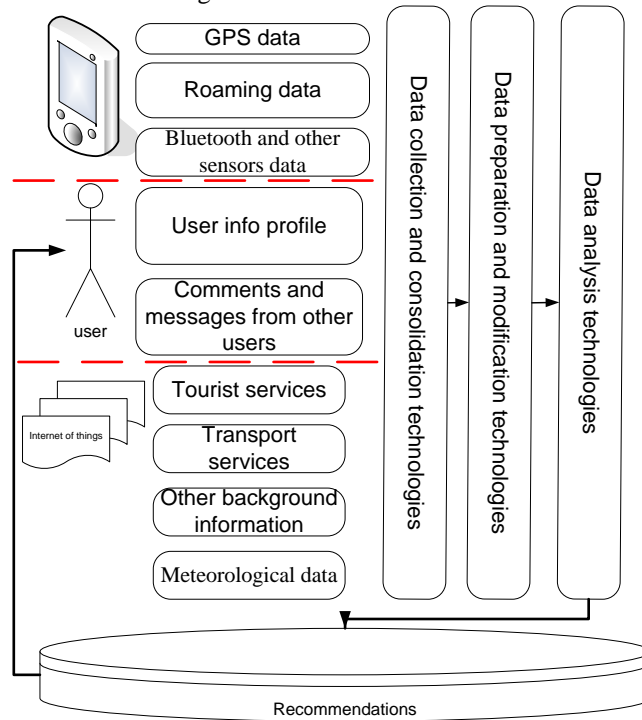

Fig. 1. The information technology of processing big data in mobile e-tourism recommender systems

As a result, to provide better outcomes, e-tourism recommender systems need not only big data collected from different sources, but also methods of their processing, which allow to analyze variously incomplete, poorly structured and unstructured information in a distributed way. These include many different techniques for analyzing large arrays of poorly structured data, based on tools borrowed from statistics and data mining (such as machine learning, visualization, knowledge mining, and more). More accurate and relevant data are obtained from the analysis of a larger and more diversified array [8].

4   Tourist as a Big Data Source for Recommender System

Information from online resources and social networks has dramatically changed the way travel is planned, maintained and displayed, providing a convenient platform for sharing user-generated data. Such data is a major category of e-tourism big data.

The results of the analysis of data generated by users are used in e-tourism recommender systems to promote tourism products [9]. The most useful for this purpose are textual data such as reviews of users for tourism services and data in various formats from personal blogs posted on social networks, as well as metadata from photos.
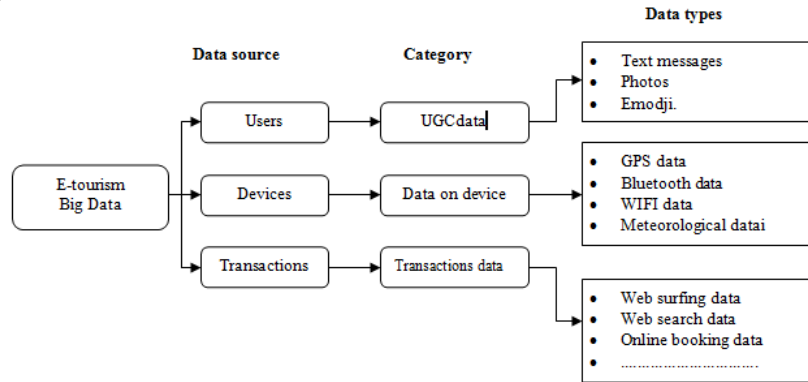

Fig. 2. Big data sources for the e-tourism recommender system

The modern information technology and social networks have led to the appearance of new ways and approaches to circulation of information in the digital space. Hashtags, emoji, geo-positioning, online access to photo and video content, live media resources complement textual content distributed by tourists.
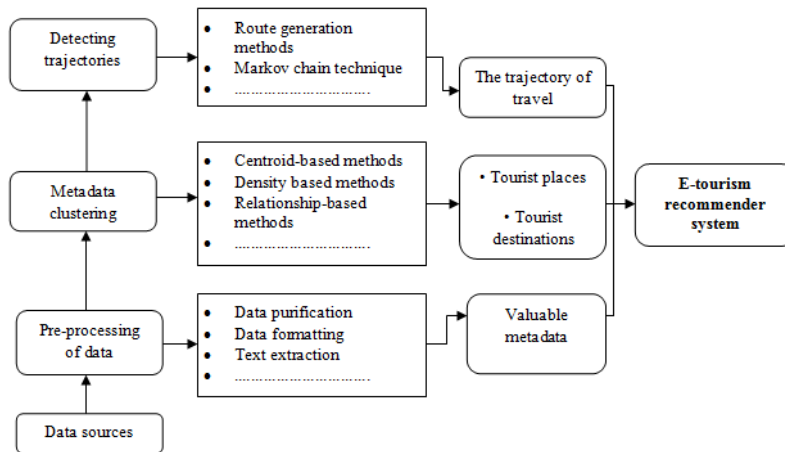

Fig. 3. The process of knowledge extraction from photo and video data in the e-tourism recommender system

The application of user-generated content in e-tourism recommender systems, including social media and video hosting services, helps to track and analyze the structure and dynamics of tourists' preferences, to obtain information about the image

and reputation of the tourism product, and to verify the tourists' behavior during the trip [10].

User-uploaded photos and video content contain a wealth of useful information that helps the recommender application to analyze personal behavior, travel experience, and preferences of the tourist. The data about geographical location added to them helps personalize recommendations. Therefore, the creation of effective tools of exporting and analyzing user-generated content distributed on social networks (in particular, the isolation of tourist content from general) improves the functionality of the e-tourism recommender systems.

5 Location-based and Trajectory Data

With the intensive development of the Internet of things, various sensors have been developed and integrated for mobile devices to provide tracking of the users' location, their trajectory, their movements, information requests and more. As follows can store huge amounts of GPS, mobile roaming, Bluetooth and WIFI data. In addition, meteorological station automatic sensors collect meteorological data, on the basis of which, using big data technology, weather conditions can be predicted at a particular location, which is relevant and useful for travel planning [11].

A whole class of location-based recommender mobile applications uses GPS and other spatial data. Their basic principle of work is to analyze requests and user behavior to further provide recommendations in the format of the optimal predicted travel destination for the tourist.

For more efficient use of GPS data in location-based mobile recommender systems, geo-location information goes through three main stages:
• Estimation the practicability and usefulness of data;
• Analysis of the users' trajectory;
• Creating recommendations based on them.

The first stage determines the feasibility and usefulness of extracting GPS data to build a recommendation and improve the performance of the recommendation system. After all, not all decisions the tourist makes regarding the priority of the location and the trajectory.
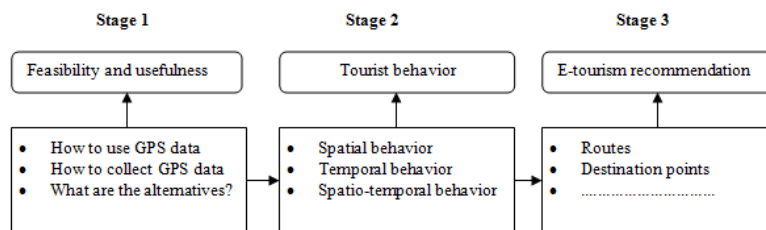

Fig. 4. Information technology process of GPS data application in tourist location-based recommender systems

The second stage determines the traveling behavior of the user of the mobile e-tourism

application, namely spatial, temporal and spatio-temporal behavior. The essence of spatial behavior is to track the GPS coordinates of the user to pinpoint his or her route. Unlike spatial behavior, the essence of temporal behavior is to determine the length of being in a particular location. As a result of the analysis of the two previously described tourist trajectories, a spatio-temporal behavior, consisting of temporal and spatial behavior, appears, the purpose of this analysis is to determine the best predicted geographical location with the longest predicted time of staying in it.

In the third stage of the estimation process, a recommendation for the tourist is generated. Based on the previous two steps, a recommendation is made regarding the potential places visited and an optimal (usually the shortest) route to that next location is formed.

With the rapid development of telecommunication technologies, roaming services provided by mobile network operators are also a tool for tracking tourist behavior.

Mobile roaming data is actively used in the e-tourism recommender systems. Mobile roaming data is collected using radio waves that are sent and received by the base station and stored automatically in the memory or log files of the mobile network operators. When a mobile phone is registered in a certain place but is used elsewhere, its user can be identified as a potential tourist.

Because traveling is a movement of people between relatively remote geographical locations, tourists typically use their mobile phone within or outside their country. For privacy reasons, users, that is, tourists and mobile network operators, are reluctant to share private information. Therefore, until recently, mobile roaming data has not been widely used in tourism research. Compared to GPS data, the use of mobile roaming data for tourism research is less useful and technologically more difficult to use for analyzing tourist behavior. Their analysis helps to track the tourist flows, the distances between the objects, the time it takes to overcome them, and the information about re-visiting a certain place.

GPS data helps to analyze the various components of the e-tourism recommendations. Specifically, GPS-based data keeps track of spatial movements throughout the journey, while mobile roaming data only provides location information of the user when the mobile phone is being actively used on the mobile network. This applies to outgoing and incoming calls, sending and receiving messages, and data transmission via the mobile Internet [12]. However, the accuracy of mobile roaming data is somewhat lower than that of GPS data.

Bluetooth technology is used to track user data over short distances, to analyze new aspects of tourist behavior, such as in museums, large infrastructure buildings such as airports, stadiums, and the like. At the same time, this technology is not used enough in the e-tourism recommender systems.
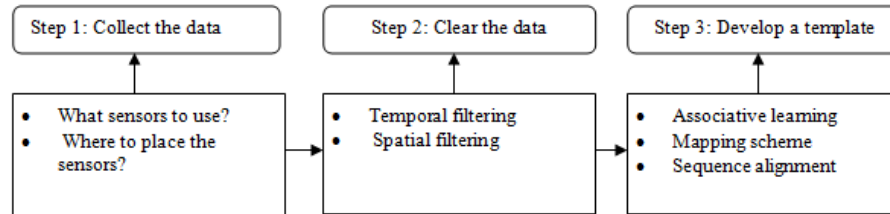
```
┌─────────────────────────┐   ┌─────────────────────────┐   ┌─────────────────────────┐
│  Step 1: Collect the data│   │  Step 2: Clear the data │   │ Step 3: Develop a template│
└─────────────────────────┘   └─────────────────────────┘   └─────────────────────────┘
```

Fig. 5. The technology of using Bluetooth data in mobile e-tourism recommender systems

Bluetooth technology has both advantages and disadvantages for the task of tracking user movements. Tracking with the user's Bluetooth location is price tag effective and convenient. Bluetooth technology is extremely convenient to use in a crowded location (such as in the middle of a building). However, the use of Bluetooth may endanger the privacy of the smartphone user or the disclosure of his private information.

Wifi communication in e-tourism recommender systems can used as an alternative to Bluetooth technology. However, Wifi as a source of users data has many advantages and disadvantages. When extracting data using Wifi, the functionality of the recommender system must take into account the limitations of its range.

The commercial use of RFID began more than 20 years ago and has proven to be useful for improving service operations such as tourist tracking and personal security.

Meteorological data is an important source of information on which recommendations for tourists are generated. Meteorological data is also a typical type of tourist big data in terms of complexity, scale and unstructured data. Meteorological data is automatically collected by meteorological sensors. The weather itself is classified according to the characteristics of air, surface, radiation, marine, agricultural, cryosphere data, physical atmospheric data, meteorological data, solar data, analytical data, meteorological catastrophes, historical, soil and vegetation data etc. This data is stored in different formats: text, images, audio, video, XML and HTML. Therefore, the processing of meteorological data in different categories and formats is challenging. On the other hand, every mobile device has software for displaying weather data both at the user's location and upon requested one.

6   Smartphone Transaction Data

The third important section of tourist big data is smartphone transaction data. Transactions are considered here in terms of travel research as Internet search queries, web page visits, online booking requests and the online purchase of tickets, different travel goods and services. Transaction data can be used in tourism applications to promote tourism products, forecasting, search engine optimization, tracking tourist behavior and marketing in terms of tourism.

Search engines are the major sources of big data for travel mobile applications that record the search operations of the Internet for content related to tourism. In particular, tourists can search for travel information through the search engines of their mobile

devices, leaving a search trail. Such traces are recorded and processed to form arrays of user interest data [13].

Also, web search data have also proven as useful for tourism marketing. This helps travel businesses and organizations get a picture of the popularity of searching for specific data by keywords.

In addition to web search data, there are other transaction data associated with visit, reservation, and online shopping transactions. However, these data are much less used in tourism research because most of them are not publicly available. However, if a gadget user gives access to such data to a recommender mobile application, it can be quite useful. Web page visit data can also help the recommender system to understand the behavior of visitors to the Internet services, such as how they interact with them. Based on a regression model, tracking direct and indirect visits can predict tourist behavior and user preferences.
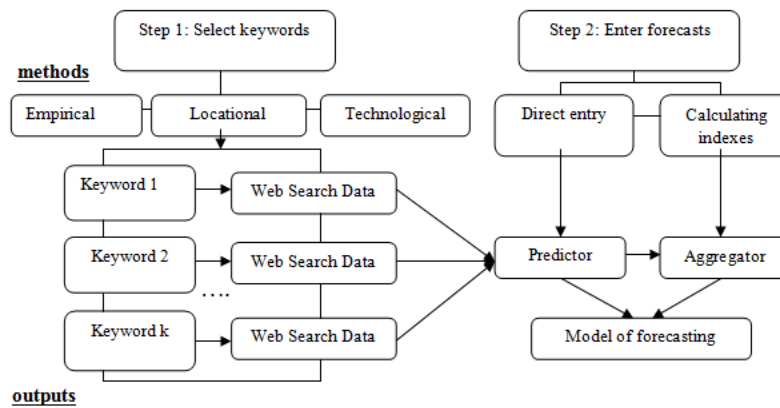

Fig. 6. The process of using web search data to predict tourists' preferences

For example, important type of user information is online hotel reservation operations, based on which you can understand the tourist behavior, determine the characteristics of how people are booking hotels, pricing and location affiliation. Also, regular electronic purchases data are also useful; analyzing them the recommender system can predict trend purchases of tourism products.

7   Discussion and Conclusion

All the categories of data described above can be collected or accumulated directly or indirectly by a user's gadget. For many tourism tasks, each category of data has a different degree of relevance and usefulness. In addition, the use of some of them may be restricted by the user who cares about privacy.

It is worth noting that alternative and interchangeable sources of the same data are considered. Therefore, when creating a mobile recommender application, it should be considered the tools and methods for collection and verification from alternative

sources.

This study is an attempt to systematize and summarize knowledge about the possibilities of using e-tourism big data in mobile e-tourism recommender systems. In particular, to analyze the sources and types of tourist data generated by the tourist gadget, that can be related to e-tourism big data.

Some solutions have been designed and methodological tools analyzed for more efficient use of various types of e-tourism data from a user's gadget to be operated by a recommender system.

The research has led to development of algorithms for searching, extracting and using content and context from mobile devices for generating recommendations in the e-tourism recommender systems.

References

1. Xiang, Z., Fesenmaier, D. R.: Big Data analytics, tourism design and smart tourism. tourism on the verge, 299–307 (2016).
2. D' Amore, M., Baggio, R., Valdani, E.: A practical approach to Big Data in tourism: A low cost Raspberry Pi cluster. Information and Communication Technologies in Tourism 2015, 169–181 (2014).
3. Sigala, M., Rahimi, R., & Thelwall, M. (Eds.). Big Data and innovation in tourism, travel, and hospitality (2019).
4. Li, J., Xu, L., Tang, L., Wang, S., Li, L.: Big data in tourism research: A literature review. Tourism Management 68, 301–323 (2018).
5. Yang, Y., Stienmetz, J. L.: Big data and tourism planning. Information Technology & Tourism 20(1-4), 189–190 (2018).
6. Cheng, M.: Media discourse on Big Data and tourism attractions in China. Perspectives on Asian Tourism, 215–229 (2018).
7. Gajdošík, T.: Big Data analytics in smart tourism destinations. A new tool for destination management organizations? WissenKompakt Medizin, 15–33 (2019).
8. Artemenko O., Kunanets O., Pasichnyk V.: E-tourism recommender systems: a survey and development perspectives. Econtechmod an international quarterly journal 6(2), 91-95 (2017).
9. Ricci F., Rokach L., Shapira B.: Recommender systems handbook: second edition. Springer Science+Business Media New York, (2015).
10. Tao, M., Nawaz, M. Z., Nawaz, S., Butt, A. H., Ahmad, H.: Users' acceptance of innovative mobile hotel booking trends: UK vs. PRC. Information Technology & Tourism (2018).
11. Scaglione, M., Johnson, C., Favre, P.: As time goes by: last minute momentum booking and the planned vacation process. Information Technology & Tourism, 21(1), 9–22 (2018).
12. Trattner, C., Oberegger, A., Marinho, L., Parra, D.: Investigating the utility of the weather context for point of interest recommendations. Information Technology & Tourism 19(1-4), 117–150 (2018).
13. Delic, A., Neidhardt, J., Nguyen, T. N., Ricci, F.: An observational user study for group recommender systems in the tourism domain. Information Technology & Tourism 19(1-4), 87–116 (2018).

14. Rzheuskyi, A., Matsuik, H., Veretennikova, N., Vaskiv R.: Selective dissemination of information – technology of information support of scientific research. Advances in Intelligent Systems and Computing III 871, 235–245 (2018).
15. Veretennikova, N., Lozytskyi, O., Vaskiv, R., Kunanets, O., Legeza, A., Lozynska, O., Kunanets, N.: Information and technology support for the training of visually impaired people. CEUR Workshop Proceedings. In: proceedings of the 8th International conference on "Mathematics. Information Technologies. Education" MoMLeT&DS, vol. 2386, pp. 307–320 (2019).
16. Kunanets, N., Matsuik, H.: Use of the smart city ontology for relevant information retrieval. In: proceedings of the 3rd International conference on computational linguistics and intelligent systems, COLINS-2019, Kharkiv, Ukraine, vol. 2362, pp. 322–333 (2019).
17. Rzheuskyi, A., Kunanets, N., Kut, V.: Methodology of research the library information services: the case of USA university libraries. Advances in Intelligent Systems and Computing II 689, 450–460 (2019).
18. Kunanets, N., Rzheuskiy, A., Veretennikova, N., Kut, V.: Information support of virtual research teams by means of cloud managers. International Journal of Intelligent Systems and Applications 10 (2), 37–46 (2018).
19. Veretennikova, N., Kunanets, N. E.: Recommendation systems as an information and technology tool for virtual research teams. Advances in Intelligent Systems and Computing II. Cham, 577-587 (2018).
20. Kunanets, N., Matsiuk, H.: Application of Saaty method while choosing thesaurus view model of the "smart city" subject domain for the improvement of information retrieval efficiency. In: 13 International Scientific and Technical Conference on Computer Science and Information Technologies (CSIT), vol. 2, 21–25 (2018).